ICM

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6—14

SECTIONS 9–11

EDITED BY D. BELIAEV AND S. SMIRNOV

EMS PRESS

ICM INTERNATIONAL CONGRESS OF MATHEMATICIANS
2022 JULY 6—14

SECTIONS 9–11

EDITED BY D. BELIAEV AND S. SMIRNOV

INTERNATIONAL MATHEMATICAL UNION
IMU

EMS PRESS

Editors

Dmitry Beliaev
Mathematical Institute
University of Oxford
Andrew Wiles Building
Radcliffe Observatory Quarter
Woodstock Road
Oxford OX2 6GG, UK

Email: belyaev@maths.ox.ac.uk

Stanislav Smirnov
Section de mathématiques
Université de Genève
rue du Conseil-Général 7–9
1205 Genève, Switzerland

Email: stanislav.smirnov@unige.ch

2020 Mathematics Subject Classification: 00B25

ISBN 978-3-98547-058-7, eISBN 978-3-98547-558-2, DOI 10.4171/ICM2022

Volume 1. Prize Lectures
ISBN 978-3-98547-059-4, eISBN 978-3-98547-559-9, DOI 10.4171/ICM2022-1
Volume 2. Plenary Lectures
ISBN 978-3-98547-060-0, eISBN 978-3-98547-560-5, DOI 10.4171/ICM2022-2
Volume 3. Sections 1–4
ISBN 978-3-98547-061-7, eISBN 978-3-98547-561-2, DOI 10.4171/ICM2022-3
Volume 4. Sections 5–8
ISBN 978-3-98547-062-4, eISBN 978-3-98547-562-9, DOI 10.4171/ICM2022-4
⟶ **Volume 5. Sections 9–11**
ISBN 978-3-98547-063-1, eISBN 978-3-98547-563-6, DOI 10.4171/ICM2022-5
Volume 6. Sections 12–14
ISBN 978-3-98547-064-8, eISBN 978-3-98547-564-3, DOI 10.4171/ICM2022-6
Volume 7. Sections 15–20
ISBN 978-3-98547-065-5, eISBN 978-3-98547-565-0, DOI 10.4171/ICM2022-7

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Published by EMS Press, an imprint of the

European Mathematical Society – EMS – Publishing House GmbH
Institut für Mathematik
Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin, Germany

https://ems.press

Typesetting using the authors’ LaTeX sources: VTeX, Vilnius, Lithuania
Printed in Germany
♾ Printed on acid free paper

CONTENTS

VOLUME 1

THE WORK OF THE FIELDS MEDALISTS AND THE IMU PRIZE WINNERS

PRIZE LECTURES

POPULAR SCIENTIFIC EXPOSITIONS

SUMMARIES OF PRIZE WINNERS' WORK

VOLUME 2

SPECIAL PLENARY LECTURES

PLENARY LECTURES

VOLUME 3

1. LOGIC

2. ALGEBRA

3. NUMBER THEORY – SPECIAL LECTURE

3. NUMBER THEORY

4. ALGEBRAIC AND COMPLEX GEOMETRY – SPECIAL LECTURE

4. ALGEBRAIC AND COMPLEX GEOMETRY

VOLUME 4

5. GEOMETRY – SPECIAL LECTURES

5. GEOMETRY

6. TOPOLOGY

7. LIE THEORY AND GENERALIZATIONS

8. ANALYSIS – SPECIAL LECTURE

8. ANALYSIS

VOLUME 5

9. DYNAMICS

10. PARTIAL DIFFERENTIAL EQUATIONS

11. MATHEMATICAL PHYSICS – SPECIAL LECTURE

11. MATHEMATICAL PHYSICS

VOLUME 6

12. PROBABILITY – SPECIAL LECTURE

12. PROBABILITY

13. COMBINATORICS – SPECIAL LECTURE

13. COMBINATORICS

14. MATHEMATICS OF COMPUTER SCIENCE – SPECIAL LECTURES

14. MATHEMATICS OF COMPUTER SCIENCE

VOLUME 7

15. NUMERICAL ANALYSIS AND SCIENTIFIC COMPUTING

16. CONTROL THEORY AND OPTIMIZATION – SPECIAL LECTURE

16. CONTROL THEORY AND OPTIMIZATION

17. STATISTICS AND DATA ANALYSIS

18. STOCHASTIC AND DIFFERENTIAL MODELLING

19. MATHEMATICAL EDUCATION AND POPULARIZATION OF MATHEMATICS

20. HISTORY OF MATHEMATICS

9. DYNAMICS

ON A CURIOUS PROBLEM AND WHAT IT LEAD TO

MIKLÓS ABÉRT

ABSTRACT

This note tells a story of an open problem on the asymptotic behavior of the minimal number of generators of groups that motivated several of my research directions.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 20F69; Secondary 22D40, 20F05, 60G55, 22E40

KEYWORDS

Rank gradient, cost, Benjamini–Schramm convergence, weak containment, point processes on locally compact groups, Heegaard genus, homology torsion growth

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6—14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3374–3387
DOI 10.4171/ICM2022/181

Published by EMS Press

In this note I will attempt to tell the story of an open problem on the minimal number of generators of groups that I am interested in for a long time and that motivated several of my research directions, sometimes in surprising ways. As stories go, the focus on the protagonist (the rank gradient problem) tends to do injustice to the other characters. This means that for some of the connected math described, even major results will get suppressed. I will also attempt to build up the character subjectively and from its birth, so not every result will be stated in its strongest form immediately, like Pallas Athene jumping out of her father's head in her full strength. In any case, the note circles around unsolved problems, which is the very opposite of this image. I will try to mitigate the damages with side stories and remarks. Finally, I believe that the truth, even mathematical, is inherently subjective and is born from a dialogue of people arriving from infinitely far. I attempt to tell this story from my own perspective but this should not be taken as a suggestion on my role in the projects I describe.

For a discrete group Γ, let $d(\Gamma)$ denote the minimal number of generators of Γ, that is, the minimal size of a subset S of Γ that generates Γ. We will also call this the *rank* of Γ, although the word "rank" is used by a lot of other notions already. We are interested in the case when Γ is also *residually finite*, that is, the intersection of its subgroups of finite index is the trivial subgroup. A rich source of residually finite groups is finitely generated matrix groups.

The rank is a rather mysterious invariant, already for finite groups. While natural (geometric) generating sets give suggestions for the rank, other generating sets may beat them to the punch. The finite symmetric group $\mathrm{Sym}(n)$ can be generated by all transpositions (i, j) and also by the neighboring transpositions $(i, i + 1)$. But it can also be generated by just 2 elements. On this track, we know that every finite simple group can be generated by at most 2 elements, but as of now, this only follows from the classification of finite simple groups.

When dealing with infinite groups, the picture does not get clearer, either. Virtually the only general way to bound the rank from below is to use the first homology, and when this does not help, one has to play it by the ears. A beautiful exception is the Grushko–Neumann theorem. The rank is not only hard to control for abstract groups. When Γ arises as the fundamental group of a nice manifold, say, one would expect that a minimal generating set, as a family of loops, will carry some geometric meaning. While there are examples when this is indeed the case, in general this is too much to hope for.

When an invariant of a residually finite group is rather unruly, one can attempt to stabilize it by looking at its growth over its subgroups of finite index and hope that this will give a more robust invariant. The biggest success story here is L^2 cohomology, or, more generally, spectral theory and representation theory, as we will discuss later on. On the geometric side, this means that instead of the defining manifold or complex M of Γ, we look at the family of finite sheeted coverings of M and try to build a geometric understanding of asymptotic homotopy on these spaces.

Rank gradient. By the Nielsen–Schreier theorem, when Γ is a free group, and H is a finite index subgroup of Γ, we have $d(H) - 1 = (d(\Gamma) - 1)|\Gamma : H|$. In other words, the number

$$r(\Gamma, H) = \frac{d(H) - 1}{|\Gamma : H|}$$

is constant $d(\Gamma) - 1$ when Γ is free, hence for an arbitrary Γ, we have the inequality $r(\Gamma, H) \le d(\Gamma) - 1$. A little exercise then shows that for $K \le H \le \Gamma$, we have $r(\Gamma, K) \le r(\Gamma, H)$. This implies that for a *chain* of finite index subgroups $\Gamma = H_0 \ge H_1 \ge \cdots$, the limit

$$\mathrm{RG}\big(\Gamma, (H_n)\big) = \lim_{n\to\infty} r(\Gamma, H_n)$$

will exist. We call this the *rank gradient* of Γ with respect to the chain (H_n). The notion comes from Marc Lackenby [32], see also an early profinite version in [36]. One can also define it to an arbitrary subset $\{H_n\}$ of finite index subgroups of Γ as

$$\mathrm{RG}\big(\Gamma, \{H_n\}\big) = \inf_n r(\Gamma, H_n).$$

Many years ago, we started to study this notion by ourselves with Nik Nikolov, proved some initial results using elementary group theory, and soon realized that we do not know a convincing example for when the rank gradient in fact depends on the chain. A nonconvincing example comes from $\Gamma = F_2 \times F_2$: normal chains in Γ with trivial intersection have rank gradient zero, but chains that only walk down on one of the factors have positive rank gradient. We could not find an example, however, when the H_n are normal subgroups of finite index and their intersection is trivial. We still cannot.

Problem 1 (Rank gradient). *Let Γ be finitely generated and let (H_n) and (K_n) be normal chains in Γ with trivial intersection. Does*

$$\mathrm{RG}\big(\Gamma, (H_n)\big) = \mathrm{RG}\big(\Gamma, (K_n)\big)?$$

What do you do when you encounter an elusive but attractive invariant? 1. Prove that it vanishes in some natural cases; 2. Try to look for translations or analogues in other fields and try to make mathematical energy flow through; 3. Connect it to some other, maybe tamer invariants; 4. Extend the notion wildly and see what happens. In what follows, I will describe some attempts of these points and where they lead.

The cost correspondence. A good starting exercise for the reader is to prove by hand that if Γ has a central element of infinite order, then the rank gradient vanishes for any normal chain with trivial intersection. After proving some starting results like this with Nik Nikolov on rank gradient, we managed to connect the rank gradient to cost.

The notion of cost was introduced by Gilbert Levitt [34] and most of the subsequent, deep work on it was done by Damien Gaboriau [24]. I will not define the notion here, just state that every probability measure preserving (p.m.p.) action of a countable group Γ has a cost, which is a real number between 1 and $d(\Gamma)$. A major question on cost is the following [24].

Problem 2 (Fixed price). *Let Γ be a countable group. Does every free p.m.p. action of Γ have the same cost?*

In his hallmark result [24], Damien Gaboriau showed that this is true for free groups. Since the cost only depends on the equivalence relation spanned by the action, it immediately follows that the free groups F_2 and F_3 are not orbit equivalent, a well-known open problem at the time.

In [9] we established the following correspondence. For a chain (Γ_n) in Γ, one can associate its coset tree $T(\Gamma, (\Gamma_n))$ as follows. The vertex set of T is the union of cosets Γ / Γ_n and the edges are defined by immediate inclusion of cosets. The group Γ acts on T by automorphisms and this action extends to the boundary ∂T of T as a continuous action. There is a natural measure on the boundary (the product measure on infinite walks) and the action preserves this measure.

Theorem 3 (Cost correspondence). *Let Γ be finitely generated and let (Γ_n) be a normal chain in Γ with trivial intersection. Then*

$$\mathrm{RG}\big(\Gamma, (\Gamma_n)\big) = \mathrm{c}\big(\Gamma, (\Gamma_n)\big) - 1,$$

where $\mathrm{c}(\Gamma, (\Gamma_n))$ *denotes the cost of* Γ *acting on* $\partial T(\Gamma, (\Gamma_n))$.

So, for chains, the rank gradient problem is a special case of the fixed price problem for profinite actions.

The existing cost theory immediately gave new vanishing results on rank gradient for a large class of groups, including amenable groups and more importantly, the so called *right angled* groups. These are groups that admit a list of generators of infinite order such that neighboring generators commute. It is an important class as it contains many nonuniform lattices, like $\mathrm{SL}(3, \mathbb{Z})$.

Looking at the cost literature, we also realized that a seemingly innocent result on cost would actually positively solve the RG problem. The question is whether the cost minus 1 is multiplicative for finite-index subrelations, just like rank minus 1 is for free groups. The result was announced to be solved at the time with versions of a preprint circulating but by now the community agrees that it should be considered unsolved. I still think that this could lead to a fruitful attack on either fixed price, or the rank gradient problem.

The rank vs. Heegaard genus problem. In our project with Nik Nikolov, we studied Marc Lackenby's work and its topological motivations [32, 33] and got aware that using his results on Heegaard genus and expansion [31], proving the vanishing of rank gradient for $\Gamma = \mathrm{SL}(2, \mathbb{Z}[i])$ would solve a famous old problem in 3-manifold theory. The problem that is still open is whether for finite volume 3-manifolds, the ratio of the Heegaard genus and the rank can get arbitrarily large. Note that for hyperbolic manifolds, it was also open for a long time whether the rank can even *differ* from the Heegaard genus. Now this is solved by Tao Li [35]. The deal for the ratio is that by [31] the Heegaard genus grows linearly for any chain of subgroups with property (τ) and it is easy to produce a normal chain in Γ with vanishing rank gradient, since it is virtually a finitely generated free by cyclic group. So if the rank gradient is independent of the chain, then a chain of principal congruence subgroups in Γ will have property (τ) and hence positive Heegaard genus growth but vanishing rank gradient, which makes the ratio of the two invariants go to infinity. On the other hand, if the rank gradient may depend on the chain, then the fixed price problem is solved negatively.

That is, we managed to show that at least one of these well-studied problems have a negative solution, but we still do not know which one(s). This is certainly a good joke, but a word of caution is due here. It could very well be that eventually *both* problems have a negative solution but for entirely different reasons, and then our bridge between them will not prove to be useful, as no one walks through it.

Homology growth and Lück approximation. The first rational homology b_Q of a group is a trivial lower bound for its rank, in fact, it is the only general lower bound people use. As a consequence, the growth of homology satisfies

$$\lim_{n\to\infty} \sup \frac{b_Q(\Gamma_n)}{|\Gamma : \Gamma_n|} \leq \mathrm{RG}\big(\Gamma, (\Gamma_n)\big).$$

The good news here is that when Γ is finitely presented, a famous theorem of Wolfgang Lück [37] implies that the limit of the left-hand side exists and is independent of the chain.

Theorem 4 (Lück approximation). *Let Γ be a finitely presented group and let (Γ_n) be a chain of normal subgroups of finite index in Γ with trivial intersection. Then we have*

$$\lim_{n\to\infty} \frac{b_Q(\Gamma_n)}{|\Gamma : \Gamma_n|} = \beta^1(\Gamma),$$

where $\beta^1(\Gamma)$ is the first L^2 Betti number of Γ.

The L^2 story that starts here is quite extensive and beautiful (see [37] and around), but for us what matters now is that for finitely presented groups we have

$$\mathrm{RG}\big(\Gamma, (H_n)\big) \geq \beta^1(\Gamma).$$

We still do not know an example where there is a proper inequality here. Note that for a classical group theorist, it sounds quite weird that the abelianization of Γ_n should asymptotically control its rank! For instance, this suggests that if the Γ_n are perfect groups, then by some miracle we should be able to generate them by fewer elements than the trivial bound.

When turning to the measured setting and use the cost correspondence, this takes the form of a question already asked in the initial paper of Gaboriau [24]. He shows that for every free action of Γ the cost of the action minus 1 is at least $\beta^1(\Gamma)$ and no one knows an example when they are not equal.

I will not state the full generality of the Lück approximation result, but need to make some side comments here. First, the proof is really about spectral convergence. It is easy to see that for normal chains with trivial intersection, the eigenvalue distribution of any locally defined operator on the finite quotient will weakly converge to the spectral measure of the same operator in the limit. The gist of Lück approximation is to show that the measure of the set $\{0\}$ will also converge. This is a tightness result that does not follow from weak convergence in general. The result was later generalized by Andreas Thom [41] for arbitrary real values instead of 0.

Graph sequences, combinatorial cost, and the Farber condition. In our project leading to [9], we also studied what happens with arbitrary instead of normal subgroups or when we ease up on having trivial intersection. These were not arbitrary questions. First, we hoped to find counterexamples easier in this bigger class. Second, in a lot of cases when one is interested in the asymptotic behavior of an invariant on a family of finite index subgroups, they do not form a chain and are not normal. For instance, in number theory we often care about the congruence subgroups $\Gamma_0(N)$: they are not normal and do not form a chain. These questions has lead to a connection to graph limit theory.

What connects two different chains of normal subgroups in Γ with trivial intersection? The best answer I know is that they are locally indistinguishable. That is, from every vertex, the corresponding Schreier graphs locally look more and more like the infinite Cayley graph of Γ. When you only ask that the same holds for *most* vertices, you get the notion of Benjamini–Schramm convergence [13].

In the homological direction, we found Michael Farber's extension of the Lück approximation theorem [22] and the subsequent work of Nicolas Bergeron and Damien Gaboriau [14] on when and how such an extension may fail. It is clear that the Farber condition is equivalent to asking that the action of Γ on the boundary of the coset tree is essentially free. We called these chains *Farber chains*.

For a fixed generating set S of Γ, one can visualize the chain (Γ_n) by looking at the sequence of finite Schreier graphs $(\mathrm{Sch}(\Gamma/\Gamma_n, S))$ and attempt to understand rank gradient using the asymptotic metric geometry of this graph sequence. Now we can state what the Farber condition is, in various ways. For a permutation action of Γ and $g \in \Gamma$, let $\mathrm{Fix}(g, \Gamma/\Gamma_n)$ denote the number of fixed points of g.

Proposition 5. *Let Γ be a group generated by the finite symmetric set S and let (Γ_n) be a sequence of subgroups in Γ. Then the following are equivalent:*

1. *For every $1 \neq g \in \Gamma$, we have*
$$\lim_{n\to\infty} \frac{\mathrm{Fix}(g, \Gamma/\Gamma_n)}{|\Gamma : \Gamma_n|} = 0 \quad \textit{(Farber condition);}$$
2. *A random conjugate of Γ_n as an invariant random subgroup weakly converges to the trivial one;*
3. $\mathrm{Sch}(\Gamma/\Gamma_n, S)$ *Benjamini–Schramm converges to* $\mathrm{Cay}(\Gamma, S)$*;*
4. *The coset actions of Γ on Γ/Γ_n form a sofic approximation of Γ.*

These equivalences are all easy once one learns the language. However, these forms are important to note, as they highlight the fields that got connected by this theory: in order, representation theory, ergodic theory, graph limits, and soficity.

An example for a Farber sequence is the above mentioned congruence subgroups $\Gamma_0(p)$ (p prime). Let me remark that the notion of Farber sequence can be naturally extended to a sequence of lattices in a fixed locally compact group. In [12] that some people call the 7 samurai paper, we prove that in a higher rank simple Lie group, *every* sequence of lattices with covolume tending to infinity is Farber! We use invariant random subgroups in the proof, a notion that was coined (but not invented) in [6]. There would be a lot to tell here, but this story is about rank gradient, so we stop at this point.

When looking at it as a graph theory problem, the rank gradient problem asks if the asymptotic rank is a *local invariant*, or whether it depends on global properties of $\mathrm{Sch}(\Gamma/\Gamma_n, S)$.

The idea of bringing graph limit theory to cost and L^2 theory is due to Gabor Elek [21], who in an early paper [20] defined the combinatorial cost of a graph sequence and proved the analogues of the known results on cost and homology growth in this setting. A quick definition of combinatorial cost is as follows. For a sequence of finite graphs, a *bi-Lipshitz rewiring* of the sequence is another graph sequence using the same vertex sets, such that there exists a constant L where the distance of every edge in one graph can be substituted by a path of length at most L in the other. The combinatorial cost is the infimum of edge densities that can be achieved by such a rewiring. Note that much later, two groups consisting of Alessandro Carderi, Damien Gaboriau, and Mikael de la Salle, and me and Laszlo Toth, respectively, independently showed that using an ultraproduct language [19] or local–global convergence [10], the combinatorial cost is, in fact, equal to the cost of a suitable limiting object. But by then, the damage was done and graph limit theory was affecting the field in various ways.

When you ease up on normality, the intersection of the subgroups really will not matter, even for chains, and it is the Farber condition that will really affect the behavior. By [24], every aperiodic p.m.p. action of an amenable group has vanishing cost, but the rank gradient correspondence only works for Farber chains. Indeed, it is not true that for an amenable group the rank gradient vanishes for any normal chain, an easy counterexample comes from the lamplighter group. However, Marc Lackenby [33] showed that for *finitely presented* amenable groups, the rank gradient vanishes for arbitrary normal chains, that is, trivial intersection is not needed there. By pushing his trichotomy theorem a bit further, together with Nik Nikolov and Andrei Jaikin-Zapirain, we managed to show that for finitely presented amenable groups, the rank gradient also vanishes for arbitrary chains [7]. This is one of the examples I know where a result on rank gradient does not seem to have an immediate cost counterpart.

Weak containment. Using the above observation on local behavior, with Gabor Elek we attempted to solve the rank gradient problem by showing that the Schreier graphs of any two normal chains in the same group can be asymptotically "massaged onto each other" by almost covering maps. We soon realized that what we look at is already investigated for p.m.p. actions by Alekos Kechris [29] under the name weak containment and is also strongly connected to the notion of local–global convergence introduced by Bela Bollobas and Oliver Riordan [16] and developed by Hamed Hatami, Laszlo Lovasz, and Balazs Szegedy [27]. Ironically, we found that, in fact, the opposite of what we attempted holds, and proved the following rigidity result in [4].

Theorem 6 (Weak containment rigidity). *If a strongly ergodic p.m.p. action of Γ weakly contains a finite action of Γ then it factors onto it. In particular, if two normal chains in Γ with property (τ) define weakly equivalent coset tree actions, then the two chains are refinements of each other.*

From the point of view of the rank gradient problem, this can be considered as a harsh no entry sign, but it does show that arithmetic lattices admit uncountably many weakly inequivalent p.m.p. actions.

Although weak containment has not yet been successful to prove new results on rank gradient, let me mention a quite elegant application by Alekos Kechris [30]. A result of Lewis Bowen [17] implies that for the free group F_n, its profinite completion weakly contains any free p.m.p. action of F_n. By the cost monotonicity result for weak containment [29], this implies that among free p.m.p. actions of F_n, the cost is *minimal* for the profinite completion. Now using the cost–rank-gradient correspondence above, one yields that the cost of any free action of F_n is at least n, hence F_n has fixed price n, giving an alternate proof for the famous starting result of Damien Gaboriau.

The graph limit language suggested another possible attempt at the rank gradient problem, using a factor of iid generating set for the Γ_n. For a Farber chain, the quotient Schreier graphs look like the infinite limiting Cayley graph from most points. So, one may apply the same rule using an iid seed, to get a cheap generating set, and since the seed was iid, the resulting cheap rewiring also works for any Farber sequence the same way. This attempt also did not work (yet) but it did eventually lead to the result with Benjy Weiss [11] that every free action of a countable group Γ weakly contains its iid actions, hence showing that iid actions of Γ have maximal cost.

Homology torsion growth. When trying to interpret the fixed price 1 result of Damien Gaboriau for right angled groups with Tsachik Gelander and Nik Nikolov in the finite setting, we realized that there is an interesting rewiring complexity notion hiding behind it and that the notion can be used to prove vanishing of the first homology torsion growth.

More precisely, when building the cheap rewiring on the finite level, using Damien Gaboriau's trick, not only the rewiring gets cheap, but at the same time its complexity also stays low. In particular, it gives cost $1 + \varepsilon$ with a bi-Lipschitz constant that is polynomial in $1/\varepsilon$. We then realized that this is enough to prove not only the vanishing of rank gradient, but also the vanishing of the first homology torsion growth.

For a finitely presented group, it is easy to show that the size of the torsion part of the abelianization of a subgroup is at most exponential in the index of the subgroup. Hence, the right growth notion to consider is

$$\mathrm{t}\big(\Gamma, (\Gamma_n)\big) = \lim_{n\to\infty} \frac{\log(\mathrm{tor}(\Gamma_n))}{|\Gamma : \Gamma_n|},$$

assuming this limit exists. Torsion homology growth is studied by various groups for various reasons, see [15] and references therein. In [5] we prove the following.

Theorem 7 (First torsion homology growth). *Let Γ be a right angled group and let (Γ_n) be a Farber sequence in Γ. Then*

$$\mathrm{t}\big(\Gamma, (\Gamma_n)\big) = 0.$$

In fact, the proof works for any Farber sequence where the above defined by-Lipschitz constant is subexponential in the index. This brought an interesting connection to the Bergeron–Venkatesh conjecture [15], that made me understand something about how unruly the notion of rank may be in reality.

A special case of the Bergeron–Venkatesh conjecture says that for a principal congruence chain in $\Gamma = \mathrm{SL}(2, \mathbb{Z}[i])$, the first torsion homology growth is a positive constant. If we believe this, and also that the rank gradient is zero for these chains (these are both tall orders of course), then the previous theorem implies that while these congruence subgroups may admit cheap generating sets, their complexity must be exponential in the error $1/\varepsilon$. That is, we may not be able to find them in a nice and geometric way as they hide deep in the cruel and dark embrace of algebra.

Vanishing theorems can be cool, but they tend to emit a somewhat pessimistic aura. After all, at the end, we reach zero. However, in the case of first homology torsion growth, currently no one can do better, as the following is still open.

Problem 8. *Is there a finitely presented group Γ and a Farber sequence (Γ_n) in Γ such that $\mathrm{t}(\Gamma, (\Gamma_n)) > 0$?*

While there are lower bounds for the torsion, currently they do not get this high.

It is a natural question whether there is a natural "higher rank" notion of being right angled so that the above vanishing theorem generalizes for higher homology torsion. Recently this was addressed in the paper [2], together with Nicolas Bergeron, Mikolaj Fraczyk, and Damien Gaboriau.

Uniform rank gradient and Poisson processes. As we discussed above, to solve the rank vs Heegaard genus problem, it would be enough to effectively estimate the rank of principal congruence subgroups of $\mathrm{SL}(2, \mathbb{Z}[i])$. Note that this approach is a blessing and a curse at the same time. Indeed, while the ambient group and its congruence subgroups seem very concrete, they are inherently number-theoretic which means that any attempt would also involve some possibly rather nontrivial number theory. In fact, the same could be said when we want to estimate the rank gradient of any discrete group, using geometric methods. Indeed, unless the group is of some quite special form, like close to free, right angled, or amenable, the geometry of its Cayley graphs seem quite complicated.

It turns out, however, that when we ask for much more, it immediately forces our hand in a good way and seems to give a much simpler image to deal with. Let G be a locally compact group, but for simplicity just concentrate on when G is a simple real Lie group. When Γ is a lattice in G, it is finitely generated, moreover, by the work of Tsachik Gelander [26], we have

$$d(\Gamma) \leq C \operatorname{vol}(G/\Gamma),$$

where C is an absolute constant. The notion of Farber sequences make perfect sense, using either Benjamini–Schramm convergence of the quotient spaces G/Γ (or X/Γ where X is the symmetric space for G) or invariant random subgroups.

Problem 9. *Let G be a semisimple Lie group and let (Γ_n) be a Farber sequence of lattices in G. Does*

$$\mathrm{RG}\big(G, (\Gamma_n)\big) = \lim \frac{d(\Gamma_n) - 1}{\mathrm{vol}(G/\Gamma_n)}$$

exist?

If it does, then the limit is independent of the sequence, since you can merge two Farber sequences and they stay Farber. This is one of the advantages over using chains of lattices. In fact, one can define Benjamini-Schramm convergence in the realm of Riemannian manifolds [3]. On this language one gets the Poisson point process on a symmetric space as the limit of independent random subsets of the finite volume manifolds.

To address this problem, with Sam Mellick [8], we recently introduced a cost theory for point processes of locally compact groups. Note that Alessandro Carderi has already introduced the cost of p.m.p. actions of locally compact groups in his nice paper [18] and used an ultraproduct language to prove that the maximal cost of a p.m.p. action of G dominates the rank gradient, at least for uniformly discrete Farber sequences of lattices. Our approach of using point processes allows us to remove his uniform discreteness assumption and answer his question whether $G \times \mathbb{Z}$ has fixed price 1.

In the paper [8] we prove that the Poisson processes have maximal cost among free point processes and that this number dominates the rank gradient of any Farber sequence in G. This is an analogue of my theorem with Benjy Weiss for discrete groups [11], as Poisson processes are arguably the substitutes of iid actions in the locally compact setting. In particular, if the cost of the Poisson process is 1, then any free point process has cost 1 and the rank gradient vanishes for any Farber sequence of lattices.

Instead of G, one can again consider Poisson processes on its symmetric space X, as they have the same cost. In particular, to settle the rank vs Heegaard genus problem, it would be enough to show that Poisson processes on the hyperbolic space H^3 have cost 1.

Problem 10 (Poisson cost). *Does the Poisson process on H^3 of intensity* 1 *have cost* 1*?*

This seems to be a much simpler and more direct geometric-stochastic question than estimating the rank of congruence subgroups directly. On the other hand, if this cost happens to be greater than 1, this would not tell anything about the rank gradient, but it would still imply the existence of a countable equivalence relation whose cost does not equal to its first L^2 Betti number, answering a question of Damien Gaboriau (see [25] on L^2 numbers of countable equivalence relations).

Apart from the case when X is the upper half-plane, as of now, nothing is known for semisimple Lie groups. The reasonable conjecture is that for every other G, the cost of the Poisson processes should vanish. When we look at homological counterparts, we still get a nontrivial task. For rational homology growth, the 7 samurai project [12] and [1] settled most of the questions. In the direction of mod p homology growth, Mikolaj Fraczyk [23] proved in a beautiful paper that when G has higher rank and property (T), then the first mod 2 homology growth vanishes for arbitrary Farber sequences of lattices. In fact, he showed that every homology class admits a cycle of length that is sublinear in the volume. The difficulty is clearly shown in the fact that for odd primes, these are still open, although it would follow from the vanishing of the cost of Poisson processes.

Further questions on rank gradient. Kazhdan's property (T) is a strong property that can be used in manifold ways, in particular, it implies that the first L^2 Betti number vanishes. So, it makes sense to ask the following.

Problem 11 (Kazhdan groups). *Does the rank gradient vanish for finitely presented, residually finite groups with property (T)?*

In other words, does it vanish for every normal chain with trivial intersection? When switching to the ergodic side, this asks whether property (T) groups have fixed price 1. This is also open, however, Tom Hutchcroft and Gabor Pete recently showed in a recent, very nice paper [28] that such groups always admit an action with cost 1, that is, the infimal cost of Γ is 1. It would be natural to use [11] here, but the processes they ingeniously generate are not factor of iid, so their result does not establish fixed price 1, and also does not seem to settle the rank gradient problem for these groups. Nevertheless, it is still tempting to try to adapt their method somehow in the finite setting and yield a vanishing result on rank gradient.

In another direction, it would be interesting to say something meaningful on groups with positive rank gradient. Marc Lackenby's trichotomy theorem [32] gives some restrictions and also his theorem that finitely presented groups with positive p-gradient are large. On the other hand, if we omit the finite presentation condition, we will have some positive rank gradient monsters lurking around, as Denis Osin [39] and Jan-Christoph Schlage-Puchta [40] showed. A specific question due to Nik Nikolov is as follows.

Problem 12. *Can a group satisfying a nontrivial identity have a positive rank gradient?*

Nik Nikolov recently managed to show that in this case, the profinite gradient does vanish [38].

ACKNOWLEDGMENTS

Being an invited lecturer at the ICM is an honor and a significant milestone at one's mathematical career. So I wish to thank all my teachers, friends, and coauthors (not a mutually exclusive list) who taught me their mathematics. The full list would be rather long, so I only name those here who influenced my general mathematical thinking in the most substantial way: László Surányi, Balázs Szegedy, Bálint Virág, Laci Pyber, and Benjy Weiss. Related to the projects and papers described in this note, I learned a lot from Nicolas Bergeron, Ian Biringer, Gabor Elek, Mikolaj Fraczyk, Damien Gaboriau, Tsachik Gelander, and Nik Nikolov. I am grateful to all of them.

REFERENCES

[1] M. Abert, N. Bergeron, I. Biringer, and T. Gelander, Convergence of normalized Betti numbers in nonpositive curvature. 2021, arXiv:1811.02520v3.

[2] M. Abert, N. Bergeron, M. Fraczyk, and D. Gaboriau, On homology torsion growth. 2021, arXiv:2106.13051.

[3] M. Abert and I. Biringer, Unimodular measures on the space of all Riemannian manifolds. 2020, arXiv:1606.03360v5, to appear in Geometry and Topology.

[4] M. Abért and G. Elek, Dynamical properties of profinite actions. *Ergodic Theory Dynam. Systems* **32** (2012), no. 6, 1805–1835.

[5] M. Abért, T. Gelander, and N. Nikolov, Rank, combinatorial cost, and homology torsion growth in higher rank lattices. *Duke Math. J.* **166** (2017), 2925–2964.

[6] M. Abért, Y. Glasner, and B. Virág, Kesten's theorem for invariant random subgroups. *Duke Math. J.* **163** (2014), no. 3, 465–488.

[7] M. Abért, A. Jaikin-Zapirain, and N. Nikolov, The rank gradient from a combinatorial viewpoint. *Groups Geom. Dyn.* **5** (2011), no. 2, 213–230.

[8] M. Abért and S. Mellick, Point processes, cost, and the growth of rank in locally compact groups. 2021, arXiv:2102.07710.

[9] M. Abért and N. Nikolov, Rank gradient, cost of groups and the rank vs Heegaard genus conjecture. *J. Eur. Math. Soc. (JEMS)* **14** (2012), no. 5, 1657–1677.

[10] M. Abért and L. Toth, Uniform rank gradient, cost, and local-global convergence. *Trans. Amer. Math. Soc.* **373** (2020), 2311–2329.

[11] M. Abért and B. Weiss, Bernoulli actions are weakly contained in any free action. *Ergodic Theory Dynam. Systems* **33** (2013), no. 2, 323–333.

[12] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, and I. Samet, On the growth of L^2-invariants for sequences of lattices in Lie groups. *Ann. of Math.* **185** (2017), 711–790.

[13] I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.* **6** (2001), no. 23, 13 pp.

[14] N. Bergeron and D. Gaboriau, Asymptotique des nombres de Betti, L^2-invariants et laminations. *Comment. Math. Helv.* **79** (2004), no. 2, 362–395.

[15] N. Bergeron and A. Venkatesh, The asymptotic growth of torsion homology for arithmetic groups. *J. Inst. Math. Jussieu* **12** (2013), no. 2, 391–447.

[16] B. Bollobas and O. Riordan, Sparse graphs: metrics and random models. *Random Structures Algorithms* **39** (2011), 1–38.

[17] L. Bowen, Periodicity and circle packing in the hyperbolic plane. *Geom. Dedicata* (2003), 213–236.

[18] A. Carderi, Asymptotic invariants of lattices in locally compact groups. 2018, arXiv:1812.02133.

[19] A. Carderi, D. Gaboriau, and M. de la Salle, Non-standard limits of graphs and some orbit equivalence invariants. 2021, arXiv:1812.00704v3.

[20] G. Elek, The combinatorial cost. *Enseign. Math. (2)* **53** (2007), no. 3–4, 225–235.

[21] G. Elek and E. Szabo, Hyperlinearity, essentially free actions and L^2-invariants, the sofic property. *Math. Ann.* **332**, no. 2, 421–441.

[22] M. Farber, Geometry of growth: approximation theorems for L^2 invariants. *Math. Ann.* **311** (1998), no. 2, 335–375.

[23] M. Fraczyk, Growth of mod-2 homology in higher-rank locally symmetric spaces. *Duke Math. J.* **171** (2022), no. 2, 247–271.

[24] D. Gaboriau, Coût des relations d'équivalence et des groupes. *Invent. Math.* **139** (2000), no. 1, 41–98.

[25] D. Gaboriau, Invariants L2 de relations d'équivalence et de groupes. *Publ. Math. Inst. Hautes Études Sci.* **95** (2002), 93–150.

[26] T. Gelander, Volume versus rank of lattices. *J. Reine Angew. Math.* **661** (2011), 237–248.

[27] H. Hatami, L. Lovasz, and B. Szegedy, Limits of local-global convergent graph sequences. *Geom. Funct. Anal.* **24** (2014), no. 1, 269–296.

[28] T. Hutchcroft and G. Pete, Kazhdan groups have cost 1. *Invent. Math.* **221** (2020), 873–891.

[29] A. Kechris, *Global aspects of ergodic group actions*. Math. Surveys Monogr. 160, American Mathematical Society, Providence, RI, 2010.

[30] A. Kechris, Weak containment in the space of actions of a free group. *Israel J. Math.* **189** (2012), 461–507.

[31] M. Lackenby, The asymptotic behaviour of Heegaard genus. *Math. Res. Lett.* **11** (2004), no. 2–3, 139–149.

[32] M. Lackenby, Expanders, rank and graphs of groups. *Israel J. Math.* **146** (2005), 357–370.

[33] M. Lackenby, Heegaard splittings, the virtually Haken conjecture and property τ. *Invent. Math.* **164** (2006), no. 2, 317–359.

[34] G. Levitt, On the cost of generating an equivalence relation. *Ergodic Theory Dynam. Systems* **15** (1995), 1173–1181.

[35] T. Li, Rank and genus of 3-manifolds. *J. Amer. Math. Soc.* **26** (2013), 777–829.

[36] A. Lubotzky, L. van den Dries, Subgroups of free profinite groups and large subfields of Q. *Israel J. Math.* **39** (1981), 25–45.

[37] W. Lück, *L^2-invariants: theory and applications to geometry and K-theory*. Ergeb. Math. Grenzgeb. (3) 44, Springer, Berlin, 2002.

[38] N. Nikolov, On profinite groups with positive rank gradient. 2022, arXiv:2104.09094v3.

[39] D. Osin, Rank gradient and torsion groups. *Bull. Lond. Math. Soc.* **43** (2011), 10–16.

[40] J-C. Schlage-Puchta, A p-group with positive rank gradient. *J. Group Theory* **15** (2012), 261–270.

[41] A. Thom, Sofic groups and Diophantine approximation. *Comm. Pure Appl. Math.* **61**, no. 8, 1155–1171.

MIKLÓS ABÉRT

Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13-15, 1053 Budapest, Hungary, abert.miklos@renyi.mta.hu

LATTICE SUBGROUPS ACTING ON MANIFOLDS

AARON BROWN

ABSTRACT

We discuss recent progress in understanding rigidity properties of smooth actions of higher-rank lattices. We primarily discuss questions of existence in low dimensions (Zimmer's conjecture), classification in the smallest possible dimension, and further classification assuming dynamical properties of the action. Two common themes arise in the proofs: (1) dynamical properties of the lattice action are mimicked by certain measures on an induced G-space; (2) such measures often exhibit additional rigidity properties. Throughout, we state some open problems and possible directions for future research.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37C85; Secondary 57M60, 22E40, 37D20, 37D25

KEYWORDS

Rigidity, Zimmer program, lattices in Lie groups, Anosov diffeomorphisms, Lyapunov exponents

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3388–3411
DOI 10.4171/ICM2022/163

1. INTRODUCTION: LATTICES, GROUP ACTIONS, AND RIGIDITY

1.1. Rigidity of linear representations

For $n \geq 2$, consider the group $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ of $n \times n$ integer matrices with determinant 1 or a more general lattice subgroup of $\mathrm{SL}(n, \mathbb{R})$. There is a stark distinction between the case $n = 2$ and $n \geq 3$; in particular, relative to various group- and representation-theoretic properties, the group $\Gamma = \mathrm{SL}(2, \mathbb{Z})$ is rather "flexible" whereas the group $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ is very "rigid" whenever $n \geq 3$.

Indeed, when $n = 2$, linear representations $\rho\colon \Gamma \to \mathrm{GL}(d, \mathbb{R})$ are very flexible. In contrast, when $n \geq 3$, linear representations $\rho\colon \Gamma \to \mathrm{GL}(d, \mathbb{R})$ exhibit many well-known rigidity properties; we highlight local rigidity of the inclusion $\iota\colon \Gamma \to \mathrm{SL}(n, \mathbb{R})$ [62, 64], local rigidity of general representations $\pi\colon \Gamma \to \mathrm{GL}(d, \mathbb{R})$ [52, 59, 65], and Mostow's strong rigidity [50, 53, 56]. The principle result that includes those above is Margulis' superrigidity theorem. Roughly, Margulis' theorem states (when $n \geq 3$) that any representation $\rho\colon \Gamma \to \mathrm{GL}(d, \mathbb{R})$ coincides—up to a compact error—with the restriction of a continuous representation $\pi\colon \mathrm{SL}(n, \mathbb{R}) \to \mathrm{GL}(d, \mathbb{R})$. Since representations of $\mathrm{SL}(n, \mathbb{R})$ are classified, this more-or-less classifies all representations of Γ.

1.2. The general setting

Throughout, G will be a connected noncompact semisimple Lie group. We will always assume the Lie algebra of G is simple and say that G is a simple Lie group. Throughout, we will typically assume that G has higher real rank. (The Lie algebra $\mathfrak{g}$ of G admits an Iwasawa decomposition $\mathfrak{g} = \mathfrak{k}\mathfrak{a}\mathfrak{n}$. The *real rank* of $\mathfrak{g}$ is $\dim(\mathfrak{a})$ and G is *higher rank* if it has real rank at least 2.) At times we may also assume G has finite center though that is not technically necessary for most results.

Such groups G admit biinvariant Haar measures. A *lattice* in G is a discrete subgroup Γ of G such that the coset space G/Γ has finite volume. A lattice Γ is *cocompact* if, in addition, the quotient G/Γ is compact; otherwise, Γ is *nonuniform*. When G is simple and is of higher real rank, we say a lattice Γ in G is a *higher-rank lattice.*

For simplicity of exposition, we formulate most results and conjectures in the case that $G = \mathrm{SL}(n, \mathbb{R})$ (though many results and conjectures hold for wider classes of groups). The real rank of $\mathrm{SL}(n, \mathbb{R})$ is $n - 1$ and thus we typically assume $n \geq 3$ to ensure we are in the higher-rank setting. The standard example of a lattice subgroup in $G = \mathrm{SL}(n, \mathbb{R})$ is the subgroup $\Gamma = \mathrm{SL}(n, \mathbb{Z})$. The subgroup $\mathrm{SL}(n, \mathbb{Z})$ is nonuniform, though we note that $\mathrm{SL}(n, \mathbb{R})$ admits cocompact lattices.

1.3. Actions on manifolds and the Zimmer program

Beyond linear representations, we might replace the vector space $\mathbb{R}^d$ with a compact manifold M and replace the finite-dimensional Lie group $\mathrm{GL}(d, \mathbb{R})$ with $\mathrm{Diff}^r(M)$, the group of all C^r-diffeomorphisms[1] of M. A homomorphism $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ then defines a

[1] If $r \geq 1$ is not integral, we write $r = k + \beta$ where $k \in \mathbb{N}$ and $\beta \in (0, 1)$ and say that $f\colon M \to M$ is C^r if it is C^k and if the kth derivatives of f are β-Hölder continuous.

C^r *action* of Γ on M. If vol is a smooth volume form on M, we also consider $\mathrm{Diff}^r_{\mathrm{vol}}(M)$, the group of volume-preserving diffeomorphisms, and study volume-preserving actions $\alpha\colon \Gamma \to \mathrm{Diff}^r_{\mathrm{vol}}(M)$.

For $\Gamma = \mathrm{SL}(2,\mathbb{Z})$, actions $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ on manifolds are again quite flexible. However, by analogy with rigidity properties of linear representations, we might ask if (possibly volume-preserving) actions of higher-rank lattices exhibit rigidity properties analogous to those that hold for linear representations. To motivate statements, it is useful to recall some standard low-dimensional, algebraically defined actions by lattices in $\mathrm{SL}(n,\mathbb{R})$.

(1) *Affine actions on tori.* Consider the case that Γ has finite index in $\mathrm{SL}(n,\mathbb{Z})$. We obtain an action $\alpha\colon \Gamma \to \mathrm{Diff}(\mathbb{T}^n)$ on the n-dimensional torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ given by $\alpha(\gamma)(x + \mathbb{Z}^n) = \gamma \cdot x + \mathbb{Z}^n$ for every matrix $\gamma \in \Gamma \subset \mathrm{SL}(n,\mathbb{Z})$. Since there exists $\gamma \in \mathrm{SL}(n,\mathbb{Z})$ with all eigenvalues outside of the unit circle, this gives an example of an affine Anosov action (see Definition 4.3). Observe that these actions preserve the Haar measure on $\mathbb{T}^d$.

(2) *Projective actions.* Given any lattice subgroup $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$, the linear action of Γ on $\mathbb{R}^n$ induces an action on the space of rays (or lines) in $\mathbb{R}^n$ through the origin. We thus obtain an action of Γ on the $(n-1)$-dimensional sphere S^{n-1} (or $\mathbb{R}P^{n-1}$). The subgroup $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ also acts on Grassmanians of higher-dimensional planes in $\mathbb{R}^n$ and on spaces of flags in $\mathbb{R}^n$. These actions are all left actions of Γ on G/Q for some parabolic subgroup $Q \subset G = \mathrm{SL}(n,\mathbb{R})$. We remark that these actions admit no Γ-invariant probability measure.

(3) *Isometric actions.* Certain cocompact lattices $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ admit representations $\pi\colon \Gamma \to \mathrm{SU}(n)$ with infinite image (see discussion in [69, SECTIONS 6.7, 6.8, WARNING 16.4.3]). The representation π then induces an isometric action of Γ on the $(2n-2)$-dimensional space $M = \mathrm{SU}(n)/\mathrm{S}(\mathrm{U}(1) \times \mathrm{U}(n-1))$.

In the early 1980s, Zimmer established a superrigidity theorem for linear cocycles over ergodic, measure-preserving actions of higher-rank Lie groups and their lattices (see [72]). The cocycle superrigidity theorem, its corollaries, and contemporaneous results of Zimmer's (see [72–77]) led Zimmer to formulate several conjectures and questions concerning (C^∞, volume-preserving) actions of higher-rank simple Lie groups and their lattices. These questions, conjectures, and more recent extensions are usually referred to as the *Zimmer program*. Roughly, the Zimmer program aims to establish analogues of rigidity results for linear representations in the setting of smooth actions on compact manifolds. See, for instance, [24] for an overview and statements of many conjectures in this area.

2. LOW DIMENSIONS AND ZIMMER'S CONJECTURE

We present some motivation, state a contemporary version of *Zimmer's conjecture*, and outline recent progress in the area. See also the article by D. Fisher in the same proceedings for related discussion.

2.1. Motivation and Zimmer's conjecture

For $n \geq 3$, let Γ be a lattice subgroup of $\mathrm{SL}(n, \mathbb{R})$. Recall the action of Γ on S^{n-1} and, assuming Γ is commensurable with $\mathrm{SL}(n, \mathbb{Z})$, the affine action of Γ on $\mathbb{T}^n$ discussed in Section 1.3. Zimmer's conjecture asserts that these represent the minimal dimensions in which nontrivial actions of such Γ could occur. To be precise, note that if $\Gamma' \subset \Gamma$ is a finite-index normal subgroup, the finite quotient group $F = \Gamma/\Gamma'$ may act on manifolds of arbitrary dimension. This induces an action of Γ that should be considered rather trivial. Assuming $\dim(M)$ is sufficiently small, Zimmer's conjecture states all actions of Γ factor through the action of a finite group.

Conjecture 2.1 (Zimmer's conjecture for lattices in $\mathrm{SL}(n, \mathbb{R})$). *For $n \geq 3$, let $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$ be a lattice subgroup. Let M be a compact manifold.*

(1) *If $\dim(M) < n - 1$, then any homomorphism $\Gamma \to \mathrm{Diff}(M)$ has finite image.*

(2) *In addition, if vol is a volume form on M and if $\dim(M) = n - 1$, then any homomorphism $\Gamma \to \mathrm{Diff}_{\mathrm{vol}}(M)$ has finite image.*

A motivation (by analogy) for this conjecture is the following corollary of Margulis' superrigidity theorem: *Let Γ be a lattice in $\mathrm{SL}(n, \mathbb{R})$ for $n \geq 3$. For any $d < n$, the image of any representation $\rho\colon \Gamma \to \mathrm{GL}(d, \mathbb{R})$ is finite.* Indeed, using that there are no nontrivial representations $\pi\colon \mathrm{SL}(n, \mathbb{R}) \to \mathrm{SL}(d, \mathbb{R})$, the image $\rho(\Gamma)$ is contained in a compact subgroup K of $\mathrm{GL}(d, \mathbb{R})$; Margulis further studies representations into compact Lie groups and shows the Lie algebra of K contains only copies of $\mathfrak{su}(n)$, the compact real form of $\mathfrak{sl}(n, \mathbb{R})$. A dimension count implies the Lie algebra of K vanishes and thus K is finite.

In the volume-preserving setting, Conjecture 2.1(2) is motivated by the following corollary of Zimmer's cocycle superrigidity theorem: *For $n \geq 3$, $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$, and $\dim(M) < n$, a volume-preserving action $\Gamma \to \mathrm{Diff}_{\mathrm{vol}}(M)$ preserves a measurable Riemannian metric on TM.* If this metric were C^0, the image $\alpha(\Gamma)$ would be contained in the compact isometry group of this metric. A dimension count again yields finiteness. Thus, if $\dim(M)$ is sufficiently small, one might expect the image $\alpha(\Gamma)$ to be contained in a compact isometry group K of M. To extend the conjecture to other groups, to each simple, noncompact Lie group G we associate 3 positive integers $v(G)$, $n(G)$, and $d(G)$ defined, roughly, as follows:

(1) $v(G)$ is the minimal dimension of G/H as H varies over all proper closed subgroups $H \subset G$. (We remark that H is a parabolic subgroup in this case.)

(2) $n(G)$ is the minimal dimension of a nontrivial linear representation of (the Lie algebra of) G.

(3) $d(G) = v(G_{\mathrm{cmt}})$ is the minimal dimension of all nontrivial homogeneous spaces of the compact real form, G_{cmt}, of G.

We also define another number, $r(G)$, first defined in [8], which arises from certain dynamical arguments. A simpler definition of $r(G)$ is the following:

(4) $r(G) = v(G')$ where G' is a maximal $\mathbb{R}$-split subgroup of G (with the same reduced restricted root system as G).

We note that $n(G)$, $d(G)$, and $v(G)$ depend only on the Lie algebra $\mathfrak{g}$ of G; $r(G)$ depends only on the restricted root system of $\mathfrak{g}$. See Tables 1, 2, and 3, in Appendix A for computations of the numbers $v(G)$, $d(G)$, $n(G)$, and $r(G)$ for various classical groups. Given the integers $n(G)$, $d(G)$, and $v(G)$, we have the following general conjecture.

Conjecture 2.2 (Zimmer's conjecture; general). *Let $\Gamma \subset G$ be a lattice in a connected higher-rank simple Lie group G. Let M be a compact manifold and let* vol *be a volume form on M.*

(1) *If $\dim(M) < \min\{n(G), d(G), v(G)\}$ then any homomorphism $\alpha \colon \Gamma \to \mathrm{Diff}(M)$ has finite image.*

(2) *If $\dim(M) < \min\{n(G), d(G)\}$ then any homomorphism $\alpha \colon \Gamma \to \mathrm{Diff}_{\mathrm{vol}}(M)$ has finite image.*

(3) *If $\dim(M) < \min\{v(G), n(G)\}$ then for any homomorphism $\alpha \colon \Gamma \to \mathrm{Diff}(M)$, the image $\alpha(\Gamma)$ preserves a Riemannian metric.*

(4) *If $\dim(M) < n(G)$ then for any homomorphism $\alpha \colon \Gamma \to \mathrm{Diff}_{\mathrm{vol}}(M)$, the image $\alpha(\Gamma)$ preserves a Riemannian metric.*

We are intentionally vague about the regularity of the action in the conjecture as it is unclear what the optimal regularity should be. In parts (3) and (4), the invariant Riemannian metric should be at least C^0. Most results discussed below require the action to be at least $C^{1+\text{Hölder}}$ though some results hold for C^1 or even C^0 actions. We note that part (3) of Conjecture 2.2 implies part (1) and part (4) implies part (2) by compactness of the isometry group of the invariant metric, superrigidity, and definition of $d(G)$.

Many prior results towards this conjecture focused on actions on the circle including [10, 32, 68] and for volume-preserving (and general measure-preserving) actions on surfaces including [29, 30, 55]. See also [31] and [22] for results on real-analytic actions and [11, 13, 14] for results on holomorphic and birational actions. There are also many results (including in the C^0 setting) for actions of specific lattices on manifolds with certain topology, where topological obstructions constrain the possible actions; a partial list of such results includes [2, 54, 66, 67, 70, 78].

2.2. Work of Brown, Fisher, and Hurtado

The series of papers [5–7] established Conjecture 2.1, Zimmer's conjecture, for C^r actions by lattices in $\mathrm{SL}(n, \mathbb{R})$.

Theorem 2.3 ([7]). *Conjecture 2.1 holds for C^r actions, $r > 1$.*

For actions by general higher-rank lattices, the same series of papers establishes the following which directly implies Theorem 2.3 (see Table 1 in Appendix A).

Theorem 2.4 ([5] cocompact case; [7] nonuniform case). *Let $\Gamma \subset G$ be a lattice in a connected higher-rank simple Lie group G. Let M be a compact manifold and let $r > 1$.*

(1) *If* $\dim(M) < r(G)$ *then any homomorphism* $\Gamma \to \mathrm{Diff}^r(M)$ *has finite image.*

(2) *In addition, if* vol *is a volume form on M and if* $\dim(M) = r(G)$ *then any homomorphism* $\Gamma \to \mathrm{Diff}^r_{\mathrm{vol}}(M)$ *has finite image.*

We outline the broad steps in the proof of Theorem 2.4. Readers interested in the case of actions by cocompact lattices in $\mathrm{SL}(n, \mathbb{R})$ may consult expository accounts in [3] and [12] for detailed proofs.

Step 1: subexponential growth. Fix a lattice subgroup Γ as in Theorem 2.4. We have that Γ is finitely generated. Given $\gamma \in \Gamma$, let $|\gamma| = |\gamma|_S$ denote the word-length of γ relative to some finite symmetric generating set S. Equip TM with a Riemannian metric.

Definition 2.5. An action $\alpha\colon \Gamma \to \mathrm{Diff}^1(M)$ has *uniform subexponential growth of derivatives* if for every $\varepsilon > 0$ there exists $C = C_\varepsilon$ such that for every $\gamma \in \Gamma$,

$$\sup_{x \in M} \left\| D_x \alpha(\gamma) \right\| \le C e^{\varepsilon|\gamma|}.$$

The following is the primary technical result established in [5–7].

Theorem 2.6 ([5, THEOREM 2.3], [7, THEOREM C]). *Let Γ and M be as in Theorem 2.4. For $r > 1$, let $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ be an action. Suppose that either*

(1) $\dim(M) < r(G)$, *or*

(2) $\dim(M) \le r(G)$ *and α preserves a smooth volume.*

Then α has uniform subexponential growth of derivatives.

Step 2: strong property (T) and averaging Riemannian metrics. The lattices Γ in Theorem 2.4 are known to have strong property (T). Strong property (T) was introduced by V. Lafforgue in [45] and shown for cocompact lattices in higher-rank groups in [16, 45] and extended to nonuniform lattices by de la Salle in [15]. An action $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ induces an action on Riemannian metrics. If $r \ge 2$, one can average over elements of this action and apply strong property (T) to obtain the following.

Theorem 2.7 ([5, THEOREM 2.4]). *Let Γ be a finitely generated group and let M be a compact manifold. For $k \ge 2$, let $\alpha\colon \Gamma \to \mathrm{Diff}^k(M)$ be an action. If α has uniform subexponential growth of derivatives and if Γ has strong property (T) then $\alpha(\Gamma)$ preserves a Riemannian metric that is $C^{k-1-\delta}$ for all $\delta > 0$.*

For $C^{1+\text{Hölder}}$ actions, the proof can be adapted to establish an analogue of Theorem 2.7. For C^1 actions, an analogue of Theorem 2.7 is obtained in [4, PROPOSITION 5].

Step 3: Margulis superrigidity. From Steps 1 and 2, the image $\alpha(\Gamma)$ is contained in a compact group K. Finiteness then follows immediately from Margulis' superrigidity, and a dimension count since (one can check) $r(G) < d(G)$; Theorem 2.4 follows.

2.3. C^1 actions

To establish a C^1 version of Conjecture 2.2, an analogue of Theorem 2.7 is given by [4, PROPOSITION 5]; it remains to establish a C^1 analogue of Theorem 2.6. However, a crucial step in the proof of Theorem 2.6 uses Pesin theory and Ledrappier–Young theory, which requires the consideration of C^r actions for $r > 1$. Still, a partial analogue of Theorem 2.6 holds under stronger constraints on the dimension of M.

Theorem 2.8 ([4]). *Let $\Gamma \subset G$ be a lattice in a connected, simple, higher-rank Lie group G. Let M be a compact manifold.*

(1) *If $\dim(M) < \operatorname{rank}(G)$, then any homomorphism $\Gamma \to \operatorname{Diff}^1(M)$ has finite image.*

(2) *In addition, if vol is a volume form on M and if $\dim(M) \leq \operatorname{rank}(G)$, then any homomorphism $\Gamma \to \operatorname{Diff}^2_{\operatorname{vol}}(M)$ has finite image.*

We note that the dimension bounds in Theorem 2.8 only coincide with the dimensions in Conjecture 2.2 in the case $G = \mathrm{SL}(n, \mathbb{R})$.

Question 2.9. Let Γ be a lattice subgroup of $G = \mathrm{Sp}(2n, \mathbb{R})$, $\mathrm{SO}(n, n)$, or $\mathrm{SO}(n, n+1)$. Do Theorem 2.6 and Conjecture 2.2(1)–(2) hold for C^1 actions of Γ?

2.4. C^0 actions and actions on the circle

Given the results of Theorem 2.4 and Theorem 2.8, it is natural to ask if any analogous results hold for actions by homeomorphisms. For actions of general higher-rank lattices, most results on C^0 actions have focused on (non-volume-preserving) actions on the circle or the interval. We mention in particular [68] where it is shown that actions of higher-$\mathbb{Q}$-rank groups Γ on the circle are finite. A recent breakthrough by B. Deroin and S. Hurtado [17] completely resolves the question of C^0 action on the circle (among many other results).

Theorem 2.10 (Corollary of [17, THEOREM 1.5]). *Let Γ be a lattice in higher-rank simple Lie group G. For every action $\alpha\colon \Gamma \to \operatorname{Homeo}(S^1)$, the image $\alpha(\Gamma)$ is finite.*

The proof of Theorem 2.10 follows somewhat the approach in the proof of Theorem 2.4 but, due to the lack of differentiability, new tools need to be developed. We mention only one novelty of working on the circle used in [17]: following [18], one may replace minimal C^0 actions with bi-Lipschitz actions.

2.5. Beyond $\mathbb{R}$-split groups

Theorem 2.4 only gives the optimal dimension bounds for Conjecture 2.2(3) (and thus Conjecture 2.2(1)) in the case of $\mathbb{R}$-split Lie groups; see Tables 1 in Appendix A. Further analysis of objects arising in the proof of Theorem 2.6 establishes the conjectured bounds in Conjecture 2.2(3) for some nonsplit groups. This was first shown for actions of lattices in $\mathrm{SL}(n,\mathbb{C})$ in [71]: *For $n \geq 3$, Conjecture 2.2(3) holds for C^r $(r > 1)$ actions of cocompact lattices in* $\mathrm{SL}(n,\mathbb{C})$. The same holds for lattices in general complex simple groups. Beyond actions by lattices in complex Lie groups, one can establish Conjecture 2.2(3) for large parameter ranges of many nonsplit Lie groups. The following (nonexhaustive list) gives some ranges where such results can be shown.

Theorem 2.11 (J. An, A. Brown, and Z. Zhang; in preparation). *Conjecture 2.2(3) holds for C^r $(r > 1)$ actions of lattices in the following Lie groups:*

(1) *all higher-rank simple complex Lie groups;*

(2) $\mathrm{SL}(n,\mathbb{H})$ *with* $n \geq 9$*;*

(3) $\mathrm{SO}^+(m,n)$ *with* $2 \leq n < m \leq \frac{1}{2}(n^2 - n + 4)$*;*

(4) $\mathrm{SU}(m,n)$ *with* $6 \leq n \leq m \leq \frac{1}{4}(n^2 - 3n + 6)$*;*

(5) $\mathrm{SO}^*(2n)$ *with* $n \geq 30$.

This naturally leads to the following question.

Question 2.12. Does Conjecture 2.2(3) hold for actions of lattices in all higher-rank simple Lie groups?

We also show some partial results towards Conjecture 2.2(4).

Theorem 2.13 (J. An, A. Brown, and Z. Zhang; in preparation). *Let Γ be a lattice in* $\mathrm{SL}(n,\mathbb{C})$ *for $n \geq 4$. If* $\dim(M) \leq n(G) - 2$ *then for any homomorphism* $\alpha\colon \Gamma \to \mathrm{Diff}^r_{\mathrm{vol}}(M)$ *$(r > 1)$, the image $\alpha(\Gamma)$ preserves a Riemannian metric.*

2.6. Dimension gaps between (3) and (4) of Conjecture 2.2

Theorem 2.4 implies all statements of Conjecture 2.2 for actions by lattices in $\mathrm{SL}(n,\mathbb{R})$ and $\mathrm{Sp}(n,\mathbb{R})$ since $r(G) = v(G) = n(G) - 1 < d(G)$. However, for the $\mathbb{R}$-split groups $G = \mathrm{SO}(n,n)$ and $G = \mathrm{SO}(n,n+1)$, we have

$$r(G) = v(G) = n(G) - 2 < d(G) = n(G) - 1 < n(G).$$

Thus, for these groups, Theorem 2.4 implies Conjecture 2.2(1)–(3) but does not imply Conjecture 2.2(4). This gap also arises for $\mathbb{R}$-split exceptional groups and many non-$\mathbb{R}$-split groups. For instance, Theorem 2.13 implies that volume-preserving actions of lattices in $\mathrm{SL}(n,\mathbb{C})$ (for $n \geq 4$) preserve a Riemannian metric if $\dim(M) \leq 2n - 2$; Conjecture 2.2(4) asserts the same should hold if $\dim(M) = 2n - 1 = n(G) - 1$.

Question 2.14. Does Conjecture 2.2(4) hold for lattices Γ in $\mathrm{SO}(n,n)$, $\mathrm{SO}(n,n+1)$, or $\mathrm{SL}(n,\mathbb{C})$, $n \geq 3$? Specifically, if $\dim(M) < n(G)$, does every volume-preserving action $\alpha\colon \Gamma \to \mathrm{Diff}^\infty_{\mathrm{vol}}(M)$ preserve a (C^0 or C^∞) Riemannian metric?

We note that every lattice Γ in $G = \mathrm{SO}(n,n)$ and $G = \mathrm{SO}(n,n+1)$ admits a nonisometric action on a compact manifold of dimension $n(G) - 1$. Indeed, there is a parabolic subgroup $Q \subset G$ with codimension $v(G) = n(G) - 2$; the left action of Γ on G/Q is nonisometric. Taking $M = (G/Q) \times S^1$, let Γ act on M naturally on the left in the first coordinate and as the identity in the second coordinate. We note that this action does not preserve any volume form on M so does not contradict Question 2.14. As a first step towards Question 2.14, we might rule out related constructions that would yield counterexamples as in the following.

Problem 2.15. Let Γ be a lattice in $G = \mathrm{SO}(n,n)$ or $G = \mathrm{SO}(n,n+1)$. Show there is no volume-preserving action of Γ on $M = (G/Q) \times S^1$ with infinite image.

As a first step towards solving Problem 2.15, one might restrict to actions that factor onto the projective action on G/Q. In a related direction, we also pose the following.

Question 2.16. For $n \geq 3$, let Γ be a lattice in $G = \mathrm{SO}(n,n)$ or $\mathrm{SO}(n,n+1)$. Suppose that $\dim(M) = n(G) - 1$ and that $\alpha \colon \Gamma \to \mathrm{Diff}^\infty(M)$ is an action that does not preserve any (C^0 or C^∞) Riemannian metric. Is there either (1) an invariant embedded G/Q in M on which the dynamics restricts to the standard action or (2) an invariant open subset $U \subset M$ restricted to which the dynamics factors onto the standard action on G/Q?

3. CLASSIFICATION IN LOWEST DIMENSIONS AND RIGIDITY OF PROJECTIVE ACTIONS

For $n \geq 3$, Theorem 2.3 implies that actions by lattices in $\mathrm{SL}(n,\mathbb{R})$ are finite when $\dim(M) < n - 1$. When $\dim(M) = n - 1$, recall the natural action of Γ on S^{n-1} or $\mathbb{R}P^{n-1}$. In work in progress, we show these to be the only actions with infinite image.

Theorem 3.1 (A. Brown, F. Rodriguez Hertz, Z. Wang; in preparation). *For $n \geq 3$, let Γ be a lattice subgroup of $\mathrm{SL}(n,\mathbb{R})$. Let M be a connected compact manifold of dimension $n - 1$. Fix $r > 1$ and let $\alpha \colon \Gamma \to \mathrm{Diff}^r(M)$ be an action with infinite image $\alpha(\Gamma)$. Then*

(1) *there is a C^r-diffeomorphism h between M and either S^{n-1} or $\mathbb{R}P^{n-1}$ such that*

(2) *for all $x \in M$ and $\gamma \in \Gamma$, $h(\alpha(\gamma)(x)) = \gamma \cdot h(x)$ where the right-hand side denotes the standard projective action of Γ on S^{n-1} or $\mathbb{R}P^{n-1}$.*

The techniques used to prove Theorem 3.1 also give local rigidity of higher-dimensional projective actions, extending the results of [38] and [44, THEOREM 17].

Theorem 3.2. *For $n \geq 3$, let $\mathcal{F}$ be a flag manifold (of flags in $\mathbb{R}^n$) and let $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ be a lattice subgroup. Then the standard action $\rho \colon \Gamma \to \mathrm{Diff}(\mathcal{F})$ is $C^{\infty,1,\infty}$-locally rigid.*

Theorem 3.2 says for any action $\alpha \colon \Gamma \to \mathrm{Diff}^\infty(\mathcal{F})$ sufficiently C^1 close to the standard projective action ρ, there exists a C^∞ diffeomorphism $h \colon \mathcal{F} \to \mathcal{F}$ such that $h \circ$

$\alpha(\gamma) = \rho(\gamma) \circ h$ for all $\gamma \in \Gamma$. The above results lead to the following question classifying all actions on flag manifolds.

Question 3.3 (Global rigidity). For $n \geq 3$, let Γ be a lattice subgroup in $\mathrm{SL}(n, \mathbb{R})$ and let $\mathcal{F}$ be a flag manifold (of flags in $\mathbb{R}^n$). Let $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(\mathcal{F})$ be an action with infinite image $\alpha(\Gamma)$. Is α smoothly conjugate to the standard projective action on $\mathcal{F}$?

4. CLASSIFICATION UNDER DYNAMICAL AND TOPOLOGICAL HYPOTHESES

4.1. Classification in dimension n

Given the classification in Theorem 3.1, it is natural to ask if it is possible to classify all (possibly volume-preserving) actions $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(M)$ when Γ is a lattice in $\mathrm{SL}(n, \mathbb{R})$ and M is a compact connected manifold of dimension n. This seems much harder since there are known examples of "exotic" actions in dimension n. In the non-volume-preserving case, there exist many nonequivalent real-analytic actions of $\mathrm{SL}(n, \mathbb{R})$ on the n-sphere constructed in [63] and the restriction to $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ yields exotic actions of Γ. Roughly, one builds a skew-product $\mathrm{SL}(n, \mathbb{R})$-action on $S^{n-1} \times (-1, 1)$ factoring onto the standard action on S^{n-1} and takes the two-point compactification. This motivates the following alternative version of Question 2.16.

Question 4.1. Let Γ be a lattice in $G = \mathrm{SL}(n, \mathbb{R})$ for $n \geq 3$. Let $\dim M = n$ and let $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(M)$ be an action with infinite image that does not preserve any volume form (or absolutely continuous measure). Does M contain an embedded projective action or an invariant open subset that factors onto the projective action on $\mathbb{R}P^{n-1}$?

In the setting of volume-preserving actions, given the affine action of $\mathrm{SL}(n, \mathbb{Z})$ on $\mathbb{T}^n$, it is possible to blowup a fixed point (or a finite Γ orbit) to obtain a smooth action on a n-manifold preserving a smooth density; in [40], A. Katok and J. Lewis showed these examples can be perturbed to preserve a smooth, nowhere vanishing density.

One might conjecture that actions of lattices $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$ in dimension n are built by gluing together modifications of standard actions such as those described above. At this time though, it seems any conjectured picture is far from understood. Thus to classify actions of $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$ in dimension n (and higher), it is natural to first impose additional dynamical or topological hypotheses. The remainder of this section discusses several results in this direction.

4.2. Toral homeomorphisms and Anosov diffeomorphisms

Given a homeomorphism $f \in \mathrm{Homeo}(\mathbb{T}^d)$, there is a unique matrix $A_f \in \mathrm{GL}(d, \mathbb{Z})$ such that any lift $\tilde{f}\colon \mathbb{R}^d \to \mathbb{R}^d$ of f is of the form $\tilde{f}(x) = A_f x + \phi(x)$ for some $\mathbb{Z}^d$-periodic $\phi\colon \mathbb{R}^d \to \mathbb{R}^d$. We call A_f the *linear data* of f and note that A_f induces an automorphism L_{A_f} on the torus $\mathbb{T}^d$. If $\alpha\colon \Gamma \to \mathrm{Homeo}(\mathbb{T}^d)$ is an action we similarly obtain $\rho\colon \Gamma \to \mathrm{GL}(d, \mathbb{Z})$ called the *linear data* of α. A matrix $A \in \mathrm{GL}(d, \mathbb{Z})$ is *hyperbolic* if no eigenvalue of A is on the unit circle. The following theorem characterizes (up to continuous semiconjugacy) maps $f\colon \mathbb{T}^d \to \mathbb{T}^d$ whose linear data A_f is hyperbolic.

Theorem 4.2 (Franks, [28]). *Let $f\colon \mathbb{T}^d \to \mathbb{T}^d$ be a homeomorphism with hyperbolic linear data A_f. There exists a continuous, surjective $h\colon \mathbb{T}^d \to \mathbb{T}^d$ such that $h \circ f = L_{A_f} \circ h$.*

We recall Anosov diffeomorphisms, which provide the main example of homeomorphisms satisfying the hypotheses of Theorem 4.2.

Definition 4.3. A C^1 diffeomorphism $f\colon M \to M$ of a compact Riemannian manifold M is *Anosov* if there is a continuous, Df-invariant splitting of the tangent bundle $TM = E^s \oplus E^u$ and constants $0 < \kappa < 1$ and $C \geq 1$ such that for every $x \in M$ and every $n \in \mathbb{N}$,

$$\left\|D_x f^n(v)\right\| \leq C\kappa^n \|v\|, \quad \text{for } v \in E^s(x), \qquad \left\|D_x f^{-n}(w)\right\| \leq C\kappa^n \|w\|, \quad \text{for } w \in E^u(x).$$

All known examples of Anosov diffeomorphisms occur on finite factors of tori and nilmanifolds. From [28,49] we have a complete classification of Anosov diffeomorphisms on tori (and nilmanifolds) up to homeomorphism.

Theorem 4.4 (Franks–Manning, [28,49]). *If $f\colon \mathbb{T}^n \to \mathbb{T}^n$ is Anosov, then f is homotopic to L_A for some hyperbolic $A \in \mathrm{GL}(n, \mathbb{Z})$; moreover, there is a homeomorphism $h\colon \mathbb{T}^n \to \mathbb{T}^n$ such that $h \circ f = L_A \circ h$.*

4.3. Global topological and smooth rigidity of Anosov actions

For simplicity, consider $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$. An action $\alpha\colon \Gamma \to \mathrm{Diff}(M)$ is *Anosov* if $\alpha(\gamma_0)$ is Anosov for some $\gamma_0 \in \Gamma$; see Definition 4.3. We state the following conjecture which is motivated in part by the works of Feres–Labourie [23] and Goetze–Spatzier [33].

Conjecture 4.5 ([24, CONJECTURE 1.3]). *If Γ is a lattice in $\mathrm{SL}(n, \mathbb{R})$ where $n \geq 3$, then any C^∞, volume-preserving, Anosov action by Γ on a compact manifold is smoothly conjugate to an action by affine automorphisms of an infranilmanifold.*

See also [35, CONJECTURE 1.1] and [40, CONJECTURE 1.1] for related conjectures. The assumption that the action preserves a volume is standard though results discussed below suggest that such a hypothesis may be unnecessary. Most progress on this conjecture requires additional strong dynamical hypotheses on the action, low dimensionality of the manifold, or assumptions on the topology of the underlying manifold.

We note that affine Anosov actions of higher-rank lattices are known to be local rigid by the work of A. Katok and R. Spatzier [44], extending many earlier results including [34,39,58]. Several partial results towards Conjecture 4.5 appear in [23,33,34,40,41,51,57]. In [9], a new topological and smooth classification of higher-rank lattice actions on tori and nilmanifolds was established. A novelty of the approach in [9] is that no invariant measure is assumed unlike many prior global rigidity results including those in [27,41,51]. For simplicity, we state the following result for actions on tori though versions on nilmanifolds also hold.

Theorem 4.6 ([9, THEOREM 1.3]). *Let Γ be a lattice in* $\mathrm{SL}(n,\mathbb{R})$ *for* $n \geq 3$. *Let* $\alpha\colon \Gamma \to \mathrm{Homeo}(\mathbb{T}^d)$ *be an action by homeomorphisms with linear data* $\rho\colon \Gamma \to \mathrm{GL}(d,\mathbb{Z})$. *Suppose*

(1) *the matrix $\rho(\gamma_0)$ is hyperbolic for some $\gamma_0 \in \Gamma$, and*

(2) *for some finite-index subgroup $\Gamma' \subset \Gamma$, the action $\alpha\colon \Gamma' \to \mathrm{Homeo}(\mathbb{T}^d)$ lifts to an action $\tilde{\alpha}\colon \Gamma' \to \mathrm{Homeo}(\mathbb{R}^d)$.*

Then there is a continuous, surjective $h\colon \mathbb{T}^d \to \mathbb{T}^d$ such that

$$h \circ \alpha(\gamma) = \rho(\gamma) \circ h \tag{4.1}$$

for all γ in a finite-index subgroup $\Gamma'' \subset \Gamma$. In particular, the action $\alpha\colon \Gamma \to \mathrm{Homeo}(\mathbb{T}^d)$ is semiconjugate to an action by affine maps of $\mathbb{T}^d$.

Sufficient conditions for the lifting hypothesis (2) are known; see [9, REMARK 1.5] and references therein. In particular, this automatically holds if $\Gamma = \mathrm{SL}(d,\mathbb{Z})$ acts on $\mathbb{T}^d$ for $d \geq 5$, Γ is cocompact, or α preserves a probability measure μ.

Assuming that $\alpha(\gamma_0)$ is Anosov for some $\gamma_0 \in \Gamma$, Theorem 4.4 implies the map h in Theorem 4.6 is a homeomorphism. For actions by higher-rank lattices, the map h is, in fact, smooth, thus classifying all Anosov actions on tori up to smooth coordinate change.

Theorem 4.7 ([9, THEOREM 1.7]). *Let Γ be a lattice in* $\mathrm{SL}(n,\mathbb{R})$ *for* $n \geq 3$. *Let* $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(\mathbb{T}^d)$ *be an action with linear data* $\rho\colon \Gamma \to \mathrm{GL}(d,\mathbb{Z})$. *Suppose that*

(1) *the diffeomorphism $\alpha(\gamma_0)$ is Anosov for some $\gamma_0 \in \Gamma$, and*

(2) *for some finite-index subgroup $\Gamma' \subset \Gamma$, the action $\alpha\colon \Gamma' \to \mathrm{Diff}^\infty(\mathbb{T}^d)$ lifts to an action $\tilde{\alpha}\colon \Gamma' \to \mathrm{Diff}^\infty(\mathbb{R}^d)$.*

Then, there is a C^∞ diffeomorphism $h\colon \mathbb{T}^d \to \mathbb{T}^d$ such that

$$h \circ \alpha(\gamma) = \rho(\gamma) \circ h$$

for all γ in a finite-index subgroup $\Gamma'' \subset \Gamma$. In particular, the action $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(\mathbb{T}^d)$ is smoothly conjugate to an action by affine maps of $\mathbb{T}^d$.

Again, similar results hold for lattices in other higher-rank simple Lie groups and for Anosov actions on nilmanifolds. To establish Theorem 4.7, we need only show the homeomorphism h in (4.1) given by Theorem 4.6 and Theorem 4.4 is C^∞. Roughly, this follows by studying the restriction of the action α to a higher-rank abelian subgroup $\Sigma \subset \Gamma$. For Anosov actions of higher-rank abelian groups, the map intertwining the action with the linear data is often smooth as shown in [25,26] with the most general result obtained in [60]. The main work to establish Theorem 4.7 is to find $\gamma \in \Gamma$ (which may be different from γ_0) with sufficiently large centralizer in Γ and for which $\alpha(\gamma)$ is Anosov.

Returning to the setting of Theorem 4.6, we might ask if it is possible to classify all (non-Anosov) C^∞ actions on tori with hyperbolic linear data; it seems plausible that all such actions are obtained by a blow-up or slow-down procedure of affine Anosov actions. This suggests the following.

Problem 4.8. Classify all C^∞ actions satisfying the hypotheses of Theorem 4.6.

Specifically, the following may give a possible approach to Problem 4.8.

Question 4.9. Let $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(\mathbb{T}^d)$ be an action satisfying the hypotheses of Theorem 4.6. Is there an α-invariant open set $U \subset \mathbb{T}^d$ such that for the map h satisfying (4.1), $h(U)$ is dense, $h\restriction_U$ is injective, and $h\restriction_U$ is smooth?

In the proof of Theorem 4.7, one shows that every Anosov action of a higher-rank lattice Γ on $\mathbb{T}^d$ preserves a volume form. It is natural to ask if the same holds for the actions as in Theorem 4.6 and ask if a weaker version of Question 4.9 holds. We note that this holds for $\mathrm{SL}(n,\mathbb{Z})$ acting on $\mathbb{T}^n$ by discussion and references in [9, THEOREM 1.6] and [42].

Question 4.10. Let $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(\mathbb{T}^d)$ be an action satisfying the hypotheses of Theorem 4.6. Does α preserve an absolutely continuous probability measure μ on $\mathbb{T}^d$? If so, is there a set $U \subset \mathbb{T}^d$ of full μ-measure such that for h satisfying (4.1), $h(U)$ has full Lebesgue measure, $h\restriction_U$ is injective, and h is smooth along Pesin unstable manifolds?

4.4. Anosov actions in dimension *n*

We return to the motiving problem of classifying actions of $\mathrm{SL}(n,\mathbb{Z})$ on n-manifolds. In [41], volume-preserving Anosov actions of $\mathrm{SL}(n,\mathbb{Z})$ on n-tori were shown to be smoothly conjugate to affine actions for $n \geq 3$. Recently, H. Lee considered the same problem but without any assumption on the topology of the underlying manifold.

Theorem 4.11 ([47, THEOREM 1.5]). *For $n \geq 3$, let Γ be a lattice in $\mathrm{SL}(n,\mathbb{R})$. Suppose $\dim(M) = n$ and let $\alpha\colon \Gamma \to \mathrm{Diff}^1_{\mathrm{vol}}(M)$ be an action such that $\alpha(\gamma_0)$ is Anosov for some $\gamma_0 \in \Gamma$. Then there is a homeomorphism $h\colon M \to \mathbb{T}^n$ such that $h \circ \alpha(\gamma) \circ h^{-1}$ is affine for every $\gamma \in \Gamma$. Moreover, if $\alpha\colon \Gamma \to \mathrm{Diff}^\infty_{\mathrm{vol}}(M)$ then h is C^∞.*

It is natural to ask if the assumption that the action preserves a volume form in Theorem 4.11 can be removed.

Conjecture 4.12. *For $n \geq 3$, let $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ be a lattice. Let $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ be an action such that $\alpha(\gamma_0)$ is Anosov for some $\gamma_0 \in \Gamma$. Then Γ preserves a smooth (nowhere vanishing) volume form on M.*

In a recent collaboration, we were able to verify this conjecture in certain situations.

Theorem 4.13 (A. Brown and H. Lee; in preparation). *Conjecture 4.12 holds for C^∞ Anosov actions on n-manifolds by cocompact lattices in $\mathrm{SL}(n,\mathbb{R})$ for $n \geq 4$.*

We also expect that Theorem 4.13 holds for nonuniform lattices. In Theorem 4.11, the only possible lattice subgroups $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ admitting Anosov actions on $\mathbb{T}^n$ are (up to conjugacy) commensurable with $\mathrm{SL}(n,\mathbb{Z})$. Combined with Theorem 4.11, this would imply that the only lattices $\Gamma \subset \mathrm{SL}(n, \mathbb{R})$ that admit a C^∞ Anosov action in dimension n are commensurable with $\mathrm{SL}(n, \mathbb{Z})$.

5. TOOLS USED IN PROOFS

5.1. Suspension space and induced G-action

Let G be a Lie group and let Γ be a lattice subgroup of G. Let M be a compact manifold and let $\alpha\colon \Gamma \to \mathrm{Diff}(M)$ be an action. A well-known construction translates between the Γ-action on M and an equivariant G-action on a fiber-bundle X over G/Γ with fibers diffeomorphic to M: On $G \times M$ consider the right Γ-action and the left G-action:

$$(g, x) \cdot \gamma = \big(g\gamma, \alpha(\gamma^{-1})(x)\big), \quad a \cdot (g, x) = (ag, x).$$

Define the quotient manifold $X := (G \times M)/\Gamma$. The G-action on $G \times M$ descends to a G-action on X. For $g \in G$ and $x \in X$, denote the action by $g \cdot x$ and denote the derivative of the diffeomorphism $x \mapsto g \cdot x$ at $x \in X$ by $D_x g\colon T_x X \to T_{g \cdot x} X$.

The space X is a fiber bundle over G/Γ. Let $\pi\colon X \to G/\Gamma$ be the projection and let $\mathcal{E} := \ker D\pi$ denote the *fiberwise tangent bundle*. That is, $\mathcal{E} = (G \times TM)/\Gamma$.

We write $\mathcal{A}\colon G \times \mathcal{E} \to \mathcal{E}$ for the *fiberwise derivative* cocycle over the G-action X: given $x \in X$, if $\mathcal{E}(x)$ is the fiber of $\mathcal{E}$ over x then $\mathcal{A}(g, x)\colon \mathcal{E}(x) \to \mathcal{E}(g \cdot x)$ is the restriction to $\mathcal{E}(x)$ of the derivative of translation by g:

$$\mathcal{A}(g, x) = D_x g \restriction_{\mathcal{E}(x)}.$$

When Γ is cocompact, we equip TX and $\mathcal{E}$ with any choice of Riemannian metric. When Γ is nonuniform, we use arithmeticity of Γ and Siegel domains in G to equip TX and $\mathcal{E}$ with Riemannian metrics adapted to the geometry of Γ in G.

5.2. Common themes

Let $A = \exp \mathfrak{a}$ be a maximal $\mathbb{R}$-split Cartan subgroup of G; when $G = \mathrm{SL}(n, \mathbb{R})$, we take $A = \{\mathrm{diag}(e^{t_1}, \dots, e^{t_n}) : t_1 + \cdots + t_n = 0\}$, the subgroup of positive diagonal matrices whence $A \simeq \mathbb{R}^{n-1}$. Most results discussed above follow from precise formulations of the following 2 heuristics. In the remainder of this section, we discuss concrete examples of these. Throughout, we always assume $\dim(A) \geq 2$.

Theme 1. *Dynamical properties of the Γ-action on M induce A-invariant probability measures μ on X factoring onto the Haar measure on G/Γ with corresponding dynamical properties.*

Theme 2. *A-invariant probability measures μ on X factoring onto the Haar measure on G/Γ are expected to be very "rigid."*

Theme 2 often leads to extra invariance or homogeneity of the measure μ. Combined with dynamical structures associated with μ in Theme 1, this often constrains possible dynamical properties of Γ on M or reveals some homogeneous structures associated with the Γ-action on M.

Below we describe one instance of Theme 1 and several instances of Theme 2. We also outline cohomological versions of Theme 1 and Theme 2 that are used in the proof of Theorem 4.6.

5.3. Theme 1 and subexponential growth

Let $X = (G \times M)/\Gamma$ denote the induced G-space and let $\mathcal{A}$ denote the corresponding fiberwise derivative cocycle. Given $a \in G$ and an a-invariant probability measure μ on X, we define the *average top Lyapunov exponent of* $\mathcal{A}$ by

$$\lambda_{\mathrm{top},a,\mu,\mathcal{A}} := \liminf_{n\to\infty} \frac{1}{n} \int \log \|\mathcal{A}(a^n, x)\| \, d\mu(x).$$

This is finite whenever the function $x \mapsto \log \|\mathcal{A}(a, x)\|$ is $L^1(\mu)$ which—by the choice of norm on $\mathcal{E}$—holds for any probability measure μ on X that factors onto the normalized Haar measure on G/Γ. We recall Definition 2.5. The main technical theorem established in the papers [5–7] is the following precise version of Theme 1 which, under the assumption that the conclusion of Theorem 2.6 fails, builds a measure on the suspension space X with certain dynamical properties.

Theorem 5.1 ([7, THEOREM D]). *Let G be a connected semisimple Lie group with finite center,*[2] *without compact factors, and with* $\operatorname{rank}_{\mathbb{R}} G \geq 2$. *Let Γ be an irreducible lattice subgroup in G, let M be a compact manifold, and let $\alpha\colon \Gamma \to \mathrm{Diff}^1(M)$ be an action. If the action α fails to have uniform subexponential growth of derivatives then there exists a maximal $\mathbb{R}$-split Cartan subgroup A of G and a probability measure μ on X such that*

(1) *μ is A-invariant,*

(2) *μ projects to the Haar measure on G/Γ, and*

(3) *for some $a \in A$, the average top Lyapunov exponent $\lambda_{\mathrm{top},a,\mu,\mathcal{A}}$ is positive.*

We remark that there are no constraints on the dimension of M in the statement of Theorem 5.1. In particular, Theorem 5.1 serves as the starting point for the proofs of Theorems 2.4 and 2.6 as well as Theorems 2.11, 2.13, and 3.1 and may serve as a starting point for future results.

5.4. Theme 2 and invariance of measures

In Theme 2, we consider an A-invariant probability measure μ on X factoring onto the Haar measure on G/Γ. One precise version of Theme 2 produces extra invariance of μ by certain subgroups of G normalized by A. See especially [8, PROPOSITION 5.1]. This has the following corollary used to prove Theorem 2.6.

Theorem 5.2. *Let G be a higher-rank simple Lie group, let Γ be a lattice in G, let M be a compact manifold, and let $\alpha\colon \Gamma \to \mathrm{Diff}^r(M)$ be an action for $r > 1$. Then*

[2] For simplicity of statement, we assume the center of G is finite though that is not necessary for applications.

(1) *if* $\dim(M) \leq r(G) - 1$, *every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant;*

(2) *if* $\dim(M) \leq r(G)$ *and* α *is volume-preserving, every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant.*

To prove Theorem 2.6, if $\dim M < n(G)$ and if μ is a G-invariant probability measure on X, then Zimmer's cocycle superrigity implies $\lambda_{\mathrm{top},a,\mu,\mathcal{A}} = 0$ for every $a \in G$. Combined with Theorems 5.1 and 5.2, we obtain a contradiction unless the conclusion of Theorem 2.6 holds.

In the setting of C^1 actions, we have the following weaker version of Theorem 5.2 which follows from mild modifications of the invariance principle in [1] (extending results of [46]).

Theorem 5.3 ([4, PROPOSITION 3]). *Let* G *be a higher-rank simple Lie group, let* Γ *be a lattice in* G*, let* M *be a compact manifold, and let* $\alpha\colon \Gamma \to \mathrm{Diff}^1(M)$ *be an action. Then*

(1) *if* $\dim(M) \leq \mathrm{rank}(G) - 1$, *every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant;*

(2) *if* $\dim(M) \leq \mathrm{rank}(G)$ *and* α *is volume-preserving, every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant.*

The appearance of $r(G)$ (rather than $v(G)$ or $n(G)$) in Theorem 5.2 is the main reason why $r(G)$ appears in Theorem 2.4. However, one might expect an analogue of Theorem 5.2 with dimension bounds corresponding to those in Conjecture 2.2 holds.

Conjecture 5.4. *Let* G *be a higher-rank simple Lie group, let* Γ *be a lattice in* G*, let* M *be a compact manifold, and let* $\alpha\colon \Gamma \to \mathrm{Diff}^\infty(M)$ *be an action.*

(1) *If* $\dim(M) \leq v(G) - 1$, *every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant.*

(2) *If* $\dim(M) \leq v(G)$ *and if* α *is volume-preserving, every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant.*

One might further conjecture the following.

(3) *If* $\dim(M) \leq n(G) - 1$ *and if* α *is volume-preserving, every* A*-invariant probability measure on* X *that projects to the Haar measure on* G/Γ *is* G*-invariant.*

For many non-$\mathbb{R}$-split groups G, (1) and (2) of Conjecture 5.4 can be established using tools of measure rigidity and cocycle superrigidity. We discuss this in the next section.

5.5. Theme 2: measure and cocycle rigidity; homogeneous structures

Let A be a maximal (connected) $\mathbb{R}$-split Cartan subgroup of G; since $A \simeq \mathbb{R}^{\mathrm{rank}(G)}$, A is a higher-rank abelian group if G is higher-rank. Measures invariant under higher-rank abelian groups (with positive entropy) are expected to exhibit some degree of homogeneity unless they factor onto an action of a rank-1 quotient of A. Such results have been established in the setting of homogenous dynamics, see especially [19–21,43,48,61], and in the setting of smooth non-linear dynamics, see especially [36,37].

To prove Theorem 2.11, it remains to establish relevant cases of Conjecture 5.4; this implies an analogous version of Theorem 2.6 and allows one to complete the outline in Section 2.2. An argument discovered by J. An shows that one may assume the A-invariant measure μ in the conclusion of Theorem 5.1 is invariant under a parabolic subgroup $Q \subset G$ containing A. When the Levi component of Q is sufficiently large, Zimmer's cocycle superrigidity constrains the combinatorics of the Lyapunov spectrum of the cocycle $\mathcal{A}$ (over the action of A and the measure μ); this, combined with [8, PROPOSITION 5.1], yields many cases of Conjecture 5.4 including (among others) the ranges in Theorem 2.11. Adaptations of the nonlinear measure rigidity arguments in [36,37] yield further constraints on the combinatorics of the Lyapunov spectrum of $\mathcal{A}$ solving additional cases of Conjecture 5.4. We summarize with following.

Theorem 5.5 (J. An, A. Brown, and Z. Zhang; in preparation).

(1) *Conjecture 5.4(1) holds for lattices in complex simple Lie groups.*

(2) *Conjecture 5.4(2) holds for lattices in* $\mathrm{SL}(n,\mathbb{C})$ *for* $n \geq 4$.

(3) *Conjecture 5.4(1) holds for lattices in the groups appearing in Theorem 2.11.*

While this establishes Conjecture 5.4 in many cases, there are many higher-rank simple groups for which Conjecture 5.4 is unresolved.

Problem 5.6. Find a new mechanism to obtain extra invariance of A-invariant measures that allows us to establish additional cases of Conjecture 5.4.

The result announced in Theorem 4.13 follows by similarly adapting the measure rigidity arguments of [36] as well as a version of Theme 1 involving topological entropy.

The proof of Theorem 3.1 follows from further adapting the techniques of measure rigidity. Very roughly, starting from the A-invariant measure μ in the conclusion of Theorem 5.1, we have that μ is invariant under a parabolic subgroup $Q \subset G$ containing A. Locally, every point $x \in X$ has a neighborhood parameterized as $U \times M$ where $U \subset G$ is an open neighborhood of the identity. If $V \subset G$ denotes the unipotent subgroup transverse to Q, one shows the restriction of the measure μ to such parameterized neighborhoods coincides with the graph of an injective, C^r function $V \to M$. These graphs then assemble coherently to give local homogeneous coordinates relative to which M admits the structure of a Γ-equivariant covering space of G/Q.

5.6. Cohomological versions of Theme 1 and Theme 2

We end this note with a reformulation of Theme 1 and Theme 2 used in the proof of Theorem 4.6. Let $\Gamma \subset \mathrm{SL}(n,\mathbb{R})$ be a lattice, let $\alpha\colon \Gamma \to \mathrm{Homeo}(\mathbb{T}^d)$ be as in Theorem 4.6, and let $\rho\colon \Gamma \to \mathrm{GL}(d,\mathbb{Z})$ be the linear data of α. Passing to compact extensions and subgroups of finite index, assume α lifts to an action $\tilde{\alpha}\colon \Gamma \to \mathrm{Homeo}(\mathbb{R}^d)$ and that ρ coincides with the restriction to Γ of a continuous representation $\rho\colon \mathrm{SL}(n,\mathbb{R}) \to \mathrm{SL}(d,\mathbb{R})$.

Cohomological reformulation. Consider the identify map $h_0\colon \mathbb{T}^d \to \mathbb{T}^d$; the defect of h_0 satisfying (4.1) determines a continuous, ρ-twisted 1-cocycle $c\colon \Gamma \times \mathbb{T}^d \to \mathbb{R}^d$, given by $c(\gamma, x) = \tilde{\alpha}(\gamma)(\tilde{x}) - \rho(\gamma)\tilde{x}$ for any lift $\tilde{x} \in \mathbb{R}^d$ of x. In particular, for $\gamma_1, \gamma_2 \in \Gamma$,

$$c(\gamma_1\gamma_2, x) = \rho(\gamma_1)c(\gamma_2, x) + c\big(\gamma_1, \alpha(\gamma_2)(x)\big). \tag{5.1}$$

Suppose that c is a coboundary; that is, suppose there exists a continuous function $\eta\colon \mathbb{T}^d \to \mathbb{R}^d$ such that for every $\gamma \in \Gamma$ and $x \in \mathbb{T}^d$, $c(\gamma, x) = \rho(\gamma)\eta(x) - \eta(\alpha(\gamma)(x))$. The function $h(x) = h_0(x) + \eta(x) = x + \eta(x)$ then satisfies (4.1).

Cohomological version of Theme 1. Rather than study the Γ-action on $\mathbb{T}^d$, we pass to the G-action on the suspension space X and define a related cocycle $\tilde{c}\colon G \times X \to \mathbb{R}^d$. While $\tilde{c}$ is continuous in every fiber of X, it is only Borel measurable over G/Γ. Nonetheless, to establish Theorem 4.6, it suffices to show $\tilde{c}$ is a coboundary: for every $g \in G$,

$$\tilde{c}(g, x) = \rho(g)\tilde{\eta}(x) - \tilde{\eta}(g \cdot x) \tag{5.2}$$

where $\tilde{\eta}\colon X \to \mathbb{R}^d$ is a measurable function that is continuous in Haar-almost every fiber. Using that $\rho(\gamma_0)$ is hyperbolic for some $\gamma_0 \in \Gamma$, a modification of the proof of Theorem 4.2 produces such a function $\tilde{\eta}\colon X \to \mathbb{R}^d$ such that (5.2) holds for all $g \in A$ and almost every fiber.

Cohomological version of Theme 2. It remains to show the function $\tilde{\eta}$ solves equation (5.2) for all $g \in G$. We identify finitely many unipotent subgroups $\mathcal{U} = \{U_1, \dots, U_\ell\}$ (specifically, 1-parameter root subgroups) of $\mathrm{SL}(n,\mathbb{R})$, each of which is normalized by A, and show

(1) the cocycle equation (5.2) holds for all $g \in U_j$, and

(2) the group G is generated by the subgroups $\mathcal{U} = \{U_1, \dots, U_\ell\}$ and A.

For the case of $G = \mathrm{SL}(n,\mathbb{R})$ we may take $\mathcal{U}$ to contain all root subgroups normalized by A. However, for certain higher-rank simple groups G (such as $\mathrm{Sp}(4,\mathbb{R})$ of real rank 2), it may be that (5.2) only holds for g in a subset of the root groups; nonetheless, these groups still generate all of G.

The above outline should apply to any ρ-twisted cocycle $c\colon \Gamma \times \mathbb{T}^d \to \mathbb{R}^k$, assuming $\rho(\gamma_0)$ is hyperbolic for some $\gamma_0 \in \Gamma$, and show c is a coboundary. However, this is a large restriction on the class of representations ρ considered; for instance, it does not include the case that ρ is the adjoint representation. Still, using that c is a cocycle for an action of a large group, it may be possible to solve the following.

Question 5.7. Let $c\colon \Gamma \times \mathbb{T}^d \to \mathbb{R}^k$ be a ρ-twisted cocycle where $\rho\colon G \to \mathrm{GL}(k,\mathbb{R})$ is a nontrivial irreducible representation (such as the adjoint). Is c a coboundary?

A. NUMEROLOGY ASSOCIATED WITH ZIMMER'S CONJECTURE

We compute the numbers $n(G)$, $d(G)$, $v(G)$, and $r(G)$ for various classical real Lie groups. These numbers depend only on the Lie algebra of G.

Lie algebra $\mathfrak{g}$	restricted root system	real rank	$n(G)$	$d(G)$	$v(G)$	$r(G)$
$\mathfrak{sl}(n,\mathbb{R})$ $n \geq 2$	A_{n-1}	$n-1$	n	$2n-2, n \neq 4$ $5, n = 4$[(a)]	$n-1$	$n-1$
$\mathfrak{sp}(2n,\mathbb{R})$ $n \geq 2$	C_n	n	$2n$	$4n-4$	$2n-1$	$2n-1$
$\mathfrak{so}(n,n+1)$ $n \geq 3$[(b)]	B_n	n	$2n+1$	$2n$	$2n-1$	$2n-1$
$\mathfrak{so}(n,n)$ $n \geq 4$[(c)]	D_n	n	$2n$	$2n-1$	$2n-2$	$2n-2$

[(a)] $\mathfrak{sl}(4,\mathbb{R}) = \mathfrak{so}(3,3)$

[(b)] $\mathfrak{so}(1,2) = \mathfrak{sl}(2,\mathbb{R})$ and $\mathfrak{so}(2,3) = \mathfrak{sp}(4,\mathbb{R})$

[(c)] $\mathfrak{so}(2,2)$ is not simple and $\mathfrak{so}(3,3) = \mathfrak{sl}(4,\mathbb{R})$

TABLE 1

Numerology appearing in Zimmer's conjecture for classical $\mathbb{R}$-split Lie algebras.

Lie algebra $\mathfrak{g}$	restricted root system	real rank	$n(G)$	$d(G)$	$v(G)$	$r(G)$
$\mathfrak{sl}(n,\mathbb{C})$ $n \geq 2$	A_{n-1}	$n-1$	$2n$	$2n-2, n \neq 4$ $5, n = 4$[(d)]	$2n-2$	$n-1$
$\mathfrak{sp}(2n,\mathbb{C})$ $n \geq 2$	C_n	n	$4n$	$4n-4$	$4n-2$	$2n-1$
$\mathfrak{so}(2n+1,\mathbb{C})$ $n \geq 3$[(e)]	B_n	n	$4n+2$	$2n$	$4n-2$	$2n-1$
$\mathfrak{so}(2n,\mathbb{C})$ $n \geq 4$[(f)]	D_n	n	$4n$	$2n-1$	$4n-4$	$2n-2$

[(d)] $\mathfrak{sl}(4,\mathbb{C}) = \mathfrak{so}(6,\mathbb{C})$

[(e)] $\mathfrak{so}(5,\mathbb{C}) = \mathfrak{sp}(4,\mathbb{C})$ and $\mathfrak{so}(3,\mathbb{C}) = \mathfrak{sl}(2,\mathbb{C})$.

[(f)] $\mathfrak{so}(6,\mathbb{C}) = \mathfrak{sl}(4,\mathbb{C})$ and $\mathfrak{so}(4,\mathbb{C})$ is not simple.

TABLE 2

Numerology appearing appearing in Zimmer's conjecture for classical complex Lie algebras.

Lie algebra $\mathfrak{g}$	restricted root system	real rank	$n(G)$	$d(G)$	$v(G)$	$r(G)$
$\mathfrak{sl}(n,\mathbb{H}), n \geq 3$	A_{n-1}	$n-1$	$4n$	$4n-2$	$4n-4$	$n-1$
$\mathfrak{so}(n,m)$ $2 \leq n \leq n+2 \leq m$	$B_n, n < m$	n	$n+m$	$n+m-1$	$n+m-2$	$2n-1$
$\mathfrak{su}(n,m)$ $2 \leq n \leq m$ $(n,m) \neq (2,2)^{(g)}$	$(BC)_n, n < m$ $C_n, n = m$	n	$2n+2m$	$2n+2m-2$	$2n+2m-3$	$2n-1$
$\mathfrak{sp}(2n,2m)$ $1 \leq n \leq m$	$(BC)_n, n < m$ $C_n, n = m$	n	$4n+4m$	$4n+4m-4$	$4n+4m-5$	$2n-1$
$\mathfrak{so}^*(2n)$ $n \geq 4$ even$^{(h)}$	$C_{\frac{1}{2}n}$	$\frac{n}{2}$	$4n$	$2n-1$	$4n-7$	$n-1$
$\mathfrak{so}^*(2n)$ $n \geq 5$ odd	$(BC)_{\frac{1}{2}(n-1)}$	$\frac{n-1}{2}$	$4n$	$2n-1$	$4n-7$	$n-2$

$^{(g)}$ $\mathfrak{su}(2,2) = \mathfrak{so}(4,2)$

$^{(h)}$ $\mathfrak{so}^*(4)$ is not simple

TABLE 3
Numerology appearing in Zimmer's conjecture for classical higher-rank nonsplit real forms.

ACKNOWLEDGMENTS

The author would like to thank all his collaborators and coauthors who contributed to the projects described in this note.

FUNDING

This work was partially supported by National Science Foundation grant DMS-1752675.

REFERENCES

[1] A. Avila and M. Viana, Extremal Lyapunov exponents: an invariance principle and applications. *Invent. Math.* **181** (2010), 115–189.

[2] M. R. Bridson, F. Grunewald, and K. Vogtmann, Actions of arithmetic groups on homology spheres and acyclic homology manifolds. *Math. Z.* **276** (2014), 387–395.

[3] A. Brown, Entropy, Lyapunov exponents, and rigidity of group actions. *Ensaois Matematicos* **33** (2019), 1–197. With appendices by D. Malicet, D. Obata, B. Santiago, M. Triestino, S. Alvarez, and M. Roldán.

[4] A. Brown, D. Damjanovic, and Z. Zhang, C^1 actions on manifolds by lattices in Lie groups. 2018, arXiv:1801.04009, to appear.

[5] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture: Subexponential growth, measure rigidity, and strong property (T). 2016, arXiv:1608.04995.

[6] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture for actions of $\mathrm{SL}(m, \mathbb{Z})$. *Invent. Math.* **221** (2020), 1001–1060.

[7] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture for non-uniform lattices and escape of mass. 2020, arXiv:2105.14541.

[8] A. Brown, F. Rodriguez Hertz, and Z. Wang, Invariant measures and measurable projective factors for actions of higher-rank lattices on manifolds. 2016, arXiv:1609.05565.

[9] A. Brown, F. Rodriguez Hertz, and Z. Wang, Global smooth and topological rigidity of hyperbolic lattice actions. *Ann. of Math. (2)* **186** (2017), 913–972.

[10] M. Burger and N. Monod, Continuous bounded cohomology and applications to rigidity theory. *Geom. Funct. Anal.* **12** (2002), 219–280.

[11] S. Cantat, Version kählérienne d'une conjecture de Robert J. Zimmer. *Ann. Sci. Éc. Norm. Supér. (4)* **37** (2004), 759–768.

[12] S. Cantat, Progrès récents concernant le programme de Zimmer [d'après A. Brown, D. Fisher et S. Hurtado]. In *Séminaire Bourbaki, Vol. 2017/2018, exposés 1136–1150*, 414, pp. 1–48, Exp. No. 1136, 2019.

[13] S. Cantat and J. Xie, Algebraic actions of discrete groups: the p-adic method. *Acta Math.* **220** (2018), 239–295.

[14] S. Cantat and A. Zeghib, Holomorphic actions, Kummer examples, and Zimmer program. *Ann. Sci. Éc. Norm. Supér. (4)* **45** (2012), 447–489.

[15] M. de la Salle, Strong property (T) for higher-rank lattices. *Acta Math.* **223** (2019), 151–193.

[16] T. de Laat and M. de la Salle, Strong property (T) for higher-rank simple Lie groups. *Proc. Lond. Math. Soc. (3)* **111** (2015), 936–966.

[17] B. Deroin and S. Hurtado, Non left-orderability of lattices in higher rank semi-simple lie groups. 2020, arXiv:2008.10687.

[18] B. Deroin, V. Kleptsyn, A. Navas, and K. Parwani, Symmetric random walks on $\mathrm{Homeo}^{+}(\mathbf{R})$. *Ann. Probab.* **41** (2013), 2066–2089.

[19] M. Einsiedler and A. Katok, Rigidity of measures—the high entropy case and non-commuting foliations. *Israel J. Math.* **148** (2005), 169–238.

[20] M. Einsiedler, A. Katok, and E. Lindenstrauss, Invariant measures and the set of exceptions to Littlewood's conjecture. *Ann. of Math. (2)* **164** (2006), 513–560.

[21] M. Einsiedler and E. Lindenstrauss, Rigidity properties of $\mathbb{Z}^d$-actions on tori and solenoids. *Electron. Res. Announc. Am. Math. Soc.* **9** (2003), 99–110 (electronic).

[22] B. Farb and P. Shalen, Real-analytic actions of lattices. *Invent. Math.* **135** (1999), 273–296.

[23] R. Feres and F. Labourie, Topological superrigidity and Anosov actions of lattices. *Ann. Sci. Éc. Norm. Supér. (4)* **31** (1998), 599–629.

[24] D. Fisher, Groups acting on manifolds: around the Zimmer program. In *Geometry, rigidity, and group actions*, pp. 72–157, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 2011.

[25] D. Fisher, B. Kalinin, and R. Spatzier, Totally nonsymplectic Anosov actions on tori and nilmanifolds. *Geom. Topol.* **15** (2011), 191–216.

[26] D. Fisher, B. Kalinin, and R. Spatzier, Global rigidity of higher rank Anosov actions on tori and nilmanifolds. *J. Amer. Math. Soc.* **26** (2013), 167–198. With an appendix by James F. Davis.

[27] D. Fisher and K. Whyte, Continuous quotients for lattice actions on compact spaces. *Geom. Dedicata* **87** (2001), 181–189.

[28] J. Franks, Anosov diffeomorphisms. In *Global Analysis (Proc. Sympos. Pure Math., Vol. XIV, Berkeley, Calif., 1968)*, pp. 61–93, Amer. Math. Soc., Providence, R.I., 1970.

[29] J. Franks and M. Handel, Area preserving group actions on surfaces. *Geom. Topol.* **7** (2003), 757–771.

[30] J. Franks and M. Handel, Distortion elements in group actions on surfaces. *Duke Math. J.* **131** (2006), 441–468.

[31] E. Ghys, Sur les groupes engendrés par des difféomorphismes proches de l'identité. *Bol. Soc. Brasil. Mat. (N.S.)* **24** (1993), 137–178.

[32] É. Ghys, Actions de réseaux sur le cercle. *Invent. Math.* **137** (1999), 199–231.

[33] E. R. Goetze and R. J. Spatzier, Smooth classification of Cartan actions of higher rank semisimple Lie groups and their lattices. *Ann. of Math. (2)* **150** (1999), 743–773.

[34] S. Hurder, Rigidity for Anosov actions of higher rank lattices. *Ann. of Math. (2)* **135** (1992), 361–410.

[35] S. Hurder, A survey of rigidity theory for Anosov actions. In *Differential topology, foliations, and group actions (Rio de Janeiro, 1992)*, pp. 143–173, Contemp. Math. 161, Amer. Math. Soc., Providence, RI, 1994.

[36] B. Kalinin and A. Katok, Measure rigidity beyond uniform hyperbolicity: invariant measures for Cartan actions on tori. *J. Mod. Dyn.* **1** (2007), 123–146.

[37] B. Kalinin, A. Katok, and F. Rodriguez Hertz, Nonuniform measure rigidity. *Ann. of Math. (2)* **174** (2011), 361–400.

[38] M. Kanai, A new approach to the rigidity of discrete group actions. *Geom. Funct. Anal.* **6** (1996), 943–1056.

[39] A. Katok and J. Lewis, Local rigidity for certain groups of toral automorphisms. *Israel J. Math.* **75** (1991), 203–241.

[40] A. Katok and J. Lewis, Global rigidity results for lattice actions on tori and new examples of volume-preserving actions. *Israel J. Math.* **93** (1996), 253–280.

[41] A. Katok, J. Lewis, and R. Zimmer, Cocycle superrigidity and rigidity for lattice actions on tori. *Topology* **35** (1996), 27–38.

[42] A. Katok and F. Rodriguez Hertz, Arithmeticity and topology of smooth actions of higher rank abelian groups. *J. Mod. Dyn.* **10** (2016), 135–172.

[43] A. Katok and R. J. Spatzier, Invariant measures for higher-rank hyperbolic abelian actions. *Ergodic Theory Dynam. Systems* **16** (1996), 751–778.

[44] A. Katok and R. J. Spatzier, Differential rigidity of Anosov actions of higher rank abelian groups and algebraic lattice actions. *Tr. Mat. Inst. Steklova* **216** (1997), 292–319.

[45] V. Lafforgue, Un renforcement de la propriété (T). *Duke Math. J.* **143** (2008), 559–602.

[46] F. Ledrappier, Positivity of the exponent for stationary sequences of matrices. In *Lyapunov exponents (Bremen, 1984)*, pp. 56–73, Lecture Notes in Math. 1186, Springer, Berlin, 1986.

[47] H. Lee, Global rigidity of actions by higher rank lattices with dominated splitting. 2020, arXiv:2010.11874.

[48] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)* **163** (2006), 165–219.

[49] A. Manning, There are no new Anosov diffeomorphisms on tori. *Amer. J. Math.* **96** (1974), 422–429.

[50] G. A. Margulis, Non-uniform lattices in semisimple algebraic groups. In *Lie groups and their representations (Proc. Summer School on Group Representations of the Bolyai János Math. Soc., Budapest, 1971)*, pp. 371–553, Halsted, New York, 1975.

[51] G. A. Margulis and N. Qian, Rigidity of weakly hyperbolic actions of higher real rank semisimple Lie groups and their lattices. *Ergodic Theory Dynam. Systems* **21** (2001), 121–164.

[52] Y. Matsushima and S. Murakami, On vector bundle valued harmonic forms and automorphic forms on symmetric riemannian manifolds. *Ann. of Math. (2)* **78** (1963), 365–416.

[53] G. D. Mostow, *Strong rigidity of locally symmetric spaces*. Annals of Mathematics Studies 78, Princeton University Press, Princeton, N.J., 1973.

[54] K. Parwani, Actions of $\mathrm{SL}(n,\mathbb{Z})$ on homology spheres. *Geom. Dedicata* **112** (2005), 215–223.

[55] L. Polterovich, Growth of maps, distortion in groups and symplectic geometry. *Invent. Math.* **150** (2002), 655–686.

[56] G. Prasad, Strong rigidity of $\mathbf{Q}$-rank 1 lattices. *Invent. Math.* **21** (1973), 255–286.

[57] N. Qian, Tangential flatness and global rigidity of higher rank lattice actions. *Trans. Amer. Math. Soc.* **349** (1997), 657–673.

[58] N. Qian and C. Yue, Local rigidity of Anosov higher-rank lattice actions. *Ergodic Theory Dynam. Systems* **18** (1998), 687–702.

[59] M. S. Raghunathan, On the first cohomology of discrete subgroups of semisimple Lie groups. *Amer. J. Math.* **87** (1965), 103–139.

[60] F. Rodriguez Hertz and Z. Wang, Global rigidity of higher rank abelian Anosov algebraic actions. *Invent. Math.* **198** (2014), 165–209.

[61] D. J. Rudolph, ×2 and ×3 invariant measures and entropy. *Ergodic Theory Dynam. Systems* **10** (1990), 395–406.

[62] A. Selberg, On discontinuous groups in higher-dimensional symmetric spaces. In *Contributions to function theory (internat. colloq. function theory, bombay, 1960)*, pp. 147–164, Tata Institute of Fundamental Research, Bombay, 1960.

[63] F. Uchida, Classification of real analytic SL(n, $\mathbf{R}$) actions on n-sphere. *Osaka Math. J.* **16** (1979), 561–579.

[64] A. Weil, On discrete subgroups of Lie groups. II. *Ann. of Math. (2)* **75** (1962), 578–602.

[65] A. Weil, Remarks on the cohomology of groups. *Ann. of Math. (2)* **80** (1964), 149–157.

[66] S. Weinberger, SL(n, $\mathbf{Z}$) cannot act on small tori. In *Geometric topology (Athens, GA, 1993)*, pp. 406–408, AMS/IP Stud. Adv. Math. 2, Amer. Math. Soc., Providence, RI, 1997.

[67] S. Weinberger, Some remarks inspired by the C^0 Zimmer program. In *Geometry, rigidity, and group actions*, pp. 262–282, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 2011.

[68] D. Witte, Arithmetic groups of higher $\mathbf{Q}$-rank cannot act on 1-manifolds. *Proc. Amer. Math. Soc.* **122** (1994), 333–340.

[69] D. Witte Morris, *Introduction to arithmetic groups*. Deductive Press, 2015.

[70] S. Ye, The action of matrix groups on aspherical manifolds. *Algebr. Geom. Topol.* **18** (2018), 2875–2895.

[71] Z. Zhang, Zimmer's conjecture for lattice actions: the SL(n, $\mathbb{C}$)-case. 2018, arXiv:1809.06224.

[72] R. J. Zimmer, Strong rigidity for ergodic actions of semisimple Lie groups. *Ann. of Math. (2)* **112** (1980), 511–529.

[73] R. J. Zimmer, Arithmetic groups acting on compact manifolds. *Bull. Amer. Math. Soc. (N.S.)* **8** (1983), 90–92.

[74] R. J. Zimmer, Volume preserving actions of lattices in semisimple groups on compact manifolds. *Publ. Math. Inst. Hautes Études Sci.* (1984), 5–33.

[75] R. J. Zimmer, Actions of semisimple groups and discrete subgroups. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, pp. 1247–1258, Amer. Math. Soc., Providence, RI, 1987.

[76] R. J. Zimmer, Lattices in semisimple groups and invariant geometric structures on compact manifolds. In *Discrete groups in geometry and analysis (New Haven, Conn., 1984)*, pp. 152–210, Progr. Math. 67, Birkhäuser Boston, Boston, MA, 1987.

[77] R. J. Zimmer, Spectrum, entropy, and geometric structures for smooth actions of Kazhdan groups. *Israel J. Math.* **75** (1991), 65–80.

[78] B. P. Zimmermann, SL(n, $\mathbb{Z}$) cannot act on small spheres. *Topology Appl.* **156** (2009), 1167–1169.

AARON BROWN

Northwestern University, Evanston, IL 60208, USA, awb@northwestern.edu

THE HOROCYCLE FLOW ON THE MODULI SPACE OF TRANSLATION SURFACES

JON CHAIKA AND BARAK WEISS

ABSTRACT

We survey some results on the dynamics of the horocycle flow on the moduli space of translation surfaces. We outline proofs of some recent results, obtained by the authors in collaboration with John Smillie, and pose some open questions.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37C85; Secondary 30F30

KEYWORDS

Horocycle flow, translation surfaces, moduli spaces

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3412–3430
DOI 10.4171/ICM2022/89

1. INTRODUCTION

The study of dynamics on moduli spaces of translation surfaces has been undergoing intensive growth over the last two decades. This subdomain of ergodic theory lies at the crossroads of dynamics of Lie group actions and geometry of surfaces and has close connections with the theory of rational billiards, interval exchange transformations, Teichmüller theory, algebraic geometry, number theory, mathematical physics, and more. The foundations of the theory were laid down by Masur and Veech in the 1980s, in work motivated by a conjecture of Keane about interval exchange transformations. Through the efforts of many mathematicians (see the ICM proceedings contributions [11,16,23,30,49], and the survey [33] about billiards in this volume), we now know a great deal about the dynamics on these spaces. As entry points we recommend the surveys [17,24,26,44,46–48].

Our focus in this survey will be results and open questions concerning the dynamics of the horocycle flow. Much of the work on the dynamics on spaces of translation surfaces has been motivated by a fruitful analogy with the study of Lie group actions on homogeneous spaces. In such a putative dictionary, the horocycle flow on moduli spaces corresponds to a unipotent flow on a homogeneous space, for which Ratner [31] famously showed that all orbit-closures and invariant measures admit a nice algebraic description. The celebrated "magic wand" theorems of Eskin, Mirzakhani, and Mohammadi [14,15], which we will discuss briefly below, may be regarded as providing positive evidence for the existence of a corresponding picture for moduli spaces of translation surfaces. However, as we will see, the emerging picture for the horocycle flow in moduli spaces is more complicated than this simple analogy might suggest.

2. DEFINITIONS AND BACKGROUND

There are several alternative points of view concerning the definitions of translation surfaces, their moduli spaces, and the $\mathrm{SL}_2(\mathbb{R})$-action on them, see Definitions 1, 4, and 5 of [24]. See the surveys mentioned above for more information, and alternative definitions, and see [4, §2] for a more detailed treatment following the point of view we will take here.

A *polygonal surface* (which we will also call a *polygonal presentation of a translation surface*) is a finite collection of polygons in the plane, equipped with a partition of the sides into pairs of parallel sides of equal length and opposite orientation, which we identify by translations.

If e, e' is a pair of identified sides, then there is a unique translation $\varphi = \varphi_{e,e'}$ with $\varphi(e) = e'$, and we say that each $x \in e$ is identified with $\varphi(x) \in e'$. The identifying maps $\{\varphi_{e,e'}\}$ generate an equivalence relation on the polygonal surface. For points in the interior of polygons, the equivalence class is a singleton; for points in the interior of a side, it is a pair of points; and for vertices, it is some finite set of vertices. The union of polygons has a topology as a subset of Euclidean space, and we endow the polygonal surface with the quotient topology for the equivalence relation just defined. Thus the polygonal surface becomes a compact oriented surface. We make the further requirement that it is connected.

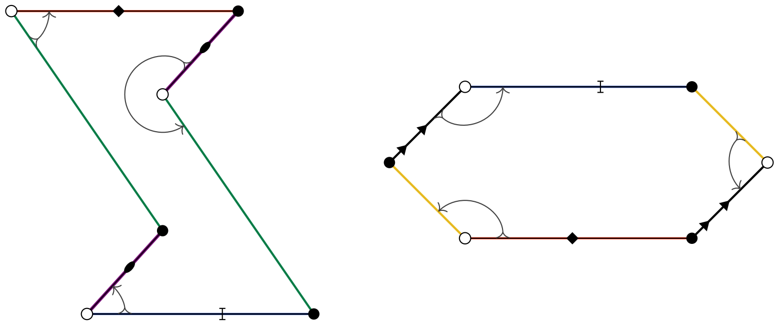


FIGURE 1

A polygonal surface. Parallel edges (in case of ambiguity, those with the same marking) are identified by translations, and the points marked with • and ∘ represent two singularities, each of order 1. The rotating arc around singularity ∘ measures its turning angle of 4π.

A polygonal surface inherits some geometric structures from the plane. Each point has a cone angle which measures the total turning angle made by a curve around the point. At points which are interior points of polygons or of edges, the turning angle is 2π, and for vertices of polygons, it is $2\pi(1+k)$ for some integer $k \geq 0$ measuring the *excess in angle*. Points for which the excess in angle is positive are called *singularities*, and the excess in angle of a singularity is its *order*. One defines the *area* of a polygonal surface, as the sum of the areas of the polygons. The surface also inherits the notion of a *straightline flow* in any direction. This is defined by extending the motion along a straight line by applying the maps $\varphi_{e,e'}$. If a straightline flow reaches a singularity, the straightline trajectory does not extend past the singularity, and thus the straightline flow in a given direction is defined for all times, only on a dense G_δ subset of the polygonal surface. A finite straightline flow trajectory which begins and ends at singular points is called a *saddle connection*. One can also measure the total horizontal and vertical displacement along an oriented path α in a polygonal surface M, i.e., the total amount traveled in the horizontal and vertical directions, when traveling along the path. We denote by $\text{hol}_M(\alpha) \in \mathbb{R}^2$ the *holonomy vector* whose components are these horizontal and vertical displacements. See Figure 2.

There is a *scissors congruence* equivalence relation on polygonal surfaces, generated by the following three operations:

(a) subdividing a polygon into two polygons by adding a diagonal (in this case the two new edges are "both sides" of the new diagonal and they are identified);

(b) the inverse operation of amalgamating two polygons separated by an edge into a larger one by deleting a diagonal; and

(c) translating polygons by translations.

See Figure 3.

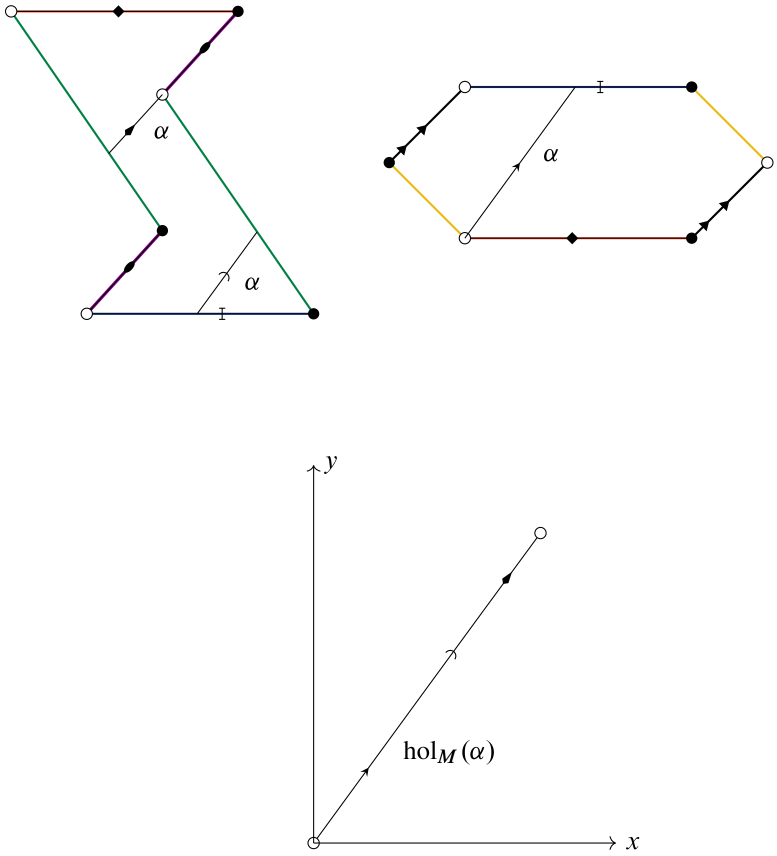


FIGURE 2
Measuring $\mathrm{hol}_M(\alpha)$ for a saddle connection α.

A scissors equivalence class of polygonal surfaces is called a *translation surface*. The number of singularities of a fixed order, the area and the holonomy vectors of piecewise linear paths are the same for polygonal surfaces that are the same up to scissors congruence, and thus make sense on translation surfaces.

The collection of all translation surfaces with a fixed number of singularities of given orders is called a *stratum*; we denote by $\mathcal{H}(a_1, \dots, a_r)$, where $a_1, \dots, a_r$ are positive integers, the stratum of translation surfaces with r singularities, of orders $a_1, \dots, a_r$. The group

$$G = \mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1 \right\}$$

acts on the plane by linear transformations. This action extends to an action on polygonal surfaces (by applying the same linear transformation to each polygon) and preserves scissors

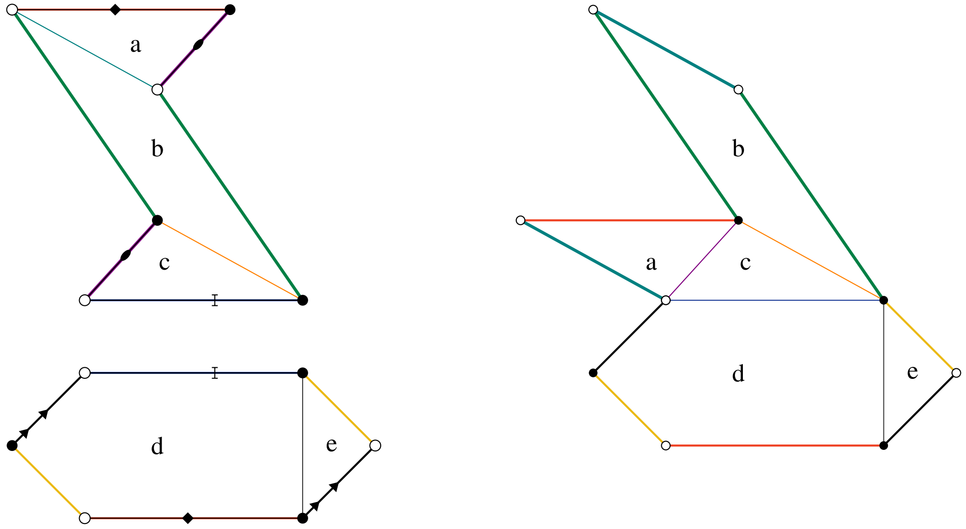


FIGURE 3
Two scissors-equivalent polygonal surfaces.

equivalence, and thus acts on each stratum. The restriction of this action to the group

$$\{g_t : t \in \mathbb{R}\}, \quad \text{where } g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$$

is called the *geodesic flow*. This action has been extensively studied, since the projections of its orbits to the moduli space of Riemann surfaces are parameterized geodesic paths with respect to the Teichmüller metric, and also since it provides a renormalization framework used for studying billiards and interval exchange transformations. Our focus will be on the *horocycle flow*, which is defined as the restriction of the G-action to the subgroup

$$U = \{u_s : s \in \mathbb{R}\}, \quad \text{where } u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}. \tag{2.1}$$

We will pay special attention to orbit-closures for this action. For this, we need to define a topology on a stratum $\mathcal{H}$. The topology we will use is a metrizable topology, characterized by the property that a sequence $\{M_j\}$ of translation surfaces converges to M as $j \to \infty$ if one can choose polygonal surfaces which are representatives for M and each of the M_j such that, for all large enough j, the polygonal surfaces have the same combinatorics (that is, the same number of polygons with the same number of sides and the same side identifications) and the vertices of the polygons comprising M_j converge to the corresponding vertices of the polygons comprising M. See [4, §2.4] for more details.

2.1. Some foundational results

The following fundamental results were proved in the 1980s and 1990s.

- The G-action preserves the area of a translation surface, and we denote by $\mathcal{H}_1(a_1, \dots, a_r)$ the collection of area-one surfaces in $\mathcal{H}(a_1, \dots, a_r)$. There is a natural smooth G-invariant measure of full support on the spaces $\mathcal{H}_1(a_1, \dots, a_r)$, derived from the Lebesgue measure in "period coordinates" via a "cone construction," which was constructed by Masur [20] and Veech [40], and is now referred to as the *Masur–Veech measure*. Masur, Veech, and Masur–Smillie [25] proved that it is finite. Kontsevich and Zorich [19] classified the connected components of strata.

- Masur [20] and Veech [38] showed that interval exchange transformations can be suspended and understood via straighline flows on translation surfaces, and that the $\{g_t\}$-action can be used to renormalize the straightline flow dynamics. They used this approach to settle a conjecture of Keane concerning the unique ergodicity of interval exchange transformations.

- Masur [20] and Veech [39] showed that the G-action is ergodic with respect to the Masur–Veech measure. This implies that there are dense orbits for the horocycle flow and the geodesic flow, and that these flows are both mixing.

- Veech [37] gave examples of $\mathbb{Z}/2\mathbb{Z}$ skew products of rotations that can be interpreted as flows on translation surfaces [26]. In these translation surfaces one has directions in which the straightline flow is minimal but not uniquely ergodic. Similar examples were also independently constructed by Sataev [32]. This phenomenon of minimality without unique ergodicity of the straightline flow will play an important role in our discussion. Masur [22] established a link between nondivergence of geodesic trajectories and unique ergodicity of foliations. Masur and Smillie [25] showed that, while they are rare, there are abundant examples of minimal and not uniquely ergodic flows on translation surfaces.

- Veech [41] gave example of surfaces whose G-orbit carries a finite G-invariant measure (such surfaces are now known as *Veech surfaces*). Using the connection to the G-dynamics, he showed that for Veech surfaces, the straightline flow dynamics admits a complete description.

- Masur [21] showed that G-orbits are never bounded, and used this to show the existence of periodic trajectories for rational billiards. On the other hand, Smillie (see [35,42]) showed that a G-orbit Gq is closed if and only if q is a Veech surface.

2.2. The analogy with Ratner's work, and the magic wand theorem

The results mentioned in Section 2.1 can be seen as counterparts of similar results in the setting of homogeneous flows. Around the end of the 20th century, several researchers began to speculate that there might be a translation surface analogue of Ratner's celebrated theorems on the action of groups generated by unipotents, acting on homogeneous spaces. In particular, see [1,11], whose authors noted the usefulness of obtaining analogues of Ratner's theorem for applications in geometry and dynamics of translation surfaces.

What makes Ratner's results so powerful is that they are able to shed light on the behavior of *every* orbit (in contrast to softer results in ergodic theory which describe the behavior of typical orbits). Indeed, this analogy led to the hope that it might be possible to completely classify all invariant measures and all orbit-closures for the G-action and the U-action, as these are the only two connected subgroups of G (up to conjugation) that are generated by unipotent one parameter subgroups. McMullen [27] established such a result for the G-action in genus two, see also Calta [6] for earlier strong results in this direction. These results gave further impetus to work in this direction. The search was officially on when Zorich published an influential survey [48] with a section titled "hope for a magic wand." For the G-action, the conjecture was confirmed in a spectacular fashion by Eskin, Mirzakhani, and Mohammadi in [14,15]. This work has revolutionized the study of dynamics on translation surfaces and has already had many applications in geometry which we do not survey here. As a sample of their results, we have the following:

Theorem 2.1 (Eskin–Mirzakhani–Mohammadi, P-genericity). *For any translation surface q, there is a measure ν whose support is the orbit-closure $\overline{Gq}$ and such that, for any compactly supported continuous test function φ on the stratum containing q,*

$$\frac{1}{T}\int_0^T\int_0^1 \varphi(g_t u_s q)\,ds\,dt \xrightarrow[T\to\infty]{} \int \varphi\, d\nu.$$

The measure ν is affine in natural coordinates, see [15, DEF. 1.1] *for a precise statement.*

These developments left open the question of whether a similar result was possible for the U-action, i.e., whether it is possible to classify all the U-orbit closures in terms of some algebraic or geometric data. Can one understand all U-invariant ergodic measures, and the asymptotic distribution of averages along any U-orbit? While the focus of this survey is on the horocycle dynamics as an interesting subject in its own right, we note that positive answers to these questions would have far-reaching consequences for some counting problems associated with billiards and flat surfaces. However, as we will see in this survey, the behavior of U-orbits in strata of translation surfaces can be quite different from the behavior of unipotent trajectories in homogeneous spaces.

3. BEHAVIOR OF INDIVIDUAL HOROCYCLE ORBITS

3.1. Some early results

Using ideas of Kerckhoff, Masur, and Smillie [18], Veech [43] showed that there is no orbit of the horocycle flow that diverges in $\mathcal{H}$. That is, for any $q \in \mathcal{H}$, there is a compact $K \subset \mathcal{H}$ such that the set of visit times

$$\{s > 0 : u_s q \in K\}$$

is unbounded. A quantitative strengthening of this result was obtained by Minsky and Weiss [28]: for any $q \in \mathcal{H}$ and any $\varepsilon > 0$, there is a compact subset $K \subset \mathcal{H}$ such that

$$\liminf_{T\to\infty} \frac{1}{T}\left|\left\{s \in [0,T] : u_s q \in K\right\}\right| > 1-\varepsilon.$$

These results are parallels of quantitative nondivergence results of Dani and Margulis for unipotent flows on homogeneous spaces (see [10]). With these results in hand, Smillie and Weiss [34] classified the minimal sets for the U-action on $\mathcal{H}$. They showed that for any $q \in \mathcal{H}$, the orbit-closure $\overline{Uq}$ contains a *minimal set*, i.e., a closed U-invariant subset containing no proper closed U-invariant subsets. Furthermore, $\overline{Uq}$ is minimal if and only if the straightline flow in the horizontal direction on the underlying surface M_q is completely periodic. It follows that any orbit-closure for the U-action contains such a horizontally completely periodic surface.

In some rather special settings, it was possible to completely classify the U-invariant measures and orbit-closures. For Veech surfaces, this follows from results in homogeneous dynamics, as was observed in [13]. The first result of this kind in a nonhomogeneous setting is due to Eskin, Marklof, and Witte Morris [12], who studied surfaces which are branched covers of Veech surfaces. This work was later extended by Calta and Wortman [7] and Bainbridge, Smillie, and Weiss [4]. In [4], a complete classification of U-invariant measures and orbit-closures is given within the eigenform loci in genus two. In these loci, which are 5-dimensional G-orbit-closures in $\mathcal{H}(1,1)$ arising in McMullen's genus-two classification, we have a complete understanding of the possible orbit-closures and invariant measures. In these examples one can observe some phenomena not present for the G-dynamics, for instance, orbit-closures which are manifolds with nonempty boundary and an infinitely generated fundamental group. Nevertheless, these partial results were all consistent with a putative "magic wand theorem for horocycles."

3.2. Recent results

The situation changed in our work [9]. In this paper we proved the following results. We recall that if μ is a measure on $\mathcal{H}$ and $q \in \mathcal{H}$, we say that *q is generic for μ* if for any compactly supported continuous function f on $\mathcal{H}$ one has

$$\frac{1}{T}\int_0^T f(u_s q)\,ds \xrightarrow[T\to\infty]{} \int_{\mathcal{H}} f\,d\mu. \tag{3.1}$$

Theorem 3.1. *Let $\mathcal{H} = \mathcal{H}(1,1)$ be the stratum of genus two surfaces with two singular points. Then:*

(1) *There is a surface $q \in \mathcal{H}$ and a G-invariant ergodic measure μ on $\mathcal{H}$ such that q is generic for μ but* $\mathrm{supp}(\mu) \subsetneq \overline{Uq}$.

(2) *There is a surface $q \in \mathcal{H}$ which is not generic for any measure.*

(3) *There is a surface $q \in \mathcal{H}$ whose orbit-closure is a fractal, in the sense that the Hausdorff dimension of $\overline{Uq}$ is not an integer.*

We make some remarks to put these results in context. The third item in Theorem 3.1 is perhaps the most striking, but the first two are also in stark contrast with a “magic wand paradigm.” Note that in (1), the support of the measure limit measure μ is not the closure of the orbit – compare with Theorem 2.1 or Ratner’s work. Also in (2) we see that there is no analogue of Theorem 2.1 for the horocycle flow.

The orbit-closure we construct in (3) has an explicit description. The precise statement requires some technical preparation and will not be discussed here, see [9, THM. 1.8].

The stratum $\mathcal{H}(1,1)$ is the simplest one in which we are able to exhibit a surface satisfying (3) but it is likely that our method can be extended to many other strata. However, we are not able to establish (1) and (3) in the stratum $\mathcal{H}(2)$. Note, however, that (2) holds in $\mathcal{H}(2)$ by the work of Chaika, Khalil, and Smillie [8].

In the next sections we will explain some of the ideas of [9], focusing on the proofs of (1) and (2).

4. TREMORS

The dynamical properties of a horocycle flow trajectory are intimately related to those of the horizontal straightline flow on the corresponding surfaces. Following [9], we will use the notation q to refer to a surface in a stratum, and M_q to refer to the underlying translation surfaces. Although they are formally identical, we will use the symbol q when the dynamical system we are considering is primarily the G-action on the stratum $\mathcal{H}$ containing q, and we will use $M = M_q$ when we are considering the dynamics of the horizontal straightline flow on the underlying translation surface. From now on, by straightline flow we always mean the horizontal straightline flow, which we will denote by $\{\phi_t\}$ (the surface on which the flow takes place will be clear from the context).

Let ν be a $\{\phi_t\}$-invariant measure on M. The simplest example is the Lebesgue measure on the individual polygons. The second simplest example is the restriction of Lebesgue measure to a polygonal subsurface which is $\{\phi_t\}$-invariant; for example, in Figure 1, the restriction of Lebesgue measure to one of the two hexagons in the picture (note that these hexagons are separated from each other by two horizontal saddle connections, and thus each is $\{\phi_t\}$-invariant). Finally, a more interesting example referred to earlier, is the case when the straightline flow is minimal but not uniquely ergodic; in that case there will be two or more mutually singular $\{\phi_t\}$-invariant measures, all supported on the entire surface M. By a standard result (see, e.g., [26]), if the straightline flow is not minimal then the surface contains a horizontally invariant polygonal subsurface, and thus this list exhausts all possible cases. Note that for almost every surface M, with respect to the measures discussed in Section 2.1, the only $\{\phi_t\}$-invariant measure (up to scaling) is Lebesgue measure.

Let σ be any nonhorizontal segment in M. Such a segment is known as a *cross-section*. We consider σ as a piece of a trajectory for a (nonhorizontal) straightline flow, thus parameterizing it by an interval, where we choose the *positive orientation* on σ so that $\mathrm{hol}_M(\sigma) = (x_\sigma, y_\sigma)$ satisfies $y_\sigma > 0$. From ν we can construct a *cross-section measure* on

σ via the formula

$$\beta_{\sigma,\nu}(A) = \beta(A) = \lim_{\varepsilon \to 0+} \frac{1}{\varepsilon} \nu\big(\{\phi_t(a) : a \in A, t \in [0, \varepsilon]\}\big).$$

This classical construction defines a bijection between straightline flow invariant measures on M and measures on σ which are invariant under the first return map to σ along horizontal lines (see [2]).[1] If $\nu = \text{Leb}$ is Lebesgue measure on M, the cross-section measure on σ (viewed as an interval via its parameterization) is a multiple of one-dimensional Lebesgue measure. The system of measures $\beta_\nu = \{\beta_{\sigma,\nu} : \sigma \text{ is a cross-section}\}$ is an example of a *transverse measure* (corresponding to ν).[2] The transverse measure corresponding to Leb will be called the *canonical transverse measure*. Similarly, if ν is the restriction of Lebesgue measure to a polygonal subsurface, the cross-section measure on each σ is the restriction of Lebesgue measure to a finite collection of subintervals. If $\{\phi_t\}$ is minimal but not uniquely ergodic, the cross-section measures are fully supported measures which may be distinct from Lebesgue measure.

Let us now express the action of the horocycle element u_s (notation as in (2.1)). Recall that u_s acts on polygonal surfaces by tilting or *shearing* polygons. The computation

$$\text{hol}_{u_s M}(\sigma) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_\sigma \\ y_\sigma \end{pmatrix} = \begin{pmatrix} x_\sigma + s y_\sigma \\ y_\sigma \end{pmatrix} = \text{hol}_M(\sigma) + \begin{pmatrix} s y_\sigma \\ 0 \end{pmatrix}$$

shows that the amount of tilting of a side σ of a polygon is proportional to its measure, with respect to the canonical transverse measure. We can tweak this definition and replace each appearance of y_σ with $\beta(\sigma)$, where β is a transverse measure on M. This idea gives rise to the *tremor map*. It takes as input a surface q, a transverse measure β, and a "time parameter" s, and produces a new translation surface $q' = \text{trem}_{s,\beta}(q)$, where $M' = M_{q'}$ is defined by assigning to each side σ of a polygonal presentation of M_q the holonomy

$$\text{hol}_{M'}(s) = \text{hol}_M(\sigma) + \begin{pmatrix} s\beta(\sigma) \\ 0 \end{pmatrix}.$$

A basic observation is that this definition makes sense. That is, there is a polygon presentation of M_q for which the adjusted segments in the above definition still give a polygonal presentation of a surface M', and, moreover, M' does not depend on a particular choice of a polygonal presentation. Furthermore,[3] it can be shown that $\text{trem}_{s,\beta}(q)$ is defined for all values of s. On the other hand, the tremor map is not a flow in the sense that the choice of β depends on the initial translation surface M_q. For most choices of M_q, the only choice for β is the canonical transverse measure, and in that case $\text{trem}_{s,\beta}(q)$ is nothing but the horocycle

[1] More precisely, for this correspondence we need σ to intersect every straightline trajectory; this always happens when the straightline flow is minimal but it will be convenient to relax this condition and define $\beta_{\sigma,\nu}(A)$ for any σ.

[2] In [9] we use a more general definition of transverse measures, but the only transverse measures we will need in this survey arise from straightline flow invariant measures via this construction.

[3] Recall that in this survey we discuss a more restrictive class of transverse measures. This assertion is false in the more general context considered in [9].

image $u_s q$. However, for surfaces M_q for which there are noncanonical transverse measures, we get other tremor paths $\{\mathrm{trem}_{s,\beta}(q) : s \in \mathbb{R}\}$.

Sometimes it will be helpful to ignore the dependence of $\mathrm{trem}_{s,\beta}(q)$ on s, and we will write $\mathrm{trem}_{\beta}(q) = \mathrm{trem}_{1,\beta}(q)$. Note that the multiple of a transverse measure by a positive scalar is also a transverse measure, and we have the obvious identity

$$\mathrm{trem}_{s\beta}(q) = \mathrm{trem}_{s,\beta}(q). \tag{4.1}$$

It is sometimes helpful to work with signed measures, which turns the set of all "signed transverse measures" into a real vector space. We call elements of this vector space *signed foliation cycles*. One can extend the definition of a tremor to the case in which β is a signed foliation cycle, and then one obtains identities like (4.1) for all $s \in \mathbb{R}$. In this more general setup, the set of transverse measures forms a convex cone C_q^+ in the vector space of signed foliation cocycles. See [9, §6] for more details.

A crucial fact for our analysis is the fact that surfaces which are obtained from one another by a tremor "have the same horizontal foliation." To make this precise, in [9, §5], we show the existence of a homeomorphism $\psi = \psi_\beta : M_q \to M_{q'}$, which is a topological conjugacy between the straightline flow on M_q and the straightline flow on $M_{q'}$, i.e.,

$$\forall t \in \mathbb{R}, \quad \phi'_t \circ \psi = \psi \circ \phi_t, \tag{4.2}$$

where ϕ_t, ϕ'_t denote respectively the straightline flows on M_q and $M_{q'}$. The pushforward map ψ_* induces a bijection between the straightline flow invariant measures on M_q and those on $M_{q'}$, and thus between the cones of transverse measures $C_q^+, C_{q'}^+$. In particular, this holds when β is canonical, i.e., when $q' \in Uq$. Thus if M_q is not horizontally uniquely ergodic, then the same holds for $M_{q'}$. Furthermore, we have the relation

$$\forall s \in \mathbb{R}, \quad u_s \mathrm{trem}_\lambda(q) = \mathrm{trem}_\lambda(u_s q), \tag{4.3}$$

where we have used λ to denote both a transverse measure on M_q, and its image under ψ_*. Formally, this is a commutation relation between the maps $q \mapsto \mathrm{trem}_\lambda(q)$ and $q \mapsto u_s(q)$. Note, however, that off of a set of measure zero, any tremor is just the horocycle flow and (4.3) is just the relation $u_{s_1} \circ u_{s_2} = u_{s_2} \circ u_{s_1}$.

5. SOME IDEAS IN THE PROOF OF THEOREM 3.1

5.1. U-orbits of tremored surfaces almost track U-orbits

The starting point for our analysis is the following observation:

Proposition 5.1. *There is a proper complete metric* dist *on $\mathcal{H}$, inducing the topology, such that the following holds. Let $q \in \mathcal{H}$ be such that M_q admits a noncanonical transverse measure $\beta = \beta_\nu$, where ν is a straightline flow invariant measure satisfying $\nu \ll$* Leb. *Let $q' = \mathrm{trem}_\beta(q)$. Then*

$$\sup_{s\in\mathbb{R}} \mathrm{dist}\big(u_s q, u_s q'\big) < \infty. \tag{5.1}$$

Note that by (4.3), $u_s q' = \mathrm{trem}_\beta(u_s q)$. A useful (but imprecise) heuristic explanation of (5.1) is that a fixed tremor can only move points a bounded distance. The metric dist appearing in Proposition 5.1 was introduced in [3].

A more detailed analysis of the function $s \mapsto \mathrm{dist}(u_s q, u_s q')$ appearing in (5.1) yields the following statement.

Theorem 5.2. *Let $\mathcal{M} = \overline{Gq} \subsetneq \mathcal{H}$ be a G orbit-closure, and let ν be the G-invariant ergodic measure on $\mathcal{M}$ (as in Theorem 2.1). Suppose q is generic for ν and M_q is horizontally minimal but not uniquely ergodic. Let β be a noncanonical transverse measure on M_q such that $q' = \mathrm{trem}_\beta(q) \notin \mathcal{M}$. Then there is $s_0 \in \mathbb{R}$ so that the surface $q_0 = u_{s_0} q$ satisfies*

$$\forall \varepsilon > 0, \quad \frac{1}{T}\left|\left\{s \in [0,T] : \mathrm{dist}\big(u_s q', u_s q_0\big) \geq \varepsilon\right\}\right| \xrightarrow[T\to\infty]{} 0. \tag{5.2}$$

Note that $q_0 \in \mathcal{M}$ is also generic for ν, since q is. If β were the canonical transverse measure then $q' = u_{s_0} q = q_0$ and (5.2) would be vacuously true. The result asserts that when $q' \notin \mathcal{M}$, the trajectory of q' nevertheless spends all but a negligible proportion of its time arbitrarily close to the trajectory of a generic point. In particular, it "falls back on $\mathcal{M}$." Since genericity is not affected by modifying a trajectory on a set of zero measure, we see that Theorem 5.2 implies (1) of Theorem 3.1, provided one can find examples of $\mathcal{M}$ and q for which the conditions of Theorem 5.2 are satisfied.

That such examples exist follows from the genericity results in [4]. Indeed, in the setting of eigenform loci in $\mathcal{H}(1,1)$ studied in that paper, the condition of having a minimal but not uniquely ergodic horizontal straightline flow does not have any effect on the asymptotic distribution of a horocycle orbit. In the simplest of these examples, $\mathcal{M}$ can be taken to be the collection of surfaces in $\mathcal{H}(1,1)$ which admit a $2:1$ branched covering of a torus (this orbit-closure is denoted by $\mathcal{E}_4$ in McMullen's classification [27]).

We now explain the idea behind the proof of (5.2). It is useful to view a transverse measure as a cohomology class. Indeed, a transverse measure (or indeed, a signed foliation cocycle) assigns a real number to any positively oriented transverse segment on M_q. One can check that the assignment $\sigma \mapsto \beta_{\sigma,\nu}(\sigma)$ (where $\beta = \{\beta_{\sigma,\nu}\}$) is a cochain representing a cohomology class in $H^1(M_q, \Sigma_q; \mathbb{R})$, where Σ_q is the set of singularities. Consider the vector bundle $\mathcal{B}$ over $\mathcal{H}$ for which the fiber over q is $H^1(M_q, \Sigma; \mathbb{R})$, that is,

$$\begin{tikzcd} H^1(M_q, \Sigma_q; \mathbb{R}) \arrow[r] & \mathcal{B} \arrow[d] \\ & \mathcal{H} \end{tikzcd}$$

The bundle $\mathcal{B}$ has a simple description as a subbundle of the tangent bundle of $\mathcal{H}$, with the pair (q, β) representing the tangent direction of the curve $s \mapsto \mathrm{trem}_{s,\beta}(q)$.[4] In particular, $\mathcal{B}$ has a natural topology, and in this topology the set of cones of transverse measures

[4] To make this description precise one should work in the category of orbifold bundles.

C_q^+ is closed (see [9, §4.1, §13]). That is, if $\beta_n \in H^1(M_{q_n}, \Sigma_{q_n}; \mathbb{R})$ are cohomology classes represented by transverse measures, and $(q_n, \beta_n) \to (q_\infty, \beta_\infty)$ as elements of $\mathcal{B}$, then β_∞ is also represented by a transverse measure on M_{q_∞}. Furthermore, the map

$$\mathcal{B} \to \mathcal{H}, \quad (q, \beta) \mapsto \mathrm{trem}_\beta(q) \tag{5.3}$$

is continuous with respect to the topology on $\mathcal{B}$.

This basic fact seems to contradict our previous heuristic that surfaces which are in the same U-orbit have the same cone of transverse measures. Indeed, suppose $q_n = u_n q$ for $u_n \in U$, and q and thus all of the q_n have a noncanonical transverse measure, but $q_n \to q_\infty$ where q_∞ has a uniquely ergodic straightline flow. Then we have that all the q_n have the same fixed cone C_q^+ of transverse measures, containing both the canonical transverse measure and a noncanonical one, while at the same time this cone of transverse measures converges to $C_{q_\infty}^+$, which is just the ray generated by the canonical transverse measure. How is this possible?

The answer is that, with respect to any reasonable metric on $\mathcal{B}$, the bijection sending C_q^+ to $C_{q_n}^+$ is far from being an isometry. One can define norms $\|\cdot\|_q$ on each $H^1(M_q, \Sigma_q; \mathbb{R})$, which are continuous in the bundle topology, with respect to which the unit length transverse measures $\{\beta \in C_{q_n}^+ : \|\beta\|_{q_n} = 1\}$ all converge to the unique unit length transverse measure in $C_{q_\infty}^+$. Thus, the cones $C_{q_n}^+$, although of dimension > 1 and in bijection with each other, "collapse" down to the ray $C_{q_\infty}^+$.

Using this idea, in the proof of Theorem 5.2 we show that, for any $\varepsilon > 0$, there is an open set $\mathcal{U} \subset \mathcal{M}$ containing all the uniquely ergodic surfaces such that, for any $q_1 \in \mathcal{U}$, the diameter of $\{\beta \in C_{q_1}^+ : \|\beta\|_{q_1} = 1\}$ is at most ε. By genericity, the orbit Uq spends all but a negligible proportion of its time in $\mathcal{U}$, and since the map (5.3) is continuous for the metric dist, (5.2) follows.

5.2. From genericity to lack of genericity

Recall that from Birkhoff's theorem, any ergodic U-invariant measure assigns full measure to its generic points. It is sometimes useful to work with *quasigeneric points* instead. These are defined as points q for which there is a sequence $T_n \to \infty$ such that for all compactly supported continuous test functions f,

$$\frac{1}{T_n} \int_0^{T_n} f(u_s q)\, ds \underset{n\to\infty}{\longrightarrow} \int f\, d\mu.$$

Note that a generic point is quasigeneric, but a point can be quasigeneric for two measures. If q_0 is quasigeneric for two measures μ and ν, we can take f for which $\int f\, d\mu \neq \int f\, d\nu$ to see that the limit $\lim_{T\to\infty} \frac{1}{T} \int_0^T f(u_s q_0)\, ds$ does not exist, and, in particular, q_0 is not generic for any measure. Thus for a given dynamical system, the condition that there are distinct invariant measures, and every point is generic for one of them, implies that there are no points which are quasigeneric for two different measures. Recall that this condition is satisfied for unipotent flows on homogeneous spaces (by Ratner's work), as well as for some averages on moduli spaces of translation surfaces (e.g., the horocycle flow in the settings of [4,7,12], or two-dimensional averages for G-invariant measures as in Theorem 2.1).

To see that this is quite restrictive, we note that the set of quasigeneric points for some measure ν is a G_δ set. Indeed, let $\{f_j\}_{j\in\mathbb{N}}$ be a dense countable collection of continuous compactly supported test functions. For each $j \in \mathbb{N}$, each $T > 0$, and each $\varepsilon > 0$, continuity of the action implies that

$$\mathcal{U}_{j,T,\varepsilon} = \left\{ q \in \mathcal{H} : \text{ for } i = 1, \dots, j, \ \left| \frac{1}{T} \int_0^T f_i(u_s q)\, ds - \int f_i \, d\mu \right| < \varepsilon \right\}$$

is open. The set of quasigeneric points can be written as

$$\bigcap_{j=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{T=k}^{\infty} \mathcal{U}_{j,T,1/j},$$

proving the claim.

By the Baire category theorem, any two dense G_δ subsets intersect. Let $\mathcal{M}$ be as in Theorem 5.2, let μ and ν be the fully supported G-invariant measures on $\mathcal{H}(1,1)$ and on $\mathcal{M}$ respectively. By Birkhoff's theorem, the set of generic points for μ is dense in $\mathcal{H}(1,1)$. Thus item (2) of Theorem 3.1 follows from Theorem 5.2 and the following:

Proposition 5.3. *The set of surfaces of the form*

$$\left\{ \mathrm{trem}_\beta(q) : q \in \mathcal{M} \textit{ is generic for } \nu \textit{ and horizontally minimal}, \beta \in C_q^+ \right\} \tag{5.4}$$

is dense in $\mathcal{H}(1,1)$.

In order to prove this statement, we recall the observation "if q' is obtained from q by a tremor then M_q and $M_{u_s q}$ have the same transverse measures," which we discussed above in connection with (4.3). We add to it the additional observation "if q' is obtained from q by the $\{g_t\}$-action then M_q and $M_{q'}$ have the same transverse measures." This is proved using a similar idea to the comparison homeomorphism of [9, §5]. Namely, the definition of the g_t action shows that if $q' = g_{t_0} q$ then there is a homeomorphism $M_q \to M_{q'}$ which intertwines the straightline flow up to a time change. That is, if $\{\phi_t\}$ and $\{\phi'_t\}$ denote respectively the horizontal straightline flows on M_q and $M_{q'}$, and $\psi : M_q \to M_{q'}$ is the map obtained by acting on a polygonal presentation as in the definition of the g_t action, then

$$\forall t \in \mathbb{R}, \quad \phi'_{e^{t_0} t} \circ \psi = \psi \circ \phi_t.$$

It follows from this that analogously to (4.3), one has

$$\forall \beta \in C_q^+,\ t \in \mathbb{R}, \quad g_t \mathrm{trem}_{s,\beta}(q) = \mathrm{trem}_{e^t s,\beta}(g_t q)$$

(where we consider β simultaneously as belonging to C_q^+ and $C_{g_t q}^+$ via the above bijection). Together with (4.3), we find that the set of surfaces $\mathcal{F}$ defined in (5.4) is invariant under both flows $\{g_t\}$, $\{u_s\}$, and hence, by Theorem 2.1, $\overline{\mathcal{F}}$ is G-invariant. Moreover, $\mathcal{M} \subsetneq \overline{\mathcal{F}}$, and examining the possibilities for $\mathcal{F}$ in McMullen's classification [27] gives that $\mathcal{F}$ is dense in $\mathcal{H}(1,1)$.

6. QUESTIONS

There are many open questions about the horocycle flow on strata of translation surfaces. We list some of them. We begin with one of the most outstanding questions in the field:

Question 1. Is there an "exotic" U-ergodic measure? For example, measures whose support has noninteger Hausdorff dimension, or fully supported measures that differ from Masur–Veech measure.

A precise statement of the above question is tricky because Calta [6] and Smillie and Weiss [36] gave examples of U-invariant ergodic measures whose support is a manifold with boundary and infinitely generated fundamental group.

The following question is motivated by renormalization dynamics:

Question 2. If ν is a U-invariant ergodic measure that is not G-invariant, are there two G-invariant ergodic measures μ_-, μ_+ so that

- $\mathrm{supp}(\mu_-) \subsetneqq \mathrm{supp}(\mu_+)$,
- $g_t \nu \underset{t \to +\infty}{\longrightarrow} \mu_+$,
- $g_t \nu \underset{t \to -\infty}{\longrightarrow} \mu_-$?

Note that we consider the zero measure to be a G-invariant ergodic measure. Even the special case of measures supported on periodic horocycles (where μ_- is the zero measure) is open and very interesting. The same question is also interesting for horocycle orbit closures.

Some basic questions on the topological dynamics of the horocycle flow are open. To us, the following is the most outstanding example.

Question 3. Is the horocycle flow recurrent as a topological dynamical system? That is, is it true that for every $q \in \mathcal{H}$ there exists a sequence $t_i \nearrow \infty$ so that $u_{t_i} q \underset{i \to \infty}{\longrightarrow} q$?

In the realm of orbit closures:

Question 4. In [9] we construct an exotic U orbit-closure, which is the orbit closure of the tremor of a translation surface in an "eigenform locus." What are all the orbit closures of tremors of translation surfaces in eigenform loci that do not have horizontal saddle connections? Do they all have the description given in [9, EQ. (1.8)]? Informally, is any such orbit-closure the set of all surfaces obtained from tremoring surfaces in a given eigenform locus by at most a certain fixed amount?

Of course, we are also interested in other horocycle orbit-closures, including those that arise from tremors of surfaces in proper G-orbit closures outside of $\mathcal{H}(1,1)$.

In special cases (see [4, 7, 12, 13]) the horocycle flow has been shown to behave much like it does in homogeneous settings. For example, every point is generic for some

U-invariant ergodic measure. These examples are all *rank-one loci* in the sense of [45]. This motivates the following general question.

Question 5. In the special setting of rank-one loci, what can be said about the behavior of the horocycle flow?

There is a growing dictionary between the earthquake flow and the horocycle flow. This dictionary was initiated by Mirzakhani [29], who used it to prove that the earthquake flow is ergodic. Calderon and Farre [5] have added to this dictionary and extended it to other actions, which has allowed them to showcase additional behavior of the earthquake flow. It is interesting to see whether some of these results can be proven directly in the setting of earthquake flows and if any arguments in the setting of earthquake flows can be used to show new behavior of horocycle flows on strata.

ACKNOWLEDGMENTS

We are grateful to John Smillie, who is our collaborator and coauthor for much of the work described in this survey. We are also grateful to CIRM, where this collaboration began.

FUNDING

J. Chaika was supported in part by a Poincare chair, a Warnock chair, NSF grants DMS-1452762 and DMS-2055354. B. Weiss was supported in part by BSF grant 2016256 and ISF grant 2919/19.

REFERENCES

[1] A. Eskin, H. Masur, Asymptotic formulas on flat surfaces. *Ergodic Theory Dynam. Systems* **21** (2001), no. 2, 443–478.

[2] W. Ambrose, Representation of ergodic flows. *Ann. of Math. (2)* **42** (1941), 723–739.

[3] A. Avila, S. Gouëzel, and J.-C. Yoccoz, Exponential mixing for the Teichmüller flow. *Publ. Math. Inst. Hautes Études Sci.* (2006), no. 104, 143–211.

[4] M. Bainbridge, J. Smillie, and B. Weiss, Horocycle dynamics: new invariants and the eigenform loci in the stratum $\mathcal{H}(1,1)$. *Mem. Amer. Math. Soc.* (to appear), arXiv:1603.00808.

[5] A. Calderon and J. Farre, Shear-shaped cocycles for measured laminations and ergodic theory of the earthquake flow. 2021, arXiv:2102.13124.

[6] K. Calta, Veech surfaces and complete periodicity in genus two. *J. Amer. Math. Soc.* **17** (2004), no. 4, 871–908.

[7] K. Calta and K. Wortman, On unipotent flows in $\mathcal{H}(1,1)$. *Ergodic Theory Dynam. Systems* **30** (2010), no. 2, 379–398.

[8] J. Chaika, O. Khalil, and J. Smillie, On the space of ergodic measures for the horocycle flow on strata of abelian differentials. 2021, arXiv:2104.00554.

[9] J. Chaika, J. Smillie, and B. Weiss, Tremors and horocycle dynamics on the moduli space of translation surfaces. 2020, arXiv:2004.04027.

[10] S. G. Dani, On invariant measures, minimal sets and a lemma of Margulis. *Invent. Math.* **51** (1979), no. 3, 239–260.

[11] A. Eskin, Counting problems and semisimple groups. In *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, pp. 539–552, Extra Vol. II, 1998.

[12] A. Eskin, J. Marklof, and D. Witte Morris, Unipotent flows on the space of branched covers of Veech surfaces. *Ergodic Theory Dynam. Systems* **26** (2006), no. 1, 129–162.

[13] A. Eskin, H. Masur, and M. Schmoll, Billiards in rectangles with barriers. *Duke Math. J.* **118** (2003), no. 3, 427–463.

[14] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127** (2018), 95–324.

[15] A. Eskin, M. Mirzakhani, and A. Mohammadi, Isolation, equidistribution, and orbit closures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Ann. of Math. (2)* **182** (2015), no. 2, 673–721.

[16] G. Forni, Asymptotic behaviour of ergodic integrals of 'renormalizable' parabolic flows. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pp. 317–326, Higher Ed. Press, Beijing, 2002.

[17] G. Forni and C. Matheus, Introduction to Teichmüller theory and its applications to dynamics of interval exchange transformations, flows on surfaces and billiards. *J. Mod. Dyn.* **8** (2014), no. 3–4, 271–436.

[18] S. Kerckhoff, H. Masur, and J. Smillie, Ergodicity of billiard flows and quadratic differentials. *Ann. of Math. (2)* **124** (1986), no. 2, 293–311.

[19] M. Kontsevich and A. Zorich, Connected components of the moduli spaces of Abelian differentials with prescribed singularities. *Invent. Math.* **153** (2003), no. 3, 631–678.

[20] H. Masur, Interval exchange transformations and measured foliations. *Ann. of Math. (2)* **115** (1982), no. 1, 169–200.

[21] H. Masur, Closed trajectories for quadratic differentials with an application to billiards. *Duke Math. J.* **53** (1986), no. 2, 307–314.

[22] H. Masur, Hausdorff dimension of the set of nonergodic foliations of a quadratic differential. *Duke Math. J.* **66** (1992), no. 3, 387–442.

[23] H. Masur, Teichmüller space, dynamics, probability. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pp. 836–849, Birkhäuser, Basel, 1995.

[24] H. Masur, Ergodic theory of translation surfaces. In *Handbook of dynamical systems. Vol. 1B*, pp. 527–547, Elsevier B. V., Amsterdam, 2006.

[25] H. Masur and J. Smillie, Hausdorff dimension of sets of nonergodic measured foliations. *Ann. of Math. (2)* **134** (1991), no. 3, 455–543.

[26] H. Masur and S. Tabachnikov, Rational billiards and flat structures. In *Handbook of dynamical systems, Vol. 1A*, pp. 1015–1089, North-Holland, Amsterdam, 2002.

[27] C. T. McMullen, Dynamics of $\mathrm{SL}_2(\mathbb{R})$ over moduli space in genus two. *Ann. of Math. (2)* **165** (2007), no. 2, 397–456.

[28] Y. Minsky and B. Weiss, Nondivergence of horocyclic flows on moduli space. *J. Reine Angew. Math.* **552** (2002), 131–177.

[29] M. Mirzakhani, Ergodic theory of the earthquake flow. *Int. Math. Res. Not. IMRN* (2008), no. 3, Art. ID rnm116, 39.

[30] M. Möller, Geometry of Teichmüller curves. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*, pp. 2017–2034, World Sci. Publ., Hackensack, NJ, 2018.

[31] M. Ratner, Interactions between ergodic theory, Lie groups, and number theory. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pp. 157–182, Birkhäuser, Basel, 1995.

[32] E. A. Sataev, The number of invariant measures for flows on orientable surfaces. *Izv. Akad. Nauk SSSR Ser. Mat.* **39** (1975), no. 4, 860–878.

[33] R. E. Schwartz, Survey lecture on billiards, to appear in the *Proceedings of the ICM*.

[34] J. Smillie and B. Weiss, Minimal sets for flows on moduli space. *Israel J. Math.* **142** (2004), 249–260.

[35] J. Smillie and B. Weiss, Finiteness results for flat surfaces: a survey and problem list. In *Partially hyperbolic dynamics, laminations, and Teichmüller flow*, pp. 125–137, Fields Inst. Commun. 51, Amer. Math. Soc., Providence, RI, 2007.

[36] J. Smillie and B. Weiss, Examples of horocycle invariant measures on the moduli space of translation surfaces, in preparation.

[37] W. A. Veech, A Kronecker–Weyl theorem modulo 2. *Proc. Natl. Acad. Sci. USA* **60** (1968), 1163–1164.

[38] W. A. Veech, Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math. (2)* **115** (1982), no. 1, 201–242.

[39] W. A. Veech, The Teichmüller geodesic flow. *Ann. of Math. (2)* **124** (1986), no. 3, 441–530.

[40] W. A. Veech, Moduli spaces of quadratic differentials. *J. Anal. Math.* **55** (1990), 117–171.

[41] W. A.Veech, Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards. *Invent. Math.* **97** (1989), no. 3, 553–583.

[42] W. A. Veech, Geometric realizations of hyperelliptic curves. In *Algorithms, fractals, and dynamics (Okayama/Kyoto, 1992)*, pp. 217–226, Plenum, New York, 1995.

[43] W. A. Veech, Measures supported on the set of uniquely ergodic directions of an arbitrary holomorphic 1-form. *Ergodic Theory Dynam. Systems* **19** (1999), no. 4, 1093–1109.

[44] Y. B. Vorobets, Plane structures and billiards in rational polygons: the Veech alternative. *Uspekhi Mat. Nauk* **51** (1996), no. 5(311), 3–42.

[45] A. Wright, Cylinder deformations in orbit closures of translation surfaces. *Geom. Topol.* **19** (2015), no. 1, 413–438.

[46] A. Wright, Translation surfaces and their orbit closures: an introduction for a broad audience. *EMS Surv. Math. Sci.* **2** (2015), no. 1, 63–108.

[47] J.-C. Yoccoz, Interval exchange maps and translation surfaces. In *Homogeneous flows, moduli spaces and arithmetic*, pp. 1–69, Clay Math. Proc. 10, Amer. Math. Soc., Providence, RI, 2010.

[48] A. Zorich, Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pp. 437–583, Springer, Berlin, 2006.

[49] A. Zorich, Geodesics on flat surfaces. In *International Congress of Mathematicians. Vol. III*, pp. 121–146, Eur. Math. Soc., Zürich, 2006.

JON CHAIKA

University of Utah, Department of Mathematics, 203, 155 S 1400 E RM 233, Salt Lake City, UT, 84112-0090, USA, chaika@math.utah.edu

BARAK WEISS

Tel Aviv University, Department of Mathematics, Tel-Aviv, 69978 Israel, barakw@tauex.tau.ac.il

TOPOLOGICAL ENTROPY AND PRESSURE FOR FINITE-HORIZON SINAI BILLIARDS

MARK F. DEMERS

ABSTRACT

This brief survey describes recent progress in our understanding of a variety of equilibrium states for finite-horizon dispersing billiard maps in two dimensions. In particular, we review formulations of topological entropy and pressure for the family of geometric potentials $-t \log J^u T$, where $J^u T$ denotes the unstable Jacobian of the map and $t \in \mathbb{R}$. We summarize recent results, proving the existence and uniqueness of related equilibrium states for some range of $t \geq 0$, including $[0, 1]$. In this family, $t = 0$ corresponds to the measure of maximal entropy, while $t = 1$ corresponds to the smooth invariant measure for the billiard map. In addition, variational principles are presented which express topological notions of pressure and entropy as the supremum of their measure-theoretic counterparts.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37C83; Secondary 37D35, 37C30, 37B40

KEYWORDS

Dispersing billiard, topological pressure, topological entropy, variational principle, transfer operator, anisotropic Banach spaces

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3432–3459
DOI 10.4171/ICM2022/52

1. INTRODUCTION

The study of mathematical billiards as prototypical examples of mechanical systems with frictionless collisions was introduced by Sinai [57] and subsequently developed by many authors. In such models, a finite number of convex obstacles are placed on a two-dimensional torus, forming the billiard table, and a point particle is set in motion, moving with constant velocity between collisions, and undergoing elastic reflections at the boundary. The billiard map is the discrete-time map which takes the particle from one collision to the next. Despite the presence of singularities (the map is discontinuous and its derivative is unbounded near tangential collisions) the map preserves a smooth invariant measure, and the ergodic properties of the map with respect to this measure have been studied extensively through a variety of techniques, including Markov partitions [14] and sieves [16], Young towers [21,60], and the spectral analysis of the associated transfer operator [32–34].

It is also possible to introduce variations on the dynamics, by varying the shape of the boundary or by including external forces such as electric fields and potentials which act on the particle between collisions [24,25,29], or twists and kicks at the moment of collision [45,61]. For small forces, the dynamics resemble those of the classical billiard [8,22,23], while for large forces the dynamics may change significantly [36,45,52]. The subject quickly becomes vast and technical, so in this note we will focus on the dynamics of the classical Sinai billiard without external forces. The book [27] by Chernov and Markarian provides an excellent introduction to the subject.

The purpose of this expository note is to introduce the reader, without delving into too many technicalities, to recent developments in the study of a family of equilibrium states for this class of billiards. Traditionally, and in all the references listed above, the focus has been on the ergodic and statistical properties of the map with respect to the Sinai–Ruelle–Bowen (SRB) measure, which in the unperturbed case has a smooth density with respect to Lebesgue measure, such as ergodicity, mixing and the Bernoulli property [37,57], rate of decay of correlations [21,60], dynamical Central Limit Theorem [14], and related limit theorems [32,46,51]. Here, instead, we outline the progress made in [3,4] regarding the family of geometric potentials, $-t \log J^u T$, $t \in \mathbb{R}$, determining the existence and uniqueness of the associated equilibrium states. The importance of this family lies in the fact that $t = 1$ corresponds to the SRB measure while $t = 0$ corresponds to the measure of maximal entropy. More generally, the parameter t has been linked to the Hausdorff dimension of certain invariant sets [12,42].

Despite this, geometric potentials have received relatively little attention in the context of billiards. For $t = 0$, the topological entropy of a finite-horizon Sinai billiard map T was studied in [20] by identifying a full Lebesgue measure set of points M_1 that can be coded via a countable Markov partition. Chernov showed that the topological entropy of T restricted to M_1 is equal to the topological entropy of the induced topological Markov chain and used this to obtain a lower bound on the growth of periodic orbits. Yet, no invariant measure achieving this topological entropy was constructed and whether the set M_1 saw the full topological entropy of the system was left open. For t near 1, the preprint [19] obtains results regarding equilibrium states for this class of geometric potentials. Yet, uniqueness is not proved, and left open is the possible connection to a topological notion of pressure.

With these questions in mind, the papers [3,4] represent a significant advance in our understanding of topological entropy and pressure and related equilibrium states. It is these ideas and the techniques involved that we hope to illuminate in this note. These, naturally, lead to further questions, several of which we formulate at the end of this review.

The paper is organized as follows. In Section 2, we present the minimal background on dispersing billiards necessary for the subsequent discussion, and state the main results from [3, 4]. In Section 3 we outline the main approach and principle estimates needed to prove the variational principle for $t > 0$, and in Section 4 we do the same for the case $t = 0$. In Section 5, we formulate some open problems related to the equilibrium states we will construct.

2. PRELIMINARIES AND STATEMENTS OF MAIN THEOREMS

A Sinai billiard table is a subset of the torus $\mathbb{T}^2$ obtained by removing finitely many pairwise disjoint, closed convex sets B_i, i.e., $Q = \mathbb{T}^2 \setminus (\bigcup_{i=1}^d B_i)$. The B_i are called scatterers and are assumed to have C^3 boundaries with strictly positive curvature $\mathcal{K}$. The billiard flow is the motion of a point particle in Q traveling at unit speed and undergoing specular reflections (angle of incidence equals angle of reflection) at collisions with the scatterers.

We introduce coordinates on ∂Q by parametrizing ∂B_i according to arclength and recording at each collision the position r and the angle φ made by the postcollision velocity vector with the outward pointing normal to the boundary. Thus the phase space for the map, $M = (\bigcup_{i=0}^d \partial B_i) \times [-\frac{\pi}{2}, \frac{\pi}{2}]$, is a union of cylinders, and for each $x = (r, \varphi) \in M$, the billiard map $T(r, \varphi) = (r_1, \varphi_1)$ maps one collision to the next. The map preserves a smooth probability measure, $d\mu_{\mathrm{SRB}} = (2|\partial Q|)^{-1} dr d\varphi$, which is ergodic, indeed Bernoulli, and enjoys exponential decay of correlations on smooth observables, as described in the Introduction.

Let $\tau(x)$ denote the (Euclidean) distance from x to $T(x)$ in Q. We say the billiard has *finite horizon* if there is no trajectory making only tangential collisions. This implies, in particular, that $\tau_{\max} := \sup \tau < \infty$. In addition, the fact that the scatterers are disjoint guarantees $\tau_{\min} := \inf \tau > 0$. Setting $\mathcal{K}_{\min} = \inf \mathcal{K} > 0$ and $\mathcal{K}_{\max} = \sup \mathcal{K} < \infty$, it follows [27, SECT. 4.4] that the stable and unstable cones in the tangent space $\mathbb{R}^2$,

$$\mathcal{C}^s = \left\{(dr, d\varphi) \;\Big|\; -\mathcal{K}_{\max} - \frac{1}{\tau_{\min}} \leq \frac{d\varphi}{dr} \leq -\mathcal{K}_{\min}\right\},$$
$$\mathcal{C}^u = \left\{(dr, d\varphi) \;\Big|\; \mathcal{K}_{\min} \leq \frac{d\varphi}{dr} \leq \mathcal{K}_{\max} + \frac{1}{\tau_{\min}}\right\}$$

are strictly invariant under DT^{-1} and DT, respectively, whenever the derivatives exist. Away from tangential collisions, T is uniformly hyperbolic, i.e., for $\Lambda := 1 + 2\tau_{\min}\mathcal{K}_{\min}$, there exists $C_0 > 0$ such that for all $n \geq 0$,

$$\begin{aligned} \|DT^n(x)v\| &\geq C_0\Lambda^n\|v\|, \quad \forall v \in \mathcal{C}^u, \quad \text{and} \\ \|DT^{-n}(x)v\| &\geq C_0\Lambda^n\|v\|, \quad \forall v \in \mathcal{C}^s. \end{aligned} \tag{2.1}$$

2.1. Singularities and distortion

Denote the set of tangential collisions by $\mathcal{S}_0 = \{x = (r, \varphi) \in M \mid \varphi = \pm\frac{\pi}{2}\}$. The singularity set for T^n, $n \in \mathbb{Z}$, is defined by

$$\mathcal{S}_n = \bigcup_{i=0}^{n} T^{-i}\mathcal{S}_0.$$

Assuming finite horizon, $\mathcal{S}_n$ comprises a finite collection of C^2 curves for each n. Indeed, the hyperbolicity of T implies an alignment property: for $n > 0$, $\mathcal{S}_n \setminus \mathcal{S}_0$ comprises decreasing curves in $\mathcal{C}^s$, while for $n < 0$, $\mathcal{S}_n \setminus \mathcal{S}_0$ comprises increasing curves in $\mathcal{C}^u$ [27, PROP. 4.45].

While the set $\bigcup_{n\in\mathbb{Z}} \mathcal{S}_n$ is dense in M [27, SECT. 4.11], its μ_{SRB}-measure is 0. Setting $M' = M \setminus \bigcup_{n\in\mathbb{Z}} \mathcal{S}_n$, it follows that the stable/unstable subspaces $E^s(x)$ and $E^u(x)$ are defined for all $x \in M'$. Thus we may define the stable/unstable Jacobians of T by

$$J^sT(x) = \left\| DT(x)|_{E^s(x)} \right\| \quad \text{and} \quad J^uT(x) = \left\| DT(x)|_{E^u(x)} \right\| \quad \forall x \in M'.$$

If x has a stable/unstable manifold of positive length (which also occurs on a full-measure set $M'' \subset M'$ [27, THM. 4.66]), then J^sT and J^uT serve as the Jacobians for a change of variables when integrating along these manifolds with respect to arclength. We let $\mathcal{W}^s$ denote the set of local stable manifolds of T with length at most $\delta_0 > 0$, which is chosen to guarantee a local complexity condition. See Lemma 3.1 for $t > 0$ and Section 4.1.1 for $t = 0$.

In fact, $DT(x)$ becomes unbounded as $T(x)$ approaches $\mathcal{S}_0$. To compensate for this, the standard technique is to introduce *homogeneity strips* that partition the space into a countable set of horizontal strips accumulating on $\mathcal{S}_0$ and which are effectively treated as singularity curves in exchange for providing some control of distortion. Specifically, choosing[1] $q > 1$ and an index $k_0 \in \mathbb{N}$, one defines for $k \geq k_0$,

$$\mathbb{H}_k = \left\{(r, \varphi) \mid (k+1)^{-q} \leq \frac{\pi}{2} - \varphi \leq k^{-q}\right\},$$

with a similar definition for $\mathbb{H}_{-k}$ approaching $\varphi = -\pi/2$. Let $\mathcal{W}^s_{\mathbb{H}} \subset \mathcal{W}^s$ denote the set of curves $W \in \mathcal{W}^s$ which lie in a single homogeneity strip. Such curves are called *weakly homogeneous* stable manifolds for T.

It follows that there exists $C_d > 0$ such that for all $n \geq 0$, if $T^i(x), T^i(y) \in T^iW \in \mathcal{W}^s_{\mathbb{H}}$ for each $0 \leq i \leq n-1$, then

$$\left| \frac{J^sT^n(x)}{J^sT^n(y)} - 1 \right| \leq C_d\, d(x, y)^{1/(q+1)}, \tag{2.2}$$

which is the desired distortion control (see [27, LEMMA 5.27] or [4, LEMMA 2.1]).

2.2. Measure-theoretic pressure for geometric potentials

Let $t \in \mathbb{R}$ and μ be an invariant probability measure for T. Define the *pressure of* μ with respect to the geometric potential $-t \log J^uT$ to be

$$P_\mu(-t\log J^uT) = h_\mu(T) - t\int_M \log J^uT\, d\mu,$$

[1] The standard choice for dispersing billiards is $q = 2$, yet here we will choose q depending on the parameter t in our potential.

where $h_\mu(T)$ denotes the Kolmogorov–Sinai entropy[2] of μ. If μ satisfies

$$P_\mu(-t \log J^u T) = P(t) := \sup\{P_\nu(-t \log J^u T) \mid \nu \in \mathcal{I}\},$$

where $\mathcal{I}$ is the set of invariant probability measures for T, then μ is called an *equilibrium state* for the potential, and $P(t)$ is the *pressure of* $-t \log J^u T$.

The theory of equilibrium states has been well established for Hölder-continuous potentials, first for Anosov and Axiom A systems [11, 53, 58], and then for nonuniformly hyperbolic systems using a variety of techniques [17, 18, 44, 50, 56]. Much less is known in the case of billiards. For $t = 1$, it is known that $P_{\mu_{\mathrm{SRB}}}(-\log J^u T) = 0$; this is the so-called Pesin entropy formula [27, THM. 3.42]. Yet, the uniqueness of this equilibrium state was not proved until [4]. For t near 1, Chen–Wang–Zhang [19] prove existence, but not uniqueness, of equilibrium states using Young towers.

One of the complications of studying equilibrium states associated with this potential is that the unstable Jacobian is not Hölder continuous; indeed, $J^u T$ is not continuous on any open set and it is not bounded (for x near $\mathcal{S}_1$, $J^u T(x) \sim 1/\cos\varphi(Tx)$). Yet, it is regular along homogeneous unstable manifolds (as the time reversal of (2.2) demonstrates) and can be approximated by smooth functions in the distributional norms we will define in Sections 3.2 and 4.2, permitting the analysis we will describe here.

In addition to proving the existence and uniqueness of equilibrium states for the family of geometric potentials, we are interested in expressing the pressure $P(t)$ in terms of topological notions of entropy for $t = 0$ and pressure for $t > 0$. We define such notions precisely in the next two subsections.

Remark 2.1. For $t < 0$, $P(t) = \infty$ if there is a periodic orbit making a grazing collision. In this case, if ν is the atomic measure supported on such a periodic orbit, then $P_\nu(t) = \infty$ as well. Thus the (possibly many) measures maximizing the pressure are simple to describe, so we will not discuss the case $t < 0$ here.

2.3. Topological entropy and variational principle for $t = 0$

Following [3], define for $n, k \geq 0$,

$$\mathcal{M}_{-k}^n = \{\text{maximal connected components of } M \setminus (\mathcal{S}_{-k} \cup \mathcal{S}_n)\}. \tag{2.3}$$

Thus elements of $\mathcal{M}_0^n$ are the (open) domains of continuity for T^n and $\mathcal{M}_{-n}^0$ plays the analogous role for T^{-n}. We define the topological entropy of T to be the exponential rate of growth of $\#\mathcal{M}_0^n$, where $\#A$ denotes the cardinality of the set A.

Definition 2.2 (Topological entropy). Define $h_* := \lim_{n\to\infty} \frac{1}{n} \log(\#\mathcal{M}_0^n)$.

The limit above exists due to the submultiplicativity of $\#\mathcal{M}_0^n$ [3, LEMMA 3.3]. Note also that if $A \in \mathcal{M}_0^n$, then $T^n A \in \mathcal{M}_{-n}^0$, so that $\#\mathcal{M}_0^n = \#\mathcal{M}_{-n}^0$ and hence $h_*(T) = h_*(T^{-1})$.

[2] Since T admits a finite generating partition, $h_\mu(T)$ is necessarily finite for any T-invariant probability measure.

One can connect h_* to the usual Bowen definitions of topological entropy via both ε-separated and ε-spanning sets, whose definitions we do not recall here. Although such definitions are usually made for continuous maps, it is a consequence of [3, THM. 2.3] that both of the Bowen definitions coincide with h_*.

The main result in [3] is the following.

Theorem 2.3 (Measure of maximal entropy and variational principle). *Let T be a finite-horizon Sinai billiard map as defined above. Under a sparse recurrence condition on the singularity set, defined in* (4.1)*, there exists a unique measure μ_0 such that*

$$h_* = h_{\mu_0}(T) = \sup_{\mu \in \mathcal{I}} h_\mu(T).$$

Moreover, μ_0 is hyperbolic and Bernoulli,[3] *has no atoms and is positive on open sets.*

A more complete set of properties for μ_0 can be found in [3, THM. 2.6]. That $h_* \geq P_\mu(0)$ for any T-invariant probability measure μ is due to a soft, classical argument (see, for example, [59, PROP. 9.10]) since the sequence $\mathcal{M}_0^n$ is related to a finite generating partition for T [3, LEMMA 3.3]. The main work of [3] is to construct an invariant measure μ_0 whose entropy equals h_*. This requires a precise understanding of the geometry of the sets $\mathcal{M}_0^n$ combined with some functional analytic techniques, whose main ideas are described in Section 4.

2.4. Topological pressure and variational principle for $t > 0$

Before defining our notion of topological pressure for the potential $-t \log J^u T$, it is convenient to consider the corresponding potential in the associated transfer operator. Indeed, arguing by analogy to smooth hyperbolic systems, the transfer operator $\tilde{\mathcal{L}}_t$ with spectral radius $e^{P(t)}$ is defined, for example, on bounded, measurable functions, by

$$\tilde{\mathcal{L}}_t f = \frac{f \circ T^{-1}}{((J^u T)^t J^s T) \circ T^{-1}}.$$

For a Sinai billiard, setting $E(x) = \sin(\angle(E^s(x), E^u(x)))$ to denote the sine of the angle between the stable and unstable subspaces at x, and denoting by $J_{\mathrm{Leb}} T$ the Jacobian of T with respect to Lebesgue measure on M, we have

$$\begin{aligned} \frac{\cos\varphi(x)}{\cos\varphi(Tx)} &= J_{\mathrm{Leb}} T(x) = J^s T(x) J^u T(x) \frac{E(Tx)}{E(x)} \\ \implies (J^u T)^t J^s T &= \left(\frac{E \cos\varphi}{(E\cos\varphi) \circ T} \right)^t (J^s T)^{1-t}. \end{aligned}$$

Since the two potentials are related by a coboundary, the associated transfer operators will have the same spectral radius, so we will study instead the operator

$$\mathcal{L}_t f = \frac{f \circ T^{-1}}{(J^s T)^{1-t} \circ T^{-1}}. \tag{2.4}$$

Remark that $J^s T \approx \cos\varphi$ so that the potential is unbounded whenever $t \neq 1$.

[3] By hyperbolic, we mean that μ_0-a.e. point has stable and unstable manifolds of positive length. By Bernoulli, we mean that it is isomorphic to a Bernoulli shift, which implies also that μ_0 is ergodic and K-mixing.

2.4.1. Weight function for the topological pressure

In order to control the evolution of $\mathcal{L}_t f$ in Section 3.2, it will be necessary to control integrals of the form

$$\int_W \mathcal{L}_t^n f \psi \, dm_W = \int_{T^{-n}W} f \psi \circ T^n \left| J^s T^n \right|^t dm_{T^{-n}W},$$

where $W \in \mathcal{W}^s$, m_W is (unnormalized) arclength on W, ψ is a Hölder-continuous test function, and f is an element of the Banach space we will construct.

In order for $(J^s T^n)^t$ to play the role of a test function, (2.2) suggests that we decompose $T^{-1}W$ into a countable collection of maximal curves $W_i^1 \in \mathcal{W}^s_{\mathbb{H}}$ and then iterate these, subdividing into homogeneous components at each step until time n. We denote this collection of curves comprising $T^{-n}W$ by $\mathcal{G}_n^{\mathbb{H}}(W)$. Then the spectral properties of $\mathcal{L}_t$ depend on the growth of

$$\sum_{W_i \in \mathcal{G}_n^{\mathbb{H}}(W)} \left| J^s T^n \right|^t_{C^0(W_i)} \qquad \text{as a function of } n \text{ and } W. \tag{2.5}$$

This toy calculation suggests the weight we use to define the topological pressure below.

2.4.2. Definition of topological pressure

Define $\mathcal{S}_0^{\mathbb{H}} = \mathcal{S}_0 \cup (\bigcup_{|k| \geq k_0} \partial H_k)$ and for $n \in \mathbb{Z}$, $\mathcal{S}_n^{\mathbb{H}} = \bigcup_{i=0}^n T^{-i} \mathcal{S}_0^{\mathbb{H}}$. This will act as an extended singularity set for T^n where we introduce artificial cuts in order to preserve bounded distortion. Let

$$\mathcal{M}_0^{n,\mathbb{H}} := \left\{ \text{maximal connected components of } M \setminus (\mathcal{S}_{n-1}^{\mathbb{H}} \cup T^{-n}\mathcal{S}_0) \right\}.$$

We define the weighted sum

$$Q_n(t) = \sum_{A \in \mathcal{M}_0^{n,\mathbb{H}}} \sup_{x \in A \cap M'} \left| J^s T^n(x) \right|^t,$$

and the topological pressure for $t > 0$ is the exponential rate of growth of $Q_n(t)$.

Definition 2.4 (Topological pressure). We let $P_*(t) := \lim_{n \to \infty} \frac{1}{n} \log Q_n(t)$.

As with Definition 2.2, the limit above exists and equals the lim inf due to the submultiplicativity of $Q_n(t)$. It follows that $Q_n(t) \geq e^{nP_*(t)}$ for each $n \geq 0$.

The first theorem from [4] says that $P_*(t)$ dominates the metric pressures.

Theorem 2.5 (Variational inequality). *Let T be a finite-horizon Sinai billiard map. Then $P_*(t)$ is a convex, continuous, decreasing function for $t > 0$, and $P(t)$ satisfies*

$$P_*(t) \geq P(t) = \sup\left\{ P_\nu(-t \log J^u T) \mid \nu \textit{ is an invariant probability measure for } T \right\}.$$

Remark that for any T-invariant measure μ, $\int \log J^u T \, d\mu = -\int \log J^s T \, d\mu$, which is useful for relating $P(t)$ with $P_*(t)$. As with Theorem 2.3, the inequality $P_*(t) \geq P(t)$ is straightforward, while the main work lies in constructing a measure μ_t such that $P_{\mu_t}(-t \log J^u T) = P_*(t)$. This again requires a detailed analysis of the growth rate of $Q_n(t)$ and the pressure of $\mathcal{G}_n^{\mathbb{H}}(W)$ from (2.5) as a function of n and $W \in \mathcal{W}^s_{\mathbb{H}}$.

2.4.3. Equilibrium state and variational principle

To prove that, in fact, $P(t) = P_*(t)$ and produce an equilibrium state, we restrict our range of t. Recalling $\Lambda > 1$ from (2.1), define

$$t_* := \sup\{t > 0 \mid P(t) > -t \log \Lambda\}.$$

Remark 2.6. The definition of t_* is motivated by the fact that Λ^{-t} controls the local growth in complexity due to singularities (and homogeneity strips), while $e^{P_*(t)}$ controls the global growth in complexity via $Q_n(t)$. Then $t < t_*$ implies the "pressure gap" condition $\Lambda^{-t} < e^{P(t)} \leq e^{P_*(t)}$.

Note that since $P(1) = 0$, $\Lambda > 1$ and $P(t)$ is decreasing, it must be that $t_* > 1$.

The main result in this setting from [4] is the following.

Theorem 2.7 (Unique equilibrium state and variational principle). *Let T be a finite-horizon Sinai billiard and let $t \in (0, t_*)$. Then $P_*(t) = P(t)$ and there exists a unique equilibrium state μ_t for the potential $-t \log J^u T$, i.e.,*

$$P_{\mu_t}(-t \log J^u T) = P_*(t) = P(t).$$

Moreover, μ_t is hyperbolic, has no atoms, is positive on every open set, and enjoys exponential decay of correlations against Hölder observables.

The main technique used in the proof of the theorem is the construction of anisotropic Banach spaces of distributions, adapted to t, on which the transfer operator $\mathcal{L}_t$ has a spectral gap. Then the measure μ_t is constructed as a product of left and right eigenvectors of $\mathcal{L}_t$, following the standard Parry construction (see, for example, [43, SECT. 4.4] for an introduction, or [41] for an application in the case of Anosov diffeomorphisms). Indeed, our control of the spectrum of $\mathcal{L}_t$ also implies the following theorem.

Theorem 2.8 (Analyticity of $P(t)$). *The pressure function $P(t)$ is analytic for $t \in (0, t_*)$, with*

$$P'(t) = \int \log J^s T \, d\mu_t = -\int \log J^u T \, dm_t < 0 \quad \textit{and}$$

$$P''(t) = \sum_{k \geq 0} \left[\int (\log J^s T \circ T^k) \log J^s T \, d\mu_t - \big(P'(t)\big)^2 \right] \geq 0.$$

Moreover, $P''(t) = 0$ if and only if $\log J^s T = f - f \circ T + P'(t)$ for some $f \in L^2(\mu_t)$.

If there exists $s \neq t \in (0, t_*)$ such that $\mu_s = \mu_t$, then $P(t)$ is affine on $(0, t_*)$ and $\log J^s T$ is μ_t-a.e. cohomologous to a constant for all $t \in (0, t_*)$.

Finally, under the sparse recurrence condition (4.1), $\lim_{t \downarrow 0} P(t) = P(0) = h_*$.

We conjecture that, in fact, $\mu_s \neq \mu_t$ for any $s \neq t$ in $(0, t_*)$, i.e., $J^s T$ cannot be cohomologous to a constant for a Sinai billiard. If it were, then the theorem would imply that $P(t)$ is affine and, by uniqueness, $\mu_t = \mu_{\mathrm{SRB}}$ for each $t \in [0, t_*)$. See Section 5.

3. IDEAS FROM THE PROOF OF THEOREM 2.7

In this section, we present some of the key ideas in the proof of Theorem 2.7. In light of Theorem 2.5, they divide into two principal parts: (1) geometric estimates that control local complexity and establish a uniform exponential rate of growth for $Q_n(t)$; (2) the functional-analytic framework needed to construct a measure μ_t with pressure $P_*(t)$.

The development of a functional-analytic framework in which to study transfer operators for hyperbolic systems has a well-established history. After some early success in the hyperbolic analytic case [54,55], intense interest was generated by the paper of Blank, Keller, and Liverani [10], which launched a series of subsequent papers developing a variety of Banach spaces for Anosov and Axiom A maps [1,7,40,41]. This was later extended to piecewise hyperbolic maps [5,6,31] and ultimately to a variety of hyperbolic billiards [3,32–34]. See [2] for a comprehensive survey or [30] for a gentle introduction to the use of such spaces. The norms we shall define in Section 3.2 are a natural adaptation of these ideas to the family of geometric potentials.

3.1. Growth lemmas and exact exponential growth of $Q_n(t)$

Our first goal is to establish precise bounds on the exponential growth of $Q_n(t)$ as well as the growth of the pressure over $\mathcal{G}_n^{\mathbb{H}}(W)$ from (2.5). In order to accomplish this, we fix $t_0 > 0$ and $t_1 < t_*$ and obtain uniform bounds for t in the closed interval $[t_0, t_1]$.

Fix $q \geq 2/t_0$ and let $\theta \in (\Lambda, 1)$ be such that $\theta^{t_1} < e^{P_*(t_1)}$. The latter choice is possible by definition of t_*, and implies by the convexity of $P_*(t)$ that $\theta^t < e^{P_*(t)}$ for all $t \in [t_0, t_1]$. Next we adapt the usual one-step expansion (see [27, LEMMA 5.56]) to our potential.

Lemma 3.1. *There exist k_0, $\delta_0 > 0$ such that*

$$\sup_{\substack{V \in \mathcal{W}^s \\ |V| \leq \delta_0}} \sum_{V_i} \left| J^s T \right|_{*,C^0(V_i)}^t < \theta^t \quad \textit{for all } t \geq t_0,$$

where V_i are the maximal, connected homogeneous components of $T^{-1}V$ and $|\cdot|_*$ denotes the sup *norm with respect to an adapted metric.*[4]

Sketch of proof. Due to the finite-horizon condition, a short stable manifold V can be cut by at most $\tau_{\max}/\tau_{\min}$ tangential collisions under T^{-1} and all but one of these collisions are nearly grazing. Near grazing collisions, $V_k \subset \mathbb{H}_k$ and, since $J^s T \sim \cos\varphi$,

$$\sum_{k \geq k_0} \left| J^s T \right|_{*,C^0(V_k)}^t \leq C \sum_{k \geq k_0} k^{-qt} \leq C' k_0^{-1} \quad \text{since } qt \geq 2.$$

Thus setting $\varepsilon = \theta - \Lambda^{-1} > 0$, we choose k_0 in the definition of homogeneity strips so large that $C' k_0^{-1} \frac{\tau_{\max}}{\tau_{\min}} < \varepsilon$. Finally, choose δ_0 so small that if $|V| \leq \delta_0$, then $T^{-1}V$ can intersect only homogeneity strips of index at least k_0 at the nearly tangential collisions. This is possible since $|T^{-1}V| \leq C|V|^{1/2}$ [27, EXERCISE 4.60]. ■

[4] The adapted metric is defined as in [27, SECT. 5.10] so that in (2.1) the constant $C_0 = 1$, i.e., the expansion is seen in one step. The lemma applies equally well to more general cone-stable curves and its time reversal to cone-unstable curves.

The one-step expansion expressed by Lemma 3.1 guarantees that the expansion provided by the weight $1/(J^sT)^t$ along stable manifolds mapped by T^{-1} is strong enough to overcome the effect of cutting by both the primary and secondary discontinuities of T^{-1}. This can be iterated inductively to obtain statements regarding the prevalence of long pieces in both $T^{-n}W$ and $\mathcal{M}_0^{n,\mathbb{H}}$, as summarized in Lemma 3.2 below.

Recalling the definition of $\mathcal{G}_n^{\mathbb{H}}(W)$ from (2.5), for $\delta_1 < \delta_0$ and $W \in \mathcal{W}^s$, define $\mathcal{G}_n^{\delta_1,\mathbb{H}}(W)$ as a decomposition of $T^{-n}W$ in an analogous manner with $\mathcal{G}_n^{\mathbb{H}}(W)$, but with pieces longer than length δ_1 subdivided into length between $\delta_1/2$ and δ_1 at each step (rather than length δ_0).

For a set $E \subset M$, we use $\operatorname{diam}^u(E)$ to denote the length of the longest cone-unstable curve in E, and $\operatorname{diam}^s(E)$ to denote the length of the longest cone-stable curve in E.

Lemma 3.2. (a) *$\forall \varepsilon > 0\ \exists \delta_1, n_1 > 0$ such that $\forall W \in \mathcal{W}^s$ with $|W| \geq \delta_1/3$ and all $n \geq n_1$,*

$$\sum_{\substack{W_i \in \mathcal{G}_n^{\delta_1,\mathbb{H}}(W_i) \\ |W_i| < \delta_1/3}} \left|J^sT^n\right|^t_{C^0(W_i)} \leq \varepsilon \sum_{W_i \in \mathcal{G}_n^{\delta_1,\mathbb{H}}(W)} \left|J^sT^n\right|^t_{C^0(W_i)}.$$

(b) *For $A \in \mathcal{M}_0^{n,\mathbb{H}}$, let $B_{n-1}(A)$ denote the connected component of $M \setminus (\bigcup_{i=0}^{n-1} T^i \mathcal{S}_0^{\mathbb{H}})$ containing $T^{n-1}A$. Define $\mathcal{A}_n(\delta) = \{A \in \mathcal{M}_0^{n,\mathbb{H}} : \operatorname{diam}^u(B_{n-1}(A)) \geq \delta/3\}$. There exist $\delta_2 \leq \delta_1$ and $c_0 > 0$ such that*

$$\sum_{A \in \mathcal{A}_n(\delta_2)} \sup_{x \in A \cap M'} \left|J^sT^n(x)\right|^t \geq c_0 Q_n(t), \quad \forall n \in \mathbb{N}, \forall t \in [t_0, 1].$$

Comments on proof. The proof of part (a) relies on iterating Lemma 3.1 combined with the following lower bound on growth valid for $t \leq 1$ and $V \in \mathcal{W}^s$,

$$\begin{aligned}\sum_{W_i \in \mathcal{G}_k^{\delta_1,\mathbb{H}}(V)} \left|J^sT^k\right|^t_{C^0(W_i)} &= \sum_{W_i \in \mathcal{G}_k^{\delta_1,\mathbb{H}}(V)} \left|J^sT^k\right|_{C^0(W_i)} \left|J^sT^k\right|^{t-1}_{C^0(W_i)} \\ &\geq C_1 \Lambda^{k(1-t)} \sum_{W_i \in \mathcal{G}_k^{\delta_1,\mathbb{H}}(V)} \frac{|T^kW_i|}{|W_i|} \geq C_1 \Lambda^{k(1-t)} |V| \delta_1^{-1},\end{aligned}$$

which guarantees that long pieces most continue to produce a sufficient number of long pieces. Once (a) is proved for $t \leq 1$, we extend it to $t \in (1, t_1]$ via interpolation (see [4, SECT. 3.4]).

(b) The proof of (b) follows the same lines as (a), using a version of Lemma 3.1 for elements of $\mathcal{M}_0^{n,\mathbb{H}}$ and a generalization of bounded distortion which says that $J^sT^n(x)$, $J^sT^n(y)$ are comparable when x, y belong to the same element of $\mathcal{M}_0^{n,\mathbb{H}}$. ■

Using Lemma 3.2, we can prove the following key results regarding the uniform growth of $W \in \mathcal{W}^s$ and a type of supermultiplicativity for $Q_n(t)$.

Proposition 3.3. (a) *$\exists c_1 > 0$ such that $\forall W \in \mathcal{W}^s$ with $|W| \geq \delta_1/3$,*

$$\sum_{W_i \in \mathcal{G}_n^{\mathbb{H}}(W)} \left|J^sT^n\right|^t_{C^0(W_i)} \geq c_1 Q_n(t), \quad \forall n \geq 1, \forall t \in [t_0, t_1].$$

(b) *$\exists c_2 > 0$ such that for all $k, n \geq 1$ and all $t \in [t_0, t_1]$, $Q_{n+k}(t) \geq c_2 Q_n(t) Q_k(t)$.*

Sketch of proof. Let $L_n^{\delta_1}(W)$ denote the elements of $\mathcal{G}_n^{\delta_1,\mathbb{H}}(W)$ longer than $\delta_1/3$. Then (b) follows from (a) and Lemma 3.2 (choosing $\varepsilon = 1/2$ there) since,

$$\sum_{W_i \in \mathcal{G}_{n+k}^{\delta_1,\mathbb{H}}(W)} |J^sT^{n+k}|^t_{C^0(W_i)} \geq C \sum_{V_j \in L_n^{\delta_1}(W)} |J^sT^n|^t_{C^0(V_j)} \sum_{W_i \in \mathcal{G}_k^{\delta_1,\mathbb{H}}(V_j)} |J^sT^k|^t_{C^0(W_i)}$$
$$\geq \frac{C}{2} \sum_{V_j \in \mathcal{G}_n^{\delta_1,\mathbb{H}}} |J^sT^n|^t_{C^0(V_j)} c_1 Q_k(t) \geq \frac{C}{2} c_1^2 Q_n(t) Q_k(t).$$

The proof of (a) relies on covering a full μ_{SRB}-measure set of M with a finite collection of Cantor rectangles, formed by maximal intersections of local stable and unstable manifolds so that each rectangle has a hyperbolic product structure. By [27, LEMMA 7.87], we may choose a finite collection of such rectangles, $\mathcal{R}(\delta_2) = \{R_i\}_{i=1}^{N_{\delta_2}}$, such that any cone-stable or cone-unstable curve of length at least $\delta_2/3$ properly crosses at least one R_i. Let

$$\mathcal{A}_n^i := \{A \in \mathcal{A}_n(\delta_2) \subset \mathcal{M}_0^{n,\mathbb{H}} \mid B_{n-1}(A) \text{ properly crosses } R_i\}.$$

By Lemma 3.2(b), there must exist i_* such that $\sum_{A \in \mathcal{A}_n^{i_*}} \sup_A |J^sT^n|^t \geq \frac{c_0}{N_{\delta_2}} Q_n(t)$.

Let $W \in \mathcal{W}^s$ with $|W| \geq \delta_1/3 \geq \delta_2/3$. W must properly cross one R_j. Since μ_{SRB} is mixing, we may ensure that $V = T^{-N}W$ properly crosses R_{i_*}, where N depends only on δ_2. This proper crossing ensures that $\sum_{W_i \in \mathcal{G}_n^{\mathbb{H}}(V)} |J^sT^n|^t_{C^0(W_i)}$ will be comparable to $\sum_{A \in \mathcal{A}_n^{i_*}} \sup_A |J^sT^n|^t$, and then, adjusting for N, we conclude that $\sum_{W_i \in \mathcal{G}_n^{\mathbb{H}}(W)} |J^sT^n|^t_{C^0(W_i)}$ grows at the rate $Q_n(t)$. ■

Remark 3.4. Proposition 3.3(a) says that the pressure of all long (in the scale δ_1) local stable manifolds grows at a uniform exponential rate (not just asymptotically the same rate).

A corollary of Proposition 3.3(b) is the exact exponential growth of $Q_n(t)$,

$$e^{nP_*(t)} \leq Q_n(t) \leq 2c_2^{-1} e^{nP_*(t)} \quad \forall n \geq 1, \forall t \in [t_0, t_1],$$

where the lower bound follows from the submultiplicativity of $Q_n(t)$ and the upper bound follows from its (approximate) supermultiplicativity. This bound is essential in proving the requisite spectral properties of $\mathcal{L}_t$ in Section 3.2.2.

3.2. Banach spaces adapted to $t \in [t_0, t_1]$

The Banach spaces adapted to the operator $\mathcal{L}_t$ for $t \in (0, t_*)$ are similar to those used in [33] for the case $t = 1$. For convenience, we identify $f \in C^1(M)$ with the measure $d\mu = f d\mu_{\mathrm{SRB}}$. With this identification, the transfer operator defined on distributions μ by

$$\mathcal{L}_t \mu(\psi) = \mu\big(\psi \circ T \cdot (J^sT)^{t-1}\big) \quad \text{for suitable test functions } \psi, \tag{3.1}$$

coincides with the pointwise definition of $\mathcal{L}_t f$ acting on measurable functions from (2.4). As in Section 3.1, we fix $[t_0, t_1] \subset (0, t_*)$ and obtain uniform estimates for $t \in [t_0, t_1]$.

3.2.1. Definition of norms

Since $\mathcal{L}_t f$ has a deregularizing effect in the stable direction, but improves regularity in the unstable direction, the norms defined below have two important properties: they integrate along local stable manifolds to average out the action of $\mathcal{L}_t f$ in the stable direction, while requiring $\mathcal{L}_t f$ to have a form of average regularity in the unstable direction (see the definition of $\| \cdot \|_u$ below). Integrating along local stable manifolds (as opposed to the cone-stable curves used in [32–34]) also allows us to take advantage of the fact that J^sT is Hölder continuous along such manifolds.

Fix $0 < \alpha \le 1/(q+1)$. For $f \in C^1(M)$, define the *weak norm* of f by

$$|f|_w = \sup_{W \in \mathcal{W}^s_{\mathbb{H}}} \sup_{\substack{\psi \in C^\alpha(W) \\ |\psi|_{C^\alpha} \le 1}} \int_W f \psi \, dm_W. \tag{3.2}$$

Define $\mathcal{B}_w$ to be the completion of $C^1(M)$ in the $| \cdot |_w$ norm.

For the strong norm, we need additional parameters. Choose

$$p > q+1 \quad \text{such that } \theta^{t_1 - 1/p} < e^{P_*(t_1)}, \beta \in (1/p, \alpha) \text{ and } \gamma < \min\{1/p, \alpha - \beta\}.$$

Define the *strong stable norm* of f by

$$\|f\|_s = \sup_{W \in \mathcal{W}^s_{\mathbb{H}}} \sup_{\substack{\psi \in C^\beta(W) \\ |\psi|_{C^\beta} \le |W|^{-1/p}}} \int_W f \psi \, dm_W.$$

The strong unstable norm measures the integral of f on two curves that are close together. To define this, we need notions of distance between curves and test functions. Since the stable cone $\mathcal{C}^s$ is bounded away from the vertical, we view $W \in \mathcal{W}^s$ as the graph of a function of the r-coordinate over an interval I_W,

$$W := \{G_W(r) \mid r \in I_W\} := \{(r, \varphi_W(r)) \mid r \in I_W\}.$$

Now given $W_1, W_2 \in \mathcal{W}^s$ defined by $\varphi_{W_1}, \varphi_{W_2}$, define

$$d(W_1, W_2) = |I_{W_1} \triangle I_{W_2}| + |\varphi_{W_1} - \varphi_{W_2}|_{C^1(I_{W_1} \cap I_{W_2})},$$

if W_1, W_2 lie in the same homogeneity strip, and $d(W_1, W_2) = \infty$ otherwise. If $d(W_1, W_2) < \infty$, we define a distance between test functions $\psi_k \in C^0(W_k)$ by

$$d_0(\psi_1, \psi_2) = |\psi_1 \circ G_{W_1} - \psi_2 \circ G_{W_2}|_{C^0(I_{W_1} \cap I_{W_2})}.$$

With these definitions, we are able to define the *strong unstable norm* of f as

$$\|f\|_u = \sup_{\varepsilon \le \varepsilon_0} \sup_{\substack{W_1, W_2 \in \mathcal{W}^s_{\mathbb{H}} \\ d(W_1, W_2) \le \varepsilon}} \sup_{\substack{|\psi_i|_{\mathcal{C}^\alpha(W_i)} \le 1 \\ d_0(\psi_1, \psi_2) = 0}} \varepsilon^{-\gamma} \left| \int_{W_1} f \psi_1 \, dm_{W_1} - \int_{W_2} f \psi_2 \, dm_{W_2} \right|,$$

where $\varepsilon_0 > 0$ is a small constant depending on the table. Finally, define $\mathcal{B}$ to be the closure of $C^1(M)$ in the *strong norm* $\| \cdot \|_{\mathcal{B}}$, defined by $\|f\|_{\mathcal{B}} = \|f\|_s + c_u \|f\|_u$, where c_u is chosen so that the inequalities in Theorem 3.7 provide contraction in the strong norm (see [4, SECT. 4.3]).

Remark 3.5. The choices of parameters are motivated as follows: $\alpha \le 1/(q+1)$ due to the Hölder exponent in (2.2). Then $\beta < \alpha$ is required for relative compactness of the unit ball of $\mathcal{B}$ in $\mathcal{B}_w$. The weight $|W|^{-1/p}$ weakens the contraction of the one-step expansion to

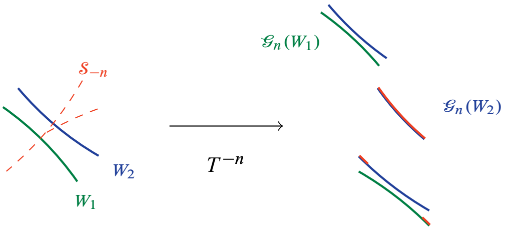


FIGURE 1
Two stable manifolds W_1 (green) and W_2 (blue) and their images under T^{-n}. Green and blue pieces are matched, while red curves are not matched due to cuts introduced by $\mathcal{S}_{-n}$.

$\theta^{t-1/p}$, so p is chosen large enough that this is still small compared to the pressure $e^{P_*(t)}$. Finally, the regularity exponent γ is chosen sufficiently small that the unmatched pieces created by discontinuities in the Lasota–Yorke inequalities are rendered negligible by the weight $|W|^{1/p}$.

Proposition 3.6. *With the correct choices of parameters above, we have a sequence of continuous inclusions, $C^1(M) \subset \mathcal{B} \subset \mathcal{B}_w \subset (C^\alpha(M))^*$.*

Moreover, the embedding of the unit ball of $\mathcal{B}$ into $\mathcal{B}_w$ is compact.

3.2.2. A spectral gap for $\mathcal{L}_t$

Since $\mathcal{B}$ is defined as the completion of $C^1(M)$ in $\|\cdot\|_{\mathcal{B}}$, a priori it is not clear that $\mathcal{L}_t$ acts continuously on $\mathcal{B}$ since J^sT is not even piecewise Hölder continuous; however, [27, THEOREM 5.66] and [4, LEMMA 4.10] show that J^sT varies sufficiently regularly on hyperbolic Cantor rectangles so that if $f \in C^1(M)$, then $\mathcal{L}_t f$ can be approximated by C^1 functions in the $\|\cdot\|_{\mathcal{B}}$-norm, i.e., $\mathcal{L}_t f \in \mathcal{B}$. Thus we are able to prove:

Theorem 3.7 ([4]). *Operator $\mathcal{L}_t$ acts continuously on $\mathcal{B}$ and satisfies the following Lasota–Yorke (or Doeblin–Fortet) inequalities: there exist $C, C_n > 0$ such that for all $f \in \mathcal{B}$, $n \geq 0$,*

$$\begin{aligned}
\left|\mathcal{L}_t^n f\right|_w &\leq C Q_n(t)|f|_w, \\
\left\|\mathcal{L}_t^n f\right\|_s &\leq C\left(\Lambda^{-(\beta-1/p)n} Q_n(t) + \theta^{(t-1/p)n}\right)\|f\|_s + C_n|f|_w, \\
\left\|\mathcal{L}_t^n f\right\|_u &\leq C Q_n(t)\left(n^\gamma \Lambda^{-\gamma n}\|f\|_u + C_n\|f\|_s\right).
\end{aligned}$$

Furthermore, $\mathcal{L}_t$ has a spectral gap: $e^{P_*(t)}$ is the eigenvalue of maximum modulus, it is simple, and the rest of the spectrum of $\mathcal{L}_t$ is contained in a disk of radius $\sigma e^{P_*(t)}$, where $\sigma < 1$ is uniform for $t \in [t_0, t_1]$.

Comments on the proof. (1) *The estimate on unmatched pieces.* For a proof of the Lasota–Yorke inequalities, the reader is referred to [4, SECT. 4.3]. Here we comment only on the control of "unmatched pieces" in the estimate of the strong unstable norm since this leads to essential changes in the case $t = 0$. We must estimate $|\int_{W_1} \mathcal{L}_t^n f \psi_1 - \int_{W_2} \mathcal{L}_t^n f \psi_2|$.

Changing variables, we see that $\mathcal{G}_n^{\mathbb{H}}(W_1)$ comprises matched pieces (that are close to a corresponding curve in $\mathcal{G}_n^{\mathbb{H}}(W_2)$), and unmatched pieces (which are not, due to cuts by the singularity set $\mathcal{S}_{-n}$). See Figure 1. The distance between matched pieces contracts due to the hyperbolicity of T, but the unmatched pieces do not contract. Yet, unmatched pieces have length at most $\Lambda^{-j}\varepsilon$ if they are cut by a singularity curve at time $-j$, so we may use the strong stable norm to estimate

$$\int_{W_i} \mathcal{L}_t^n f\psi = \int_{V_j} \mathcal{L}_t^{n-j} f\psi \circ T^j \left|J_{V_j}T^j\right|^t \leq \Lambda^{-j/p}\varepsilon^{1/p}\left\|\mathcal{L}_t^{n-j} f\right\|_s \left|J_{V_j}T^j\right|_{C^0}^t$$

In this sense, $\|\cdot\|_s$ acts as a “weak norm” for $\|\cdot\|_u$ to control unmatched pieces. This is the reason why the weight $|W|^{-1/p}$ must be included in the definition of $\|\cdot\|_s$, and is an essential difference with the case $t = 0$ in Section 4.2.

(2) *Quasicompactness of $\mathcal{L}_t$.* The Lasota–Yorke inequalities imply that the spectral radius of $\mathcal{L}_t$ on $\mathcal{B}$ is at most $e^{P_*(t)}$ and its essential spectral radius $< e^{P_*(t)}$ if $\theta^t < e^{P_*(t)}$. This is the pressure gap condition guaranteed by choice of θ for all $t \in [t_0, t_1]$. In order to conclude quasicompactness, however, we need a *lower bound* on the spectral radius. This follows from Proposition 3.3(a). Indeed, let $W \in \mathcal{W}_{\mathbb{H}}^s$ with $|W| \geq \delta_1/3$, and choose $\psi \equiv 1$. For any $n \geq 1$,

$$\begin{aligned}\int_W \mathcal{L}_t^n 1 &= \sum_{W_i \in \mathcal{G}_n^{\mathbb{H}}(W)} \int_{W_i} \left|J^sT^n\right|^t \geq e^{-C_d}\frac{\delta_1}{3} \sum_{W_i \in L_n^{\delta_1}(W)} \left|J^sT^n\right|_{C^0(W_i)}^t \geq Cc_1 Q_n(t) \\ &\geq C'e^{nP_*(t)}.\end{aligned}$$

Thus $\|\mathcal{L}^n 1\|_s \geq C'e^{nP_*(t)}$, and so the spectral radius of $\mathcal{L}$ is $e^{P_*(t)}$.

(3) *A spectral gap for $\mathcal{L}_t$.* Exact exponential growth of $Q_n(t)$ (see Remark 3.4) implies $\|\mathcal{L}_t^n\|_{\mathcal{B}} \leq CQ_n(t) \leq C'e^{nP_*(t)}$, so that the peripheral spectrum of $\mathcal{L}_t$ has no Jordan blocks. Then the expression

$$\nu_t(\psi) := \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} e^{-kP_*(t)}\mathcal{L}_t 1(\psi) \tag{3.3}$$

defines a finite Borel measure in $\mathcal{B}$ satisfying $\mathcal{L}_t\nu_t = e^{P_*(t)}\nu_t$, where, according to our identification of functions with densities with respect to μ_{SRB}, we set

$$\mathcal{L}_t 1(\psi) := \int_M \psi\mathcal{L}_t 1 \, d\mu_{\text{SRB}}. \tag{3.4}$$

Using (3.3) and the uniform control provided by Proposition 3.3, one shows that all eigenvectors corresponding to the peripheral spectrum are measures absolutely continuous with respect to ν_t and all eigenvalues are roots of unity. Finally, the topological mixing of T implies that there can be no other eigenvalues of modulus $e^{P_*(t)}$. ■

3.2.3. An equilibrium state and a variational principle

Let ν_t be defined as in (3.3) and let $\tilde{\nu}_t \in \mathcal{B}^*$ denote the analogous construction with the dual operator $\mathcal{L}_t^*$. Define

$$\mu_t(\psi) = \frac{\langle \nu_t, \psi\tilde{\nu}_t\rangle}{\langle \nu_t, \tilde{\nu}_t\rangle}, \quad \psi \in C^\alpha(M).$$

The normalization $\langle \nu_t, \tilde{\nu}_t \rangle \neq 0$ by Proposition 3.3(a). Then it is a standard calculation that μ_t is an invariant probability measure for T, and due to the spectral gap of $\mathcal{L}_t$, μ_t enjoys exponential decay of correlations against Hölder observables.

The facts that μ_t has no atoms, gives 0 weight to any C^1 curve, is positive on open sets, and has stable and unstable manifolds of positive length all follow from the regularity of $\nu_t \in \mathcal{B}$.

Finally, we comment on the entropy of μ_t and conclude the variational principle stated in Theorem 2.7. To this end, define the Bowen balls for T^{-n} by

$$B(x,n,\varepsilon) = \big\{ y \in M : d(T^{-i}x, T^{-i}y) \leq \varepsilon, \forall i \in [0,n] \big\}. \tag{3.5}$$

Proposition 3.8 (Measure of Bowen balls). *There exists $C > 0$ such that for all $x \in M$, $n \geq 1$, and $y \in B(x,n,\varepsilon)$,*

$$\mu_t\big(B(x,n,\varepsilon)\big) \leq Ce^{-nP_*(t)+t\log J^sT^n(T^{-n}y)}.$$

Then [13, MAIN THEOREM] implies that for μ_t-a.e. $x \in M$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} -\frac{1}{n} \log \mu_t\big(B(x,n,\varepsilon)\big) = h_{\mu_t}(T).$$

This, together with Proposition 3.8, implies

$$h_{\mu_t}(T) \geq P_*(t) - t \int \log J^sT \, d\mu_t = P_*(t) + t \int \log J^uT \, d\mu_t.$$

But $P_*(t) \geq h_{\mu_t}(T) - t \int \log J^uT \, d\mu_t$ since $P_*(t) \geq P(t)$ by Theorem 2.5. We conclude that $P_*(t) = h_{\mu_t}(T) - t \int \log J^uT \, d\mu_t = P(t)$, which implies both that μ_t is the measure maximizing the pressure and that the topological pressure satisfies a variational principle, despite the effect of singularities.

The last item of Theorem 2.7, the uniqueness of μ_t, uses the concept of a tangent measure. The argument exploits in particular the differentiability of the pressure for $t > 0$. We refer the interested reader to [4, SECT. 5.5].

4. IDEAS FROM THE PROOF OF THEOREM 2.3

In this section, we provide a parallel presentation to Section 3 for the case $t = 0$, i.e., the construction of the measure of maximal entropy. As before, we divide the ideas into two parts: (1) geometric estimates to control local complexity and a uniform rate of growth for $\#\mathcal{M}_0^n$; (2) a functional-analytic framework needed to construct the equilibrium state μ_0.

In contrast to Section 3, we cannot use homogeneity strips and must drastically alter the weights in the strong norm. These changes are sufficiently severe to prevent us from obtaining a spectral gap for $\mathcal{L}_0$ and exponential mixing for μ_0.

4.1. Complexity and exact exponential growth of $\#\mathcal{M}_0^n$

Recall the toy calculation in Section 2.4.1 for deriving the correct weight for the topological pressure. If we consider (2.5) with $t = 0$, we have $\#\mathcal{G}_n^{\mathbb{H}}(W) = \infty$ whenever

$T^{-n}W$ crosses infinitely many homogeneity strips. Thus we cannot use homogeneity strips when studying the case $t = 0$. Moreover, the one-step expansion Lemma 3.1 does not hold.

Instead, we use the linear complexity bound due to Bunimovich. For $x \in M$, let $N(\mathcal{S}_n, x)$ denote the number of singularity curves in $\mathcal{S}_n$ that meet at x. Define $N(\mathcal{S}_n) = \sup_{x\in M} N(\mathcal{S}_n, x)$.

Lemma 4.1 ([15]). *Assume finite horizon. There exists $K > 0$, depending only on the configuration of scatterers, such that $N(\mathcal{S}_n) \le Kn$ for all $n \ge 1$.*

Sketch of proof from [28]. Suppose $x, x' \in M$ lie on a straight billiard trajectory with one or more tangential collisions between them. Let A, A' be neighborhoods of x, x' in M, partitioned into sectors $A_1, \dots A_k \subset A$ and $A'_1, \dots A'_k \subset A$ such that $T^{n_j} A_j = A'_j$. Define $\hat{T}|_{A_j} := T^{n_j}$. See Figure 2.

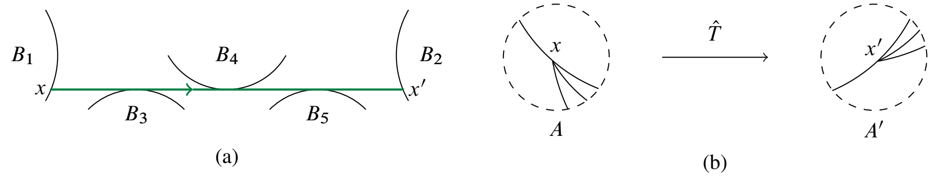


FIGURE 2
(a) A trajectory with multiple tangencies. (b) Neighborhood A of x with elements of $\mathcal{S}_n$ and neighborhood A' of x' with their images in $\mathcal{S}_{-n}$.

We prove the statement by induction. For $n = 1$ it is trivial. Now assume $N(\mathcal{S}_{n-1}) \le K(n-1)$ for some $K > 0$. Let $N(\mathcal{S}_i | A'_j, x')$ denote the number of curves in $\mathcal{S}_i$ passing through x' and lying in A'_j. Since curves in $\mathcal{S}_i \setminus \mathcal{S}_0$ (stable) and $\mathcal{S}_{-n} \setminus \mathcal{S}_0$ (unstable) are uniformly transverse, each sector created by $\mathcal{S}_i$ can only intersect one sector created by $\mathcal{S}_{-n}$. Then pulling back the picture from x' to x and recalling that k is the number of tangencies meeting at x, we have (here we are using continuity of the flow)

$$N(\mathcal{S}_n, x) \le k + \sum_j N(\mathcal{S}_{n-n_j} | A'_j, x') \le k + \sum_j N(\mathcal{S}_{n-1} | A'_j, x'),$$

and, using the inductive assumption on $n-1$, this yields $N(\mathcal{S}_n, x) \le k + K(n-1)$, which is less than Kn if $k \le K$. Due to the finite horizon condition, the number of tangencies intersecting at a point $x \in M$ has a finite upper bound depending only on the table. Thus choosing K to be this upper bound completes the proof of the lemma. ■

4.1.1. Fragmentation lemmas

Choose $n_0 \in \mathbb{N}$ such that $n_0^{-1} \log(Kn_0 + 1) < h_*$. Due to Lemma 4.1, we may choose $\delta_0 > 0$ such that any stable curve of length $\le \delta_0$ is cut into at most $Kn_0 + 1$ pieces by $\mathcal{S}_{-n_0}$. We use this choice of δ_0 in our definition of $\mathcal{W}^s$, the set of local stable manifolds with which we work. (In Section 4.2.2 we will shrink δ_0 further depending on the parameters in our norms.) This choice of δ_0 will ensure that the growth in $\mathcal{G}_n(W)$ due to local complexity will be slower than e^{nh_*}. We make this precise below.

For $\delta \le \delta_0$, let $\mathcal{G}_n^\delta(W)$ denote the collection of curves in $T^{-n}W$ analogous to $\mathcal{G}_n^{\mathbb{H}}(W)$, but without using homogeneity strips, and with pieces longer than δ subdivided into curves between length $\delta/2$ and δ at each step. Define $L_n^\delta(W) = \{W_i \in \mathcal{G}_n^\delta(W) : |W| \ge \delta/3\}$ and $Sh_n^\delta(W) = \mathcal{G}_n^\delta(W) \setminus L_n^\delta(W)$.

Lemma 4.2 ([3]). *For all $\varepsilon > 0$, there exist $n_1, \delta > 0$ such that, for all $n \ge n_1$,*

$$\#Sh_n^\delta(W) \le \varepsilon \#\mathcal{G}_n^\delta(W) \quad \textit{for all } W \in \hat{\mathcal{W}}^s \textit{ with } |W| \ge \delta/3.$$

Idea of proof. Recalling (2.1), choose $\varepsilon > 0$ and n_1 such that $3C_0^{-1}(Kn_1 + 1)\Lambda^{-n_1} < \varepsilon$. Choose $\delta > 0$ such that if $|W| < \delta$ then $T^{-n_1}W$ comprises at most $Kn_1 + 1$ connected components of length at most δ_0. Then $Sh_{n_1}^\delta(W)$ contains at most $Kn_1 + 1$ elements. On the other hand, $|T^{-n_1}W| \ge C_0\Lambda^{n_1}\delta/3$, where $\Lambda = 1 + 2\mathcal{K}_{\min}\tau_{\min}$ is from (2.1). Thus $\#\mathcal{G}_{n_1}^\delta(W) \ge C_0\Lambda^{n_1}/3$, and so $\#Sh_{n_1}^\delta(W) \le \varepsilon\#\mathcal{G}_{n_1}^\delta(W)$ by the choice of n_1.

The argument can be iterated, grouping each collection of pieces at time kn_1 by the most recent time jn_1, $j \le k$, that each piece was contained in an element of $L_{jn_1}^\delta(W)$. ■

As in Lemma 3.2(b), some control of short pieces can also be extended to elements of $\mathcal{M}_0^n$ and $\mathcal{M}_{-n}^0$. Let $\delta_1, n_1 \ge n_0$ correspond to $\varepsilon = 1/4$ in Lemma 4.2. Define

$$L_s(\mathcal{M}_0^n) = \{A \in \mathcal{M}_0^n : \operatorname{diam}^s(A) \ge \delta_1/3\} \quad \text{and}$$
$$L_u(\mathcal{M}_{-n}^0) = \{B \in \mathcal{M}_{-n}^0 : \operatorname{diam}^u(B) \ge \delta_1/3\}.$$

Lemma 4.3 ([3]). *There exists $c_0 > 0$ such that, for all $n \ge 1$,*

$$\#L_s(\mathcal{M}_0^n) \ge c_0\delta_1\#\mathcal{M}_0^n \quad \textit{and} \quad \#L_u(\mathcal{M}_{-n}^0) \ge c_0\delta_1\#\mathcal{M}_{-n}^0.$$

4.1.2. Uniform bounds on growth

As in Section 3.1, the fragmentation lemmas above imply uniform bounds on the growth of $\#\mathcal{G}_n(W)$ and $\#\mathcal{M}_0^n$.

Proposition 4.4. (a) *There exists $c_1 > 0$ such that, for any $W \in \mathcal{W}^s$ with $|W| \ge \delta_1/3$,*

$$\#\mathcal{G}_n(W) \ge c_1\#\mathcal{M}_0^n \quad \forall n \ge 1.$$

(b) *There exists $c_2 > 0$ such that for all $k, n \ge 1$,*

$$\#\mathcal{M}_0^{n+k} \ge c_2\#\mathcal{M}_0^n \cdot \#\mathcal{M}_0^k.$$

Idea of proof. Claim (b) follows from (a) and Lemma 4.2 since $\#\mathcal{M}_0^{n+k} \ge 2\delta_0^{-1}\#\mathcal{G}_{n+k}(W)$ and

$$\#\mathcal{G}_{n+k}(W) \ge \sum_{V_j \in L_n^{\delta_1}(W)} \#\mathcal{G}_k(V_j) \ge \#L_n^{\delta_1}(W)c_1\#\mathcal{M}_0^k \ge \frac{3c_1}{4}\#\mathcal{G}_n^{\delta_1}(W)\#\mathcal{M}_0^k \ge \frac{3c_1^2}{4}\#\mathcal{M}_0^n\#\mathcal{M}_0^k.$$

The proof of (a) follows the same lines as the proof of Proposition 3.3(a), covering M with a finite number N_{δ_1} of Cantor rectangles depending on the length scale δ_1. Then

Lemma 4.3 implies at least one of these rectangles, R_{i_*}, is fully crossed (in the unstable direction) by at least $\frac{c_0\delta_1}{N_{\delta_1}}\#\mathcal{M}_{-n}^0$ "long" elements of $\mathcal{M}_{-n}^0$. Any $W \in \mathcal{W}^s$ of length at least $\delta_1/3$ crosses one rectangle R_j. Then there exists N, depending only on δ_1, such that $T^{-N}W$ properly crosses R_{i_*} in the stable direction. Thus $T^{-N-n}W$ intersects at least $\frac{c_0\delta_1}{N_{\delta_1}}\#\mathcal{M}_0^n$ elements of $\mathcal{M}_0^n$ since $\#\mathcal{M}_0^n = \#\mathcal{M}_{-n}^0$. Adjusting for N (which only affects c_1) proves (a). ■

Remark 4.5. Proposition 4.4(b) implies the exact exponential growth of $\#\mathcal{M}_0^n$,

$$e^{nh_*} \le \#\mathcal{M}_0^n \le 2c_2^{-1}e^{nh_*} \quad \text{for all } n \ge 1.$$

As in Section 3.2.2, this will be essential to controlling the peripheral spectrum of $\mathcal{L}_0$.

A second corollary of our uniform bounds is the uniform growth rate of $|T^{-n}W|$ in terms of the topological entropy h_*, i.e., there exists $C > 0$ such that, for all $W \in \mathcal{W}^s$ with $|W| \ge \delta_1/3$,

$$Ce^{nh_*} \le \left|T^{-n}W\right| \le C^{-1}e^{nh_*} \quad \text{for all } n \ge n_1.$$

This is precisely the rate of growth one sees in smooth hyperbolic systems, despite the fact that in this context h_* also counts cuts due to discontinuities.

To prove this bound, the previous remark, together with Proposition 4.4(a), gives $Ce^{nh_*} \le \#\mathcal{G}_n(W) \le C^{-1}e^{nh_*}$. But $|T^{-n}W| \le \delta_0\#\mathcal{G}_n(W)$ since curves in $\mathcal{G}_n(W)$ have length at most δ_0, proving the upper bound. Finally, the lower bound follows from Lemma 4.2 with $\varepsilon = 1/4$,

$$\left|T^{-n}W\right| = \sum_{W_i \in \mathcal{G}_n^{\delta_1}} |W_i| \ge \frac{\delta_1}{3}\#L_n^{\delta_1}(W) \ge \frac{\delta_1}{4}\#\mathcal{G}_n^{\delta_1}(W).$$

4.2. Banach spaces adapted to $t = 0$

We define $\mathcal{L}_0$ acting on functions as in (2.4) and on distributions as in (3.1).

Unfortunately, the Hölder weight $|W|^{1/p}$ in the strong stable norm from Section 3.2.1 is disastrous when $t = 0$. This is because if $W \in \mathcal{W}^s$ and $T^{-1}W$ has a single component near a tangential collision so that $|T^{-1}W| \sim |W|^{1/2}$, then, if $\psi = |W|^{-1/p}$,

$$\int_W \mathcal{L}_0 f\psi = \int_{T^{-1}W} f\psi \circ T \le \|f\|_s \frac{|T^{-1}W|^{1/p}}{|W|^{1/p}} \sim \|f\|_s|W|^{-1/2p}.$$

Taking the supremum over $W \in \mathcal{W}^s$ yields ∞ and hence $\mathcal{L}_0$ is not a bounded operator.

Yet, we cannot abandon the weight entirely due to the need to control the unmatched pieces in the Lasota–Yorke estimates, see Figure 1 and the proof of Theorem 3.7. These considerations force us to adopt a weak logarithmic weight $|\log|W||$ in the definition of $\|\cdot\|_s$, which in turn forces a logarithmic modulus of continuity in $\|\cdot\|_u$. This last change prevents a genuine contraction in the Lasota–Yorke inequality, which prevents us from proving that $\mathcal{L}_0$ is quasicompact with a spectral gap.

Nevertheless, under a sparse recurrence condition to the singular set (4.1), we show that the spectral radius of $\mathcal{L}_0$ on $\mathcal{B}$ is e^{h_*} and we obtain left and right eigenvectors of $\mathcal{L}_0$ as limit points using compactness, from which we construct the measure μ_0 with entropy h_*.

4.2.1. Sparse recurrence to singularities

In order to control the evolution of $\mathcal{L}_0^n$ in the strong norm, we shall need the following condition on the rate of recurrence to the singular set $\mathcal{S}_0$, which corresponds to tangential, or grazing, collisions. Note that our results up until now have not needed this condition.

Choose $n_0 \in \mathbb{N}$ and an angle φ_0 close to $\pi/2$. Let $s_0 \in (0,1)$ be the smallest number such that any orbit of length n_0 has at most $s_0 n_0$ collisions with $|\varphi| \geq \varphi_0$. The finite horizon condition guarantees that we can always choose n_0 and φ_0 so that $s_0 < 1$. Indeed, if there are no trajectories with three consecutive tangencies on the table (a generic condition), then one may choose n_0 and φ_0 so that $s_0 \leq \frac{2}{3}$. Our assumption is the following:

$$h_* > s_0 \log 2. \tag{4.1}$$

The $\log 2$ comes from the fact that if W is a local stable manifold that makes a nearly tangential collision under T^{-1} then $|T^{-1}W| \sim |W|^{1/2}$. Thus our assumption ensures that the growth due to tangential collisions along sufficiently long orbit segments does not exceed the exponential rate of growth given by h_*.

We remark that there is no known table for which the condition $h_* > s_0 \log 2$ fails. Indeed, since $h_* \geq h_{\mu_{\mathrm{SRB}}}$ and $h_{\mu_{\mathrm{SRB}}} = \int \log J^u T \, d\mu_{\mathrm{SRB}}$ by the Pesin entropy formula, it suffices to check that $\chi^+_{\mu_{\mathrm{SRB}}} > s_0 \log 2$, where $\chi^+_{\mu_{\mathrm{SRB}}}$ is the positive Lyapunov exponent of T with respect to μ_{SRB}, in order to conclude that (4.1) holds. Using this criterion, all examples computed numerically in [9] for a triangular lattice and [38] for a rectangular lattice satisfy (4.1). Furthermore, it is possible to prove analytically that (4.1) holds for large open sets of such billiard configurations. See [3, SECT. 2.4] for a more detailed discussion.

4.2.2. Definition of norms

Choose $\alpha, \beta, \gamma > 0$, and $p > 1$ such that

$$\beta < \alpha \leq 1/3, \quad 2^{s_0 p} < e^{h_*}, \quad \gamma < p.$$

Enlarge n_0 so that

$$\frac{1}{n_0} \log(K n_0 + 1) < h_* - p s_0 \log 2,$$

where K is from Lemma 4.1. Choose $\delta_0 > 0$ as in Section 4.1.1 so that any stable manifold of length $\leq \delta_0$ is cut into at most $K n_0 + 1$ pieces by $\mathcal{S}_{-n_0}$.

The weak norm $|\cdot|_w$ and $\mathcal{B}_w$ are defined precisely as in (3.2), so we focus on the strong norm. For $f \in C^1(M)$, define the *strong stable norm* of f by

$$\|f\|_s = \sup_{W \in \mathcal{W}^s} \sup_{\substack{\psi \in \mathcal{C}^\beta(W) \\ |\psi|_{\mathcal{C}^\beta(W)} \leq |\log|W||^p}} \int_W f \psi \, dm_W$$

Recalling the distance between curves $d(W_1, W_2)$ and between test functions $d_0(\psi_1, \psi_2)$ from Section 3.2.1, we define the *strong unstable norm* of f by

$$\|f\|_u = \sup_{\varepsilon \leq \varepsilon_0} \sup_{\substack{W_1, W_2 \in \mathcal{W}^s \\ d(W_1, W_2) \leq \varepsilon}} \sup_{\substack{|\psi_i|_{\mathcal{C}^\alpha(W_i)} \leq 1 \\ d_0(\psi_1, \psi_2) = 0}} |\log \varepsilon|^\gamma \left| \int_{W_1} f \psi_1 - \int_{W_2} f \psi_2 \right|.$$

The *strong norm* of f is defined to be $\|f\|_{\mathcal{B}} = \|f\|_s + \|f\|_u$, and $\mathcal{B}$ is the completion of $C^1(M)$ in the $\|\cdot\|_{\mathcal{B}}$ norm.

4.2.3. Spectrum of $\mathcal{L}_0$ and construction of an invariant measure

Proposition 3.6 still holds true with these new norms and most importantly, the unit ball of $\mathcal{B}$ is compact in $\mathcal{B}_w$. However, due to the logarithmic modulus of continuity in the definition of $\|\cdot\|_u$, the strong unstable norm does not contract. Indeed, recalling Figure 1, when we compare $\int_{W^1} \mathcal{L}_0 f \psi_1 - \int_{W^2} \mathcal{L}_0 f \psi_2$, the matched pieces $W_i^k \in \mathcal{G}_n(W^k)$ have contracted to a distance $d(W_i^1, W_i^2) \le C\Lambda^{-n}\varepsilon$. Yet, the contraction in the norm is given by $\frac{|\log C\Lambda^{-n}\varepsilon|^\gamma}{|\log\varepsilon|^\gamma}$, and taking the supremum over $\varepsilon > 0$ yields 1. The inequalities we can prove are the following.

Proposition 4.6. *Assume $h_* > s_0 \log 2$. There exists $C > 0$ such that, for all $f \in \mathcal{B}$, $n \ge 0$,*

$$\begin{aligned}
&\left|\mathcal{L}^n f\right|_w \le C|f|_w \#\mathcal{M}_0^n,\\
&\left\|\mathcal{L}^n f\right\|_s \le C\left(\sigma^n\|f\|_s + |f|_w\right)\#\mathcal{M}_0^n, \quad \textit{for some } \sigma < 1,\\
&\left\|\mathcal{L}^n f\right\|_u \le C\left(\|f\|_u + \|f\|_s\right)\#\mathcal{M}_0^n.
\end{aligned}$$

Although the bounds of Proposition 4.6 are not sufficient to prove the quasicompactness of $\mathcal{L}_0$ on $\mathcal{B}$, they, together with Proposition 4.4, do provide good control of $\|\mathcal{L}_0^n\|_{\mathcal{B}}$.

Using Remark 4.5, we have $\|\mathcal{L}_0^n\|_{\mathcal{B}} \le Ce^{nh_*}$, for all $n \ge 1$. Moreover, our lower bounds on $\#L_n^{\delta_1}(W)$ from Lemma 4.2 and $\#\mathcal{G}_n(W)$ from Proposition 4.4 imply that

$$\left\|\mathcal{L}_0^n 1\right\|_s \ge \left|\mathcal{L}_0^n 1\right|_w \ge \int_W \mathcal{L}_0^n 1 \ge \sum_{W_i \in L_n^{\delta_1}(W)} |W_i| \ge \frac{\delta_1}{3}\frac{3}{4}\#\mathcal{G}_n^{\delta_1}(W) \ge Ce^{nh_*}. \tag{4.2}$$

These estimates imply not only that the spectral radius of $\mathcal{L}_0$ on $\mathcal{B}$ is e^{h_*}, but also that the sequence $e^{-nh_*}\mathcal{L}_0^n 1$ is uniformly bounded away from 0 and ∞ in the strong norm. We now use this fact to construct an eigenmeasure for $\mathcal{L}_0$ with eigenvalue e^{h_*}.

By the observation above, for $n \ge 1$ the sequence

$$\nu_n = \frac{1}{n}\sum_{k=0}^{n-1} e^{-kh_*}\mathcal{L}_0^k 1 \quad \text{is uniformly bounded in } \mathcal{B}.$$

Since any ball of finite size in $\mathcal{B}$ is compact in $\mathcal{B}_w$, a subsequence converges in $\mathcal{B}_w$. Let $\nu_0 \in \mathcal{B}_w$ be a limit point of ν_n. A priori, ν_0 is only a distribution; yet, recalling (3.4), the calculation

$$\left|\nu_0(\psi)\right| \le \lim_{j\to\infty}\frac{1}{n_j}\sum_{k=0}^{n_j-1} e^{-kh_*}\left|\mathcal{L}_0^k 1(\psi)\right| \le |\psi|_\infty \nu_0(1)$$

shows that, indeed, ν_0 can be extended as a bounded operator on continuous functions, i.e., ν_0 is a measure and, indeed, a nonnegative measure since the ν_n are nonnegative. A similar calculation shows that $\mathcal{L}_0\nu_0 = e^{h_*}\nu_0$.

Similarly, let $\tilde{\nu}_0 \in (\mathcal{B}_w)^*$ be a limit point of the sequence

$$\frac{1}{n}\sum_{k=0}^{n-1} e^{-kh_*}(\mathcal{L}_0^*)^k(d\mu_{\mathrm{SRB}}),$$

which is again a measure. Define the pairing

$$\mu_0(\psi) = \frac{\langle \nu_0, \psi \tilde{\nu}_0 \rangle}{\langle \nu_0, \tilde{\nu}_0 \rangle}, \quad \text{for } \psi \in C^1(M).$$

Since $\mathcal{L}_0 \nu_0 = e^{h_*} \nu_0$ and $\mathcal{L}_0^* \tilde{\nu}_0 = e^{h_*} \tilde{\nu}_0$, it is a standard calculation that $\mu_0(\psi \circ T) = \mu_0(\psi)$, i.e., μ_0 is an invariant probability measure for T. We remark that, by definition of ν_0 and $\tilde{\nu}_0$, the normalization $\langle \nu_0, \tilde{\nu}_0 \rangle$ can be computed as the average of the terms $e^{-kh_*} \int_M \mathcal{L}_0^k 1 \, d\mu_{\text{SRB}}$. Thus the fact that $\langle \nu_0, \tilde{\nu}_0 \rangle \neq 0$ follows from the lower bound (4.2) (see [3, PROOF OF PROP. 7.1]).

4.2.4. Properties of μ_0

The key observation for proving all subsequent properties of μ_0 is that, although $\nu_0 \in \mathcal{B}_w$, it inherits stronger regularity as a limit point of the sequence $(\nu_n)_{n \in \mathbb{N}}$, which is uniformly bounded in the $\| \cdot \|_{\mathcal{B}}$-norm. In particular, the convergence of (ν_{n_j}) to ν in $\mathcal{B}_w$ implies

$$\lim_{j \to \infty} \sup_{W \in \mathcal{W}^s} \sup_{|\psi|_{C^\alpha(W)} \leq 1} \left(\int_W \nu \psi \, dm_W - \int_W \nu_{n_j} \psi \, dm_W \right) = 0,$$

and, since $\|\nu_{n_j}\|_u \leq C$ for some $C > 0$, we conclude that $\|\nu\|_u \leq C$ as well. Similarly, $\int_W \nu \leq C |\log |W||^{-p}$ from the uniform bound on $\|\nu_{n_j}\|_s$. This regularity then opens the door to a host of properties for μ_0.

(1) Hyperbolicity. For $k \in \mathbb{Z}$, $\varepsilon > 0$, letting $\mathcal{N}_\varepsilon(\mathcal{S}_k)$ denote the ε-neighborhood of $\mathcal{S}_k$ in M, the strong norm bound implies that there exists $C_k > 0$ such that

$$\nu_0\big(\mathcal{N}_\varepsilon(\mathcal{S}_k)\big) \leq C_k |\log \varepsilon|^{-p} \quad \text{and} \quad \mu_0\big(\mathcal{N}_\varepsilon(\mathcal{S}_k)\big) \leq C_k |\log \varepsilon|^{-p}. \tag{4.3}$$

This implies in turn that μ_0 is T-adapted, i.e., $\int_M -\log d(x, \mathcal{S}_{\pm 1}) \, d\mu_0(x) < \infty$, and that μ_0-a.e. $x \in M$ has a stable and unstable manifold of positive length. The same is true for ν_0.

(2) Ergodicity. Since μ_0 is hyperbolic, we may cover a full measure set of M with Cantor rectangles comprising intersections of stable and unstable manifolds, and study the properties of μ_0 on each rectangle. In particular, the fact that $\|\nu_0\|_u < \infty$ allows us to prove the following (but note that μ_0 itself is singular with respect to Lebesgue measure).

Lemma 4.7 (Absolute continuity of holonomy). *On each Cantor rectangle R, the holonomy map sliding along unstable manifolds in R is absolutely continuous with respect to the conditional measures of μ_0 on stable manifolds.*

Using a Hopf argument and the above lemma, we show that each Cantor rectangle R belongs to one ergodic component. Then since T is topologically mixing, we can force images of rectangles to overlap and thus conclude that (T^n, μ_0) is ergodic for all n.

(3) Mixing and Bernoulli property. The local product structure of the Cantor rectangles, together with a global argument showing that a full measure set of points on each component

of M can be connected by a network of stable and unstable manifolds, enables us to prove that (T, μ_*) is K-mixing,[5] following techniques of Pesin [48,49].

Then adapting the approach of [26] (carried out there for μ_{SRB}), which uses all the properties we have established thus far: K-mixing, hyperbolicity, the absolute continuity of Lemma 4.7, and our bounds on $\mu_0(\mathcal{N}_\varepsilon(\mathcal{S}_{\pm 1}))$, we prove that the partition $\mathcal{M}_{-1}^1$ is *very weakly Bernoulli*. Since $\bigvee_{n=-\infty}^{\infty} T^{-n}(\mathcal{M}_{-1}^1)$ generates the full σ-algebra for T, this implies by [47] that (T, μ_0) is Bernoulli.

(4) Entropy of μ_0. For $x \in M$, define the ε-Bowen ball for T^{-n} as in (3.5). Using the fact that ν_0 scales by e^{nh_*} under a change of variables, we are able to prove:

Proposition 4.8 (Measure of Bowen balls). *There exists $C > 0$ such that, for all $x \in M$ and $n \geq 1$,*

$$\mu_0\big(B(x, n, \varepsilon)\big) \leq Ce^{-nh_*}.$$

As in Section 3.2.3, [13, MAIN THEOREM] implies that for μ_0-a.e. $x \in M$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} -\frac{1}{n} \log \mu_0\big(B(x, n, \varepsilon)\big) = h_{\mu_0}(T^{-1}) = h_{\mu_0}(T).$$

This, together with Proposition 4.8, implies $h_{\mu_0}(T) \geq h_*$. But $h_* \geq h_\mu(T)$ for all T-invariant probability measures as stated in Section 2.3. We conclude that $h_* = h_{\mu_0}(T)$, so μ_0 has maximal entropy.

4.2.5. Uniqueness of μ_0

Finally, we discuss the proof of uniqueness of the measure of maximal entropy from Theorem 2.3. This is essentially a modification of the classical Bowen argument, which uses a uniform lower bound on the measure of Bowen balls,

$$\forall \varepsilon > 0, \quad \exists C > 0 \text{ such that for } \mu_0\text{-a.e. } x \in M, \mu_*(B(x, n, \varepsilon)) \geq Ce^{-nh_*}$$

(see, for example, [43, SECT. 20.3]).

Unfortunately, this lower bound fails for billiards due to the rate of approach of typical points to the singularity set. We can prove, rather, that $\forall \eta > 0$ and μ_0-a.e. $x \in M$,

$$\exists C = C(\eta, x) > 0 \quad \text{such that } \mu_0(B(x, n, \varepsilon)) \geq Ce^{-n(h_* + \eta)}. \tag{4.4}$$

But even this arbitrarily small error in the exponent is not sufficient for the Bowen argument. Instead, we prove a version of the lower bound that "most" $x \in M$ "often" belong to an element of $\mathcal{M}_0^j$ satisfying good lower bounds.

To make this precise, let $\bar{n} \in \mathbb{N}$ be such that $(K\bar{n} + 1)^{1/\bar{n}} < e^{h_*/2}$. Then using Lemma 4.2, choose $\delta_2 > 0$ such that if $A \in \mathcal{M}_{-k}^n$ satisfies

$$\max\big\{\operatorname{diam}^u(A), \operatorname{diam}^s(A)\big\} \leq \delta_2,$$

[5] If $\mathcal{A}$ denotes the Borel sigma-algebra on M, then K-mixing means that there exists a sub-sigma algebra $K \subset \mathcal{A}$ such that (1) $K \subset TK$; (2) $\bigvee_{n=0}^{\infty} T^n K = \mathcal{A}$; (3) $\bigcap_{n=0}^{\infty} T^{-n} K = \{X, \emptyset\}$.

then $A \setminus \mathcal{S}_{\pm\bar{n}}$ consist of at most $K\bar{n}+1$ connected components. Define

$$Sh_0^{2n} := \left\{A \in \mathcal{M}_0^{2n} : \forall j, 0 \le j \le n/2, T^j A \subset E \in \mathcal{M}_0^{2n-j} \text{ such that } \operatorname{diam}^s(E) < \delta_2\right\},$$

with a similar definition for Sh_{-2n}^0 with $\operatorname{diam}^u(E)$ replacing $\operatorname{diam}^s(E)$. These are the "persistently short" elements of $\mathcal{M}_0^{2n}$ and $\mathcal{M}_{-2n}^0$, respectively, which have not belonged to a "long" element within the past $n/2$ iterates.

The next lemma demonstrates that persistently short pieces make up a small proportion of $\mathcal{M}_0^{2n}$ and that long elements that also have long images satisfy strong lower bounds.

Lemma 4.9. (a) *Let* $B_{2n} = \{A \in \mathcal{M}_0^{2n}$: *either* $A \in Sh_0^{2n}$ *or* $T^{2n}A \in Sh_{-2n}^0\}$. *There exists* $C > 0$ *such that, for all* $n \ge 1$, $\#B_{2n} \le Ce^{7nh_*/4}$.

(b) *For all* $k \ge 1$, *if* $E \in \mathcal{M}_0^k$ *with* $\operatorname{diam}^s(E) \ge \delta_2$ *and* $\operatorname{diam}^u(T^kE) \ge \delta_2$,

$$\textit{then} \quad \mu_0(E) \ge C_{\delta_2} e^{-kh_*}, \quad \textit{for some } C_{\delta_2} > 0.$$

Some comments on the proof. Claim (a) follows by iterating the complexity bound given by Lemma 4.2, using the fact that, by the choice of $\bar{n}$ and δ_2, persistently short pieces cannot grow at a rate faster than $(K\bar{n}+1)^{n/(2\bar{n})} < e^{nh_*/4}$ over the most recent $n/2$ iterates.

Claim (b) rests on the fact that if E is long in the stable direction and T^kE is long in the unstable direction, then both E and T^kE cross Cantor rectangles of a fixed size, depending on δ_2. Then the lower bound (4.2) is used to derive (b). ■

The importance of Lemma 4.9 lies in the fact that if $A \in G_{2n} := \mathcal{M}_0^{2n} \setminus B_{2n}$, then there exists $j, k \le n/2$ such that $T^jA \subset E \in \mathcal{M}_0^{2n-j-k}$ and E satisfies Lemma 4.9(b), i.e., $\mu_0(E) \ge C_{\delta_2}e^{-(j+k)h_*}$. Thus, apart from a set of "bad" elements B_{2n} whose size is relatively small, most elements of $\mathcal{M}_0^{2n}$ belong to the "good" set G_{2n} and are contained in a larger set that has good lower bounds. This, together with a time shift to group elements of $\mathcal{M}_0^{2n}$ according to their good counterparts in $\mathcal{M}_0^{2n-j-k}$, is sufficient to adapt the Bowen argument for uniqueness. The reader interested in more details is referred to [3, SECT. 7].

5. OPEN QUESTIONS

We conclude by formulating several open questions relating to the family of geometric potentials we have discussed.

(1) Is $\mu_0 = \mu_{\mathrm{SRB}}$, *or, more generally, is* $\mu_s = \mu_t$ *for* $s \neq t$*?* If there exist $s, t > 0$, $s \neq t$, such that $\mu_s = \mu_t$, then Theorem 2.8 implies that $P(t)$ is affine on $(0, t_*)$, so that μ_{SRB} would be the equilibrium state for all $t \in (0, t_*)$, and assuming the sparse recurrence condition (4.1), $\mu_{\mathrm{SRB}} = \mu_0$ as well.

This seems highly unlikely. Indeed, suppose that z is a periodic orbit with no grazing collisions and let χ_z^+ be its positive Lyapunov exponent. Then our estimates on Bowen balls such as Proposition 4.8 and (4.4), in addition to analogous ones for μ_{SRB}, imply that $h_* = \chi_z^+$ [3, PROP. 7.13]. Thus if we can find two periodic orbits with different Lyapunov exponents, we can conclude that $\mu_0 \neq \mu_{\mathrm{SRB}}$, and in turn $\mu_s \neq \mu_t$ for all $s \neq t$. There are no known Sinai billiard tables in which all periodic orbits have the same Lyapunov exponent, yet it is not proved that this cannot happen. See [35, SECT. 4.4] for a related class of models (a type of open billiard) for which such anomalous behavior has been effectively ruled out.

(2) Can one establish a rate of mixing for μ_0? While exponential mixing for μ_t, $t \in (0, t_*)$, follows from the spectral gap for $\mathcal{L}_t$, no such gap is available for $\mathcal{L}_0$. The question arises whether this is a consequence of the technique or whether there is a genuine failure of exponential mixing at $t = 0$. The fact that the pressure $P(t)$ is finite for $t \geq 0$ and infinite for $t < 0$ whenever there is a periodic orbit with a grazing collision suggests that there is, indeed, a phase transition at $t = 0$, so a loss of exponential mixing would not be out of place. On the other hand, for many expanding systems, the measure of maximal entropy has a faster rate of mixing than the SRB measure, not a slower one.

(3) What are other limit theorems and properties of μ_0? Once a rate of mixing has been established, other limits theorems might follow, for example, a dynamical Central Limit Theorem, which generally requires a summable rate of decay of correlations. Other limit theorems might include invariance principles or large deviation estimates. These are all available for μ_t, $t > 0$, by spectral techniques (see [39] or [32, SECT. 6]), but not for μ_0 at this time.

(4) Can one find a finite horizon Sinai billiard table such that the sparse recurrence condition fails? In other words, can one find a table with $h_* \leq s_0 \log 2$? If so, does a measure of maximal entropy still exist and is it T-adapted, i.e., does it satisfy bounds of the form (4.3)?

(5) Does $\mu_t \to \mu_0$ as $t \to 0$? For $t \in (0, t_*)$, continuity of μ_t and differentiability of $P(t)$ follow from perturbation theory. Assuming the sparse recurrence condition (4.1), Baladi and Demers [4, PROP. 5.5] prove that $\lim_{t \downarrow 0} P(t) = P(0) = h_*$, yet the question of whether the equilibrium states converge remains open.

(6) Is $P(t)$ analytic for all $t > 0$ or is there a phase transition at some $t_\star > 1$? If so, how does $t_\star$ depend on the configuration of scatterers? It is clear from the definition of t_* that it is not an optimal condition for most billiard tables since the hyperbolicity constant $\Lambda = 1 + 2\mathcal{K}_{\min}\tau_{\min}$ is a lower bound, which in general may not be attained along most or even all orbits.

A more refined attempt would be to define $\chi_{\min} \geq \Lambda$ to be the minimal positive Lyapunov exponent over all periodic orbits. Then we could define

$$t_\star = \sup\{t > 0 : P(t) > -t\chi_{\min}\},$$

and try to show that the spectral techniques described here go through for all $t < t_\star$ (note that $t_\star \geq t_*$). This should involve in particular working with higher iterates of T and proving a version of (2.1) with Λ replaced by $\chi_{\min}$. (In some works on the thermodynamic formalism, the value of $t_\star$ is called the freezing point of the geometric family, in analogy with $1/t$ being thought of as temperature.)

Some inspiration for $\chi_{\min}$ being the correct quantity to use can be found in finite-horizon billiards in a triangular lattice. All scatterers on such tables are circles of equal radius R, thus the positive Lyapunov exponent of a period 2 orbit between two scatterers at minimal distance from one another is precisely $\Lambda = 1 + \frac{2\tau_{\min}}{R}$. In this case, $\Lambda = \chi_{\min}$ and so $t_* = t_\star$. If δ is the atomic invariant measure supported on this period 2 orbit, then $P_\delta(t) = -t \log \Lambda$, so certainly $P(t) \geq -t \log \Lambda$ for all $t > 0$. Yet, it is not known even in this special case whether in fact $P(t) = -t \log \Lambda$ at some $t = t_\star$, or whether $t_\star = \infty$.

ACKNOWLEDGMENTS

I would like to thank Viviane Baladi, with whom the papers [3, 4] surveyed in this note were written. I would also like to thank Carlangelo Liverani, who has taught me much about transfer operators and the construction of Banach spaces, as well as Hong-Kun Zhang and Péter Bálint, who have repeatedly explained to me many things about billiards.

FUNDING

This work was partially supported by U. S. National Science Foundation grants DMS 1800321 and DMS 2055070.

REFERENCES

[1] V. Baladi, Anisotropic Sobolev spaces and dynamical transfer operators: C^∞ foliations. In *Algebraic and topological dynamics*, edited by S. Kolyada, Y. Manin, and T. Ward, pp. 123–136, Contemp. Math., Amer. Math. Society, 2005.

[2] V. Baladi, The quest for the ultimate anisotropic Banach space. *J. Stat. Phys.* **166** (2017), 525–557.

[3] V. Baladi and M. F. Demers, On the measure of maximal entropy for finite horizon Sinai billiard maps. *J. Amer. Math. Soc.* **33** (2020), no. 2, 381–449.

[4] V. Baladi and M. F. Demers, Thermodynamic formalism for dispersing billiards. 2020, arXiv:2009.10936.

[5] V. Baladi and S. Gouezel, Good Banach spaces for piecewise hyperbolic maps via interpolation. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26** (2009), 1453–1481.

[6] V. Baladi and S. Gouezel, Banach spaces for piecewise cone hyperbolic maps. *J. Mod. Dyn.* **4** (2010), 91–137.

[7] V. Baladi and M. Tsujii, Anisotropic Hölder and Sobolev spaces for hyperbolic diffeomorphisms. *Ann. Inst. Fourier (Grenoble)* **57** (2007), 127–154.

[8] P. Balint and I. P. Toth, Correlation decay in certain soft billiards. *Comm. Math. Phys.* **243** (2003), 55–91.

[9] F. Baras and P. Gaspard, Chaotic scattering and diffusion in the Lorentz gas. *Phys. Rev. E (3)* **51** (1995), no. 6, 5332–5352.

[10] M. Blank, G. Keller, and C. Liverani, Ruelle–Perron–Frobenius spectrum for Anosov maps. *Nonlinearity* **15** (2001), no. 6, 1905–1973.

[11] R. Bowen, Some systems with unique equilibrium states. *Math. Syst. Theory* **8** (1974), 193–202.

[12] R. Bowen, Hausdorff dimension of quasi-circles. *Publ. Math. IHÉS* **50** (1979), 11–26.

[13] M. Brin and A. Katok, On local entropy. In *Geometric Dynamics (Rio de Janeiro, 1981)*, pp. 30–38, Lecture Notes in Math. 1007, Springer, Berlin, 1983.

[14] L. Bunimovich and Ya. G. Sinai, Statistical properties of Lorentz gas with periodic configuration of scatterers. *Comm. Math. Phys.* **78** (1980/1981), 479–497.

[15] L. Bunimovich, Ya. G. Sinai, and N. Chernov, Markov partitions for two-dimensional hyperbolic billiards. *Russian Math. Surveys* **45** (1990), 105–152.

[16] L. Bunimovich, Ya. G. Sinai, and N. Chernov, Statistical properties of two-dimensional hyperbolic billiards. *Russian Math. Surveys* **46** (1991), 47–106.

[17] K. Burns, V. Climenhaga, T. Fisher, and D. J. Thompson, Unique equilibrium states for geodesic flows in nonpositive curvature. *Geom. Funct. Anal.* **28** (2018), no. 5, 1209–1259.

[18] J. Buzzi, S. Crovisier, and O. Sarig, Measures of maximal entropy for surface diffeomorphisms. To appear in *Ann. of Math.* **195** (2022), arXiv:1811.02240.

[19] J. Chen, F. Wang, and H.-K. Zhang, Markov partition and thermodynamic formalism for hyperbolic systems with singularities. 2019, arXiv:1709.00527.

[20] N. Chernov, Topological entropy and periodic points of two-dimensional hyperbolic billiards. *Funct. Anal. Appl.* **25** (1991), no. 1, 39–45.

[21] N. Chernov, Decay of correlations in dispersing billiards. *J. Stat. Phys.* **94** (1999), 513–556.

[22] N. Chernov, Sinai billiards under small external forces. *Ann. Henri Poincaré* **2** (2001), 197–236.

[23] N. Chernov, Sinai billiards under small external forces II. *Ann. Henri Poincaré* **9** (2008), 91–107.

[24] N. Chernov, G. Eyink, J. Lebowitz, and Ya. G. Sinai, Derivation of Ohm's law in a deterministic mechanical model. *Phys. Rev. Lett.* **70** (1993), 2209–2212.

[25] N. Chernov, G. Eyink, J. Lebowitz, and Ya. G. Sinai, Steady-state electrical conduction in the periodic Lorentz gas. *Comm. Math. Phys.* **154** (1993), 569–601.

[26] N. Chernov and C. Haskell, Nonuniformly hyperbolic K-systems are Bernoulli. *Ergodic Theory Dynam. Systems* **16** (1996), no. 1, 19–44.

[27] N. Chernov and R. Markarian, *Chaotic Billiards*. Math. Surveys Monogr. 127, AMS, Providence, RI, 2006.

[28] N. Chernov and L.-S. Young, Decay of correlations for Lorentz gases and hard balls. In *Hard ball systems and the Lorentz gas*, edited by D. Sasz, pp. 89–120, Encyclopaedia Math. Sci. 101, Springer, Berlin, 2000.

[29] N. Chernov, H-K. Zhang, and P. Zhang, Electrical current for Sinai billiards under general small forces. *J. Stat. Phys.* **153** (2013), 1065–1083.

[30] M. F. Demers, A gentle introduction to anisotropic Banach spaces. *Chaos Solitons Fractals* **116** (2018), 29–42.

[31] M. F. Demers and C. Liverani, Stability of statistical properties in two-dimensional piecewise hyperbolic maps. *Trans. Amer. Math. Soc.* **360** (2008), no. 9, 4777–4814.

[32] M. F. Demers and H.-K. Zhang, Spectral analysis of the transfer operator for the Lorentz gas. *J. Mod. Dyn.* **5** (2011), 665–709.

[33] M. F. Demers and H.-K. Zhang, A functional analytic approach to perturbations of the Lorentz gas. *Comm. Math. Phys.* **324** (2013), no. 3, 767–830.

[34] M. F. Demers and H.-K. Zhang, Spectral analysis of hyperbolic systems with singularities. *Nonlinearity* **27** (2014), 379–433.

[35] J. De Simoi, V. Kaloshin, and M. Leguil, Marked length spectral determination of analytic chaotic billiards with axial symmetries. 2019, arXiv:1905.00890.

[36] V. Donnay, Non-ergodicity of two particles interacting via a smooth potential. *J. Stat. Phys.* **96** (1996), no. 5–6, 1021–1048.

[37] G. Gallavotti and D. S. Ornstein, Billiards and Bernoulli schemes. *Comm. Math. Phys.* **38** (1974), 83–101.

[38] P. L. Garrido, Kolmogorov–Sinai entropy, Lyapunov exponents, and mean free time in billiard systems. *J. Stat. Phys.* **88** (1997), no. 3–4, 807–824.

[39] S. Gouëzel, Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.* **38** (2010), no. 4, 1639–1671.

[40] S. Gouëzel and C. Liverani, Banach spaces adapted to Anosov systems. *Ergodic Theory Dynam. Systems* **26** (2006), no. 1, 189–217.

[41] S. Gouézel and C. Liverani, Compact locally maximal hyperbolic sets for smooth maps: fine statistical properties. *J. Differential Geom.* **79** (2008), 433–477.

[42] G. Iommi, Multifractal analysis for countable Markov shifts. *Ergodic Theory Dynam. Systems* **25** (2005), 1881–1907.

[43] A. Katok and B. Hasselblatt, *An introduction to the modern theory of dynamical systems*. Cambridge University Press, 1986.

[44] Y. Lima and C. Matheus, Symbolic dynamics for non-uniformly hyperbolic surface maps with discontinuities. *Ann. Sci. Éc. Norm. Supér.* **51** (2018), 1–38.

[45] R. Markarian, E. J. Pujals, and M. Sambarino, Pinball billiards with dominated splitting. *Ergodic Theory Dynam. Systems* **30** (2010), 1757–1786.

[46] I. Melbourne and M. Nicol, Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc.* **360** (2008), 6661–6676.

[47] D. S. Ornstein and B. Weiss, Geodesic flows are Bernoullian. *Israel J. Math.* **14** (1973), 184–198.

[48] Ya. B. Pesin, Lyapunov characteristic exponents and smooth ergodic theory. *Russian Math. Surveys* **32** (1977), 55–114.

[49] Ya. B. Pesin, Dynamical systems with generalized hyperbolic attractors: hyperbolic, ergodic and topological properties. *Ergod. Theory Dyn. Syst.* **12** (1992), no. 1, 123–152.

[50] Ya. B. Pesin, S. Senti, and K. Zhang, Thermodynamics of towers of hyperbolic type. *Trans. Amer. Math. Soc.* **368** (2016), 8519–8552.

[51] L. Rey-Bellet and L.-S. Young, Large deviations in nonuniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems* **28** (2008), 587–612.

[52] V. Rom-Kedar and D. Turaev, Big islands in dispersing billiard-like potentials. *Phys. D* **130** (1999), no. 3–4, 187–210.

[53] D. Ruelle, *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics*. Addison-Wesley Publishing Co., Reading, MA, 1978.

[54] H. H. Rugh, The correlation spectrum for hyperbolic analytic maps. *Nonlinearity* **5** (1992), no. 6, 1237–1263.

[55] H. H. Rugh, Fredholm determinants for real-analytic hyperbolic diffeomorphisms of surfaces. In *XIth International Congress of Mathematical Physics (Paris, 1994)*, pp. 297–303, 1994, Internat. Press, Cambridge, MA, 1995.

[56] O. Sarig, Bernoulli equilibrium states for surface diffeomorphisms. *J. Mod. Dyn.* **5** (2011), no. 3, 593–608.

[57] Ya. G. Sinai, Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. *Uspekhi Mat. Nauk* **25** (1970), no. 2, 141–192.

[58] Ya. G. Sinai, Gibbs measures in ergodic theory. *Russian Math. Surveys* **27** (1972), no. 4, 21–70.

[59] P. Walters, *An introduction to ergodic theory*. Grad. Texts in Math. 79, Springer, Berlin, 1982.

[60] L.-S. Young, Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **147** (1998), 585–650.

[61] H.-K. Zhang, Current in periodic Lorentz gases with twists. *Comm. Math. Phys.* **306** (2011), 747–776.

MARK F. DEMERS

Department of Mathematics, Fairfield University, Fairfield, CT 06824, USA, mdemers@fairfield.edu

GEOMETRIC METHODS IN HOLOMORPHIC DYNAMICS

ROMAIN DUJARDIN

ABSTRACT

In this text we review a selection of contemporary research themes in holomorphic dynamics. The main topics that will be discussed are geometric (laminar and woven) currents and their applications, bifurcation theory in one and several variables, and the problem of wandering Fatou components.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37F80; Secondary 32U40, 37F46, 37F44, 37F10

KEYWORDS

Laminar and woven currents, stability and bifurcations, wandering Fatou components

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3460–3482
DOI 10.4171/ICM2022/60

Published by EMS Press

Holomorphic dynamics was once part of classical complex analysis, but since its rebirth in the 1980s it keeps enlarging its scope, integrating new ideas, and developing new interactions. Some main tendencies of contemporary holomorphic dynamics are the convergence between its one- and higher-dimensional aspects and its ever deeper interconnection with algebraic and arithmetic dynamics. As a consequence, there is an endless diversification of the available mathematical techniques. Besides the classical methods from dynamics and complex analysis, its modern toolbox now comprises sophisticated tools and ideas imported from complex geometry, pluripotential theory (and its latest advances for currents of higher bidegree), algebraic geometry and commutative algebra, non-Archimedean analysis and geometry, arithmetic geometry (in particular, arithmetic equidistribution theory), Teichmüller theory, geometric group theory, etc. Conversely, each of these domains benefits from its interaction with holomorphic dynamics, by gaining new problems and examples. Many (though not all!) of these connections were reported in recent ICMs [21,25,34,40,71,77]. Our purpose here is to present a few contemporary research themes whose common thread—if one were to find one—is an emphasis on "soft" geometric techniques, such as the basic geometry of analytic subsets in $\mathbb{C}^n$. These represent only a tiny piece of the domain, reflecting, of course, the author's own taste and research interests. The main topics that will be discussed are geometric currents, bifurcation theory, and the problem of wandering Fatou components. The reader will soon notice that these three subjects are largely interrelated. Many open problems have also been included, as a motivation for future investigations.

Let us describe in more detail the contents of this paper. Section 1 is a short survey on positive closed currents with "geometric structure". The use of geometric currents in holomorphic dynamics was pioneered by Bedford, Lyubich, and Smillie in their seminal work [9] on complex Hénon maps. Since then they have turned into a very versatile tool, with many applications. Here we intend to give the flavor of a few specific results and how they are used in dynamical problems, so this part of the paper will be a bit more technical than the remaining sections.

Holomorphic dynamics is equally about the dynamics of a holomorphic map f and about the evolution of this dynamical behavior when f depends on certain parameters. The basic stability/bifurcation theory of rational maps in one variable was designed by Mañé, Sad, Sullivan, and Lyubich [69,70,73] in the 1980s, who showed that one-dimensional rational maps are generically structurally stable, using surprisingly elementary arguments. For the quadratic family $z^2 + c$, $c \in \mathbb{C}$, the bifurcation locus is the celebrated Mandelbrot set, whose intricate structure was thoroughly studied since then, using a variety of combinatorial and geometric methods. This research area was profoundly renewed in the 2000s by the systematic investigation of higher-dimensional phenomena, and in particular with the introduction of bifurcation currents by DeMarco [32]. The bifurcation theory of holomorphic dynamical systems is nowadays a very active research domain, and a meeting point between the communities of one and several variable dynamicists. We relate this continuing story in Section 2.

Finally, one recent breakthrough is the construction of wandering Fatou components in higher-dimensional polynomial dynamics, which at the same time solves an old problem and raises many questions. We review these recent developments in Section 3.

Let us conclude this introduction with a little notice. Some important theorems will be mentioned only in passing, while others are isolated within numbered environments: this is meant to keep the reading flow, not to reflect a hierarchy of importance. Likewise, the list of references is already quite long, but not exhaustive, and we apologize in advance for any serious omission.

1. GEOMETRIC CURRENTS

1.1. Definitions

This part assumes some familiarity with positive currents and pluripotential theory (see, e.g., Demailly [31] for basics). All the definitions here are local, so we work in some bounded open set $\Omega \subset \mathbb{C}^k$. Let T be a positive closed current of bidimension (p,p) in Ω. Following Bedford, Lyubich, and Smillie [9], we say that T is *locally uniformly laminar* if there exists a lamination by complex submanifolds of dimension p embedded in Ω such that the restriction of T to any flow box B of the lamination is of the form

$$T|_B = \int_\tau [\Delta_t] d\nu(t). \tag{1.1}$$

Here τ is a global transversal in the flow box B, the Δ_t are the plaques of the lamination in the flow box, and ν is a positive measure on τ. The word "uniform" here refers to the local uniformity of the geometry of the plaques Δ_t We say that T is *laminar* if there exists a sequence of open subsets Ω_k, together with a sequence of currents T_k, locally uniformly laminar in Ω_k, such that T_k increases to T. The Ω_k should be thought of as a union of many small polydisks, whose complement has a small mass. The key word in the definition is "increases." Intuitively, this definition should be understood as follows: T_k represents all the disks contained in T of some given size (say 2^{-k}); then, to T_k we add $T_{k+1} - T_k$ which is made of disks of size $2^{-(k+1)}$ (which may have nonempty boundary in Ω_k, but form a lamination in $\Omega_{k+1} \subset \Omega_k$), and so on. The sequence T_k is not canonical, and has to be understood as the choice of a "representation" of T as a laminar current. From this we can deduce another representation of T as an integral over an abstract family of compatible holomorphic disks, namely

$$T = \int_{\mathcal{A}} [D_\alpha] d\mu(\alpha). \tag{1.2}$$

Here *compatible* means that two disks can only intersect along some relatively open subset, but there is no further restriction on the geometry of the D_α. Even if this definition is rather restrictive, it can lead to pathological examples, and for dynamical applications we will have to constrain it further (see the notion of "strongly approximable" current below).

It was observed by Dinh [39] that in many situations it is more natural to let the disks admit nontrivial intersections. One then defines *uniformly woven* currents by replacing "lamination" by "web" in (1.1), where a web is locally given by a family of disks of dimension p with uniformly bounded volume or, more generally, a family of holomorphic chains of dimension p with uniformly bounded volume (any such family is precompact for the Hausdorff topology, so it makes sense to define a measure on a set of such disks). Then, *woven* currents are defined from uniformly woven ones as in the laminar case. A difference between laminar and woven currents is that in the woven case the measures in (1.1) and (1.2) are not determined by T (e.g., the standard Kähler form in $\mathbb{C}^2$ admits several representations as a uniformly woven current), so a woven current has to be thought of as "marked" by such a measure μ. It is not completely obvious to show that not every positive closed current is woven; we leave this as an exercise to the reader!

There is no unified reference for the basic properties of laminar and woven currents. Besides [9] and [39], the information in this paragraph was extracted from various papers, notably by De Thélin and the author [28, 30, 37, 43, 44, 46]. In the following we use the word *geometric* as a synonym of "laminar or woven."

1.2. Construction and approximation

Positive closed currents often appear as limits of sequences of normalized currents of integration. Furthermore, by a classical theorem of Lelong, any positive closed current of bidegree $(1,1)$ is locally of this form. In this section we explain how, under appropriate hypotheses, a geometric structure can be extracted from such an approximation.

Still working locally in some open set $\Omega \subset \mathbb{C}^k$, endowed with its standard Hermitian structure, we say that a submanifold V of dimension p in Ω *has size r at $x \in V$* if it contains a graph over a ball of radius r of its tangent space $T_x V$, relative to the orthogonal projection to $T_x V$, with slope (i.e., the norm of the derivative of the graphing map) bounded by 1. In particular, V has no boundary in $B(x, cr)$ for some constant c depending only on p and k. This notion of size makes sense in any compact complex manifold, up to uniform constants, by choosing a finite covering by coordinate charts and a Hermitian metric. Note that we may relax this definition by allowing V to be an analytic set: then V can have several irreducible components at x, some of which being of size r.

If V is any submanifold (or subvariety) of Ω, possibly with boundary, and $r > 0$, we denote by V^r the set of $x \in V$ such that V has size r at x. In this way we get a tautological decomposition, $V = V^r \cup (V \setminus V^r)$, which is reminiscent of the thin–thick decomposition of hyperbolic manifolds.

Assume now that V_n is a sequence of p-dimensional subvarieties of volume v_n, such that $v_n^{-1}[V_n]$ converges to a positive closed current T. If $\mathrm{Vol}(V_n^r) \geq v_n(1 - \varepsilon(r))$ where ε is a function independent of n and such that $\varepsilon(r) \to 0$ as $r \to 0$, then one may extract a subsequence so that $v_n^{-1}[V_n^r]$ converges to a geometric current $T^r \leq T$ with the mass estimate $\mathbf{M}(T - T^r) \leq \varepsilon(r)$. This endows T with a geometric structure: if $p \leq k-2$, we obtain a woven current and, if $p = k-1$, this current is laminar. Indeed, if $p = k-1$, by the persistence of proper intersections, the limiting graphs cannot intersect nontrivially. (Note that when $p \leq k-2$, intersections can appear at the limit even if the V_n are submanifolds. Conversely, if in codimension 1 we allow the V_n to admit self-intersections, we obtain woven currents also in this case.)

A technically convenient option is to further assume that the disks constituting V_n^r are submanifolds (without boundary) in a subdivision of Ω by cubes of size cr (for some constant $c > 0$). This is consistent with the manner in which the V_n^r are constructed in practice, and the resulting definition is equivalent (see [46]). In this way the limiting currents T^r are uniformly geometric in the cubes of this subdivision.

There are several easily checkable geometric and/or topological criteria ensuring this condition, which sometimes give an explicit bound on $\varepsilon(r)$:

- If $\psi : \mathbb{C} \to X$ is an entire curve in a projective manifold, then by Ahlfors' theory of covering surfaces, for well-chosen sequences $R_n \to \infty$, $V_n := \psi(D(0, R_n))$ satisfies $v_n^{-1}[\partial V_n] \to 0$ and $\mathrm{Vol}(V_n^r) \geq v_n(1 - \varepsilon(r))$ for $\varepsilon(r) = O(r^2)$. Thus the cluster values of $v_n^{-1}[V_n]$ are closed woven currents; if, in addition, ψ is injective and $\dim(X) = 2$, then they are laminar (Bedford–Lyubich–Smillie [9], Cantat [24]).

- If V_n is a sequence of algebraic curves in a projective surface whose geometric genus is $O(v_n)$, then $\mathrm{Vol}(V_n^r) \geq v_n(1 - \varepsilon(r))$ for $\varepsilon(r) = O(r^2)$, therefore the limiting currents of $v_n^{-1}[V_n]$ are woven; under a mild additional condition on the singularities of V_n, they are laminar (Dujardin [43]).

- If $\iota_n : \mathbb{P}^p \to X$ is a sequence of holomorphic mappings of generic degree 1 to a projective manifold X of dimension $k > p$ and $V_n = \iota_n(\mathbb{P}^p)$, then the limiting currents of $v_n^{-1}[V_n]$ are woven (Dinh [39]). In addition, $\varepsilon(r) = O(r^2)$ [46].

- If V_n is a sequence of smooth curves in the unit ball in $\mathbb{C}^2$, whose genus is $O(v_n)$, then the limiting currents of $v_n^{-1}[V_n]$ are laminar (De Thélin [28]). A version of this result in arbitrary dimension is given by De Thélin in [30].

In all these papers, the geometric structure is obtained by projecting V_n in several directions and keeping only from V_n the graphs over these directions with bounded diameter or volume. The bound $\varepsilon(r) = O(r^2)$ plays an important role in applications as we shall see below.

1.3. Geometric intersection

The main interest of geometric currents is the possibility of a geometric interpretation of their wedge products. This technique was introduced in [9], and it was systematized and generalized in several subsequent works. Such results are so far essentially available in dimension 2; again since the problem is local, we work in some open set $\Omega \subset \mathbb{C}^2$, say a ball. If T_1 and T_2 are closed positive $(1,1)$ currents in Ω, we say that the wedge product $T_1 \wedge T_2$ is well defined if $u_1 \in L^1_{\mathrm{loc}}(T_2)$, where u_i is a local potential of T_i, in which case we set $T_1 \wedge T_2 = dd^c(u_1 T_2)$. This condition and the resulting wedge product are actually symmetric in T_1 and T_2. We also say that such a current is *diffuse* if it gives no mass to curves.

For uniformly laminar and woven currents, geometric intersection is easy and basically follows from Fubini's theorem. Indeed, assume that T_1 and T_2 are uniformly geometric $(1,1)$-currents in Ω, which locally in Ω admit the representation $T_i = \int [\Delta_t^i] d\nu_i(t)$. Then, if the wedge product $T_1 \wedge T_2$ is well defined, locally we have that

$$T_1 \wedge T_2 = \int [\Delta_t^1 \cap \Delta_s^2] d\nu_1(t) d\nu_2(s), \tag{1.3}$$

where $[\Delta_t^1 \cap \Delta_s^2]$ is the sum of point masses at isolated intersection points, counting multiplicities (see [38, 44]). In addition, if T_1 and T_2 are laminar and diffuse, nontransverse intersections do not contribute to the integral, so we can restrict to transverse intersections. Note the intermediate "semigeometric intersection" result

$$T_1 \wedge T_2 = \int \big([\Delta_t^1] \wedge T_2\big) d\nu_1(t), \tag{1.4}$$

which makes sense for an arbitrary positive closed current T_2.

Now assume that T is a geometric positive closed current in $\Omega \subset \mathbb{C}^2$ and S is an arbitrary positive closed current in Ω such that the wedge product $S \wedge T$ is well defined. We say that $T \wedge S$ is *semigeometric* if there is a representation $T = \lim_{r\to 0} T^r$ as an increasing limit of uniformly geometric currents, such that $T^r \wedge S$ increases to $T \wedge S$ as $r \to 0$. Thanks to (1.4), $T^r \wedge S$ admits a geometric interpretation. If now S itself is a geometric current, we say that the wedge product $T \wedge S$ is *geometric* if there are representations $T^r \nearrow T$ and $S^r \nearrow S$ such that $T^r \wedge S^r$ (which has a geometric interpretation by (1.3)) increases to $T \wedge S$.

We say that a geometric current is *strongly approximable* if there is a representation $T^r \nearrow T$ where T^r is uniformly geometric in a subdivision Ω^r of Ω into cubes of size r, and $\varepsilon(r) = \mathbf{M}(T - T^r) = O(r^2)$. As we have seen in Section 1.2, this estimate is commonly satisfied in practice. (Technically, some freedom on the choice of Ω^r is also necessary, but we do not dwell on this point.) The sharpest version of the geometric intersection theorem for geometric currents in dimension 2 is the following:

Theorem 1.1 (Dujardin [38, 44, 45]). *Let S and T be closed positive $(1,1)$ currents in $\Omega \subset \mathbb{C}^2$, such that the wedge product $T \wedge S$ is well defined. Assume that T is a strongly approximable geometric current. Then, if S has locally bounded potentials, or if $T \wedge S$ gives no mass to pluripolar sets, then $T \wedge S$ is semigeometric.*

A consequence of this theorem, which is often as useful as the result itself, is that if T was obtained as the limit of $v_n^{-1}[V_n]$ as in Section 1.2, then $v_n^{-1}[V_n^r] \wedge S$ is close to $T \wedge S$ for small r and large n.

Applying Theorem 1.1 to $T \wedge S$ and $S \wedge T$, we get:

Corollary 1.2. *If in Theorem 1.1 both S and T are strongly approximable geometric currents and $T \wedge S$ gives no mass to pluripolar sets, then $T \wedge S$ is geometric.*

The main open problem at this stage is the extension of these results to higher dimensions.

Question 1.3. Is there a version of Theorem 1.1 for geometric currents of arbitrary codimension?

While the case of uniformly geometric currents and the case where T is of bidimension $(1,1)$ follow without serious difficulties (see [46] and [47] for details), the general case remains a challenge so far. The crucial mass estimate $\mathbf{M}(T - T^r) = O(r^2)$ is known to hold in some significant cases (see [46]), but it does not appear to be sufficient to conclude for currents of arbitrary bidimension.

1.4. Dynamical applications

The first application of laminar currents by Bedford, Lyubich, and Smillie [9] was to prove that certain intersections are nonempty. A typical example is the following: assume that we are given an entire curve $\psi : \mathbb{C} \to X$ in some projective manifold, and let T be a closed current obtained from ψ by Ahlfors' construction. Let S be a current of bidegree $(1,1)$ with bounded potentials. If we know that $\int T \wedge S > 0$ (for instance, for cohomological reasons), then by Theorem 1.1, this intersection is semigeometric, therefore $S|_{\psi(D(0,R_n))}$ is nonzero for large n. (A version of this result which does not apply to laminarity was proved by Dinh and Sibony [42].) This fact (as well as some variants) plays an important role in the dynamics of automorphisms and birational maps on complex surfaces, where it is used as a tool to create intersections between stable and unstable manifolds. This is used in [9] to establish that any saddle point belongs to the support of the maximal entropy measure; this technique also appears in the work of Cantat, Favre, Lyubich, and the author [24, 26, 52, 53], among others. Note also that the failure of Theorem 1.1 for unbounded potentials can be viewed as the main reason why the uniqueness of the measure of maximal entropy for general birational maps of surfaces remains an unsolved problem.

Another use of geometric intersection, which was initiated in [45], concerns the dynamical analysis of wedge products of dynamically defined currents. Indeed, suppose that f is a self-map of some complex manifold X, and $f^n(L)$ is a sequence of iterated curves such that $d^{-n} f^n(L)$ converges to a geometric current T, with a control of the asymptotic geometry of $f^n(L)$ as in Section 1.2. Assume also that S is some invariant current of bidegree $(1,1)$: $f^*S = dS$ and that $T \wedge S$ is a semigeometric intersection. Then for large n, the action of f^k on the bounded geometry part of $d^{-n}[f^n(L)] \wedge S$ is a good approximation of the action of f^k on $T \wedge S$, and its expansion properties "in the direction of T" can be analyzed geometrically by "soft" methods, such as counting disjoint disks of size r and length–area estimates (see below Theorem 2.4 for a worked out example). This idea was used in various contexts by De Thélin and others [29, 36, 37, 45, 46].

1.5. Foliations

Foliated Ahlfors currents play an important role in the work of Brunella and McQuillan on singular holomorphic foliations (see, e.g., [20]). Geometric intersection has been applied in foliation theory to prove the vanishing of certain self-intersections. For a positive current directed by a holomorphic foliation on a compact Kähler surface, this vanishing can in turn be used to infer dynamical properties of the foliation such as the nonexistence of invariant transverse measures (for closed currents) or the uniqueness of harmonic measures (for dd^c-closed currents), according to a Hodge-theoretic formalism for dd^c-closed currents devised by Fornæss and Sibony [58]. Proving that the self-intersection of harmonic currents directed by holomorphic foliations vanishes is a very difficult problem in the presence of singularities. On $\mathbb{P}^2$ this can be treated by regularizing with global automorphisms, the general case makes use of the theory of densities of Dinh and Sibony (see [41]). Here we want to mention a more elementary-looking problem:

Question 1.4. Does there exist a diffuse (closed) uniformly laminar current on $\mathbb{P}^2$?

The expected answer to the question is "no," since it is generally expected that there does not exist a Riemann surface lamination embedded in $\mathbb{P}^2$. The above question is supposed to be the "easy case" of this deep conjecture (since it deals with laminations with transverse measures), and it admits a straightforward approach: if T is such a current, then $T \wedge T = 0$ because of the laminar structure, which is impossible on $\mathbb{P}^2$. This approach works well as soon as $T \wedge T$ is well defined in the sense of pluripotential theory (but it does not for a curve!), or when the holonomy of the induced lamination is Lipschitz [58]. But in general the holonomy of a Riemann surface lamination in $\mathbb{C}^2$ (that is, a holomorphic motion) is less regular and, surprisingly enough, the problem is still open so far. (See Kaufmann [66] for a discussion of the higher-dimensional case.)

2. BIFURCATION THEORY IN ONE AND SEVERAL DIMENSIONS

Let $(f_\lambda)_{\lambda\in\Lambda}$ be a family of rational maps on $\mathbb{P}^1$ of degree d, holomorphically parameterized by some complex manifold Λ. Then the well-known Fatou–Julia decomposition of the phase space is mirrored by a stability–bifurcation dichotomy of the parameter space. The proper definition of stability in this context was found simultaneously by Mañé–Sad–Sullivan and Lyubich [69, 70, 73]: the family $(f_\lambda)_{\lambda\in\Lambda}$ is *J-stable* over some domain $\Omega \subset \Lambda$ if one of the following equivalent conditions holds over Ω:

(i) the periodic points of (f_λ) do not collide or, equivalently, the nature (attracting, repelling, indifferent) of each periodic point remains the same in the family;

(ii) the Julia set $\lambda \mapsto J_\lambda$ moves continuously for the Hausdorff topology;

(iii) for any two parameters λ, λ' in Ω, $f_\lambda|_{J_\lambda}$ is topologically conjugate to $f_{\lambda'}|_{J_{\lambda'}}$;

(iv) the orbits of the critical points f_λ do not bifurcate.

The equivalence between these properties relies on the notion of *holomorphic motion* (also known as *holomorphic families of injections*) of a subset of the Riemann sphere, and the simple, yet powerful idea of automatic extension of a holomorphic motion to its closure (the "λ-lemma"). Condition (iv), together with the finiteness of the critical set, easily implies that in any such parameterized family (f_λ), *the stability locus is open and dense in* Λ. In other words, *one-dimensional polynomial and rational maps are generically stable*.

For the emblematic family $f_\lambda(z) = z^2 + \lambda$ of quadratic polynomials, the bifurcation locus is the boundary of the Mandelbrot set M (connectivity locus). Even if its interior is empty, ∂M is still quite large, as shown by the following famous result of Shishikura [80]: *∂M has Hausdorff dimension 2.* This property was extended to arbitrary families of rational maps by Tan Lei and McMullen [67,75]. The basic technical tool underlying Shishikura's theorem is the phenomenon of *parabolic implosion*, which will also play an important role below. Note that is still unknown whether ∂M has zero or positive Lebesgue measure.

This research area was renewed in the last 20 years as the result of several tendencies: (1) the use of positive closed currents, and (2) the move towards higher dimensions (both in dynamical and parameter spaces). In the next few pages, we review some of these developments; in particular, we will see how these influential one-dimensional results translate to new settings. Lack of space prevents us from giving a complete treatment, and some important results will barely be mentioned. Also, we do not discuss the profound connection with arithmetic dynamics, for which the reader is referred, e.g., to [34], or bifurcations of Kleinian groups (see [35,48]).

2.1. Bifurcation currents in one-dimensional dynamics

Let as above $(f_\lambda)_{\lambda\in\Lambda}$ be a holomorphic family of rational maps of degree d. The following addition to the list of equivalent conditions to stability was found by DeMarco [33]:

(v) the Lyapunov exponent of the unique measure of maximal entropy $\chi(\mu_{f_\lambda})$ is a pluriharmonic function of λ.

The *bifurcation current* is then defined by $T_{\mathrm{bif}} := dd^c_\lambda \chi(\mu_{f_\lambda})$. For the family of quadratic polynomials, T_{bif} ($= \mu_{\mathrm{bif}}$, see below) is the harmonic measure of the Mandelbrot set.

The original definition of the bifurcation current in [32] can be interpreted geometrically as follows (see [51]). Consider the fibered dynamical system in $\Lambda \times \mathbb{P}^1$ defined by $\hat{f} : (\lambda, z) \mapsto (\lambda, f_\lambda(z))$. It admits a natural invariant current $\hat{T}$ of bidegree $(1,1)$, satisfying $\hat{f}^*\hat{T} = d\hat{T}$, whose restriction to a generic vertical line $\{\lambda\} \times \mathbb{P}^1$ is the maximal entropy measure μ_{f_λ}. Now, take a holomorphically moving (or "marked") point $\lambda \mapsto a(\lambda)$ in $\mathbb{P}^1$, and denote by Γ_a its graph in $\Lambda \times \mathbb{P}^1$. If $\pi : \Lambda \times \mathbb{P}^1 \to \Lambda$ is the natural projection, we obtain a current in Λ associated to a by slicing $\hat{T}$ by Γ_a and projecting down to Λ: $T_a := \pi_*(\hat{T} \wedge [\Gamma_a])$. If in a holomorphic family (f_λ), the critical points are marked by holomorphic functions $\lambda \mapsto c_i(\lambda)$ (this is always possible up to replacing Λ by some branched cover), we thus obtain the corresponding bifurcation currents T_{c_i}. It turns out that $T_{\mathrm{bif}} = \sum T_{c_i}$: this follows from a variant of the Manning–Przytycki formula for the Lyapunov exponent $\chi(\mu_{f_\lambda})$, which in the case of polynomials is written as

$$\chi(\mu_f) = \log d + \sum_i G_f(c_i),$$

where G_f is the dynamical Green function (which satisfies $dd^c G_f = \mu_f$).

Bifurcation currents have turned into a fundamental tool for exploring higher dimensional issues in parameter spaces. Here is a sample problem: consider a critically marked family $(f_\lambda, c_i(\lambda))$ and suppose that for some parameter $\lambda_0 \in \Lambda$, the critical point $c_1(\lambda)$ bifurcates at $\lambda = \lambda_0$. Then a simple application of Montel's theorem shows that there is a sequence of parameters $\lambda_n \to \lambda_0$ such that, for $\lambda = \lambda_n$, c_1 is preperiodic. Now assume that several (say all) critical points bifurcate at λ_0: is it then possible to approximate λ_0 by parameters such that the corresponding critical points are preperiodic? Of course, in this question one has to discard a few "trivial" obstructions, e.g., when $\dim(\Lambda)$ is too small, so that there are not enough degrees of freedom to hope for an independent behavior of the critical points. Still after excluding these counterexamples, the answer to this problem is "no" (see [51, EXAMPLE 6.13]), the fundamental reason for this being the failure of Montel's theorem in higher dimension. Using currents is a known way of circumventing this problem in higher-dimensional dynamics, and, as a matter of fact, the following theorem holds:

Theorem 2.1 (Bassanelli–Berteloot [8], Dujardin–Favre [51]). *Let $(f_\lambda)_{\lambda\in\Lambda}$ be a holomorphic family of rational maps of degree $d \geq 2$. Then for every $k \leq \dim(\Lambda)$,*

$$\operatorname{Supp}(T_{\text{bif}}^k) \subset \overline{\{\lambda,\, f_\lambda \textit{ admits } k \textit{ periodic critical points}\}}. \tag{2.1}$$

(This result was actually not stated explicitly in [8, 51], see [48] for this formulation. The converse inclusion is studied below.)

When Λ is the moduli space $\mathcal{P}_d$ of polynomials of degree d with marked critical points (which is a finite quotient of $\mathbb{C}^{d-1}$) or the moduli space $\mathcal{M}_d$ of rational maps of degree d with marked critical points (which is of dimension $2d - 2$), we define the *bifurcation measure* μ_{bif} to be the maximal exterior power of T_{bif}, that is, $\mu_{\text{bif}} = T_{\text{bif}}^{d-1}$ or $\mu_{\text{bif}} = T_{\text{bif}}^{2d-2}$, respectively. The following neat dynamical characterization of $\operatorname{Supp}(\mu_{\text{bif}})$ can be obtained:

Theorem 2.2 (Dujardin–Favre [51], Buff–Epstein [22]). *For $\Lambda = \mathcal{P}_d$ or $\mathcal{M}_d$, the support of μ_{bif} is the closure of (non-Lattès) strictly postcritically finite parameters, that is, parameters for which all critical points are preperiodic to a repelling cycle.*

A version of this result for intermediate powers of T_{bif} was obtained in [47], which explains to what extent the converse inclusion in (2.1) holds.

Sketch of proof. The most delicate point is to show that any non-Lattès postcritically finite parameter λ_0 belongs to $\operatorname{Supp}(\mu_{\text{bif}})$. To fix the ideas, assume that $\Lambda = \mathcal{M}_d$. Observe that λ_0 is an intersection point of a family of $(2d - 2)$ hypersurfaces of the form

$$\{\lambda \in \mathcal{M}_d,\ f_\lambda^n(c_i(\lambda)) = f_\lambda^{n+k}(c_i(\lambda))\}$$

(one for each critical point). The proof in [22] is based on two important ideas. The first one consists in proving that these hypersurfaces are smooth and transverse at λ_0: this is based on Teichmüller-theoretic ideas. Then, using this transversality, a version of Tan Lei's transfer principle between dynamical and parameter space allows comparing the mass of μ_{bif} in a carefully scaled small polydisk about λ_0 with the mass of $\mu_{f_{\lambda_0}}$ near the $f^n(c_i)$, and conclude that this mass is positive. ∎

In the space of polynomials of degree d, Theorem 2.2, together with other characterizations of $\operatorname{Supp}(\mu_{\text{bif}})$, e.g., in terms of landing of parameter rays, makes $\operatorname{Supp}(\mu_{\text{bif}})$ the natural analogue of the boundary of the Mandelbrot set for polynomials of higher degree.

This motivates an investigation of its topological and geometric properties. First, it is a compact set, which, for $d \geq 3$, is strictly contained in the boundary of the locus $\mathcal{C}_d$ of polynomials with connected Julia set. A topological consequence of Theorem 2.1 is that $\operatorname{Supp}(\mu_{\mathrm{bif}})$ is contained in the closure of $\operatorname{Int}(\mathcal{C}_d)$; on the other hand, it is unknown whether $\mathcal{C}_d$ is the closure of its interior. Gauthier [59] extended Shishikura's theorem to show that $\operatorname{Supp}(\mu_{\mathrm{bif}})$ *has maximal Hausdorff dimension at each of its points*. Let us also note that by using advanced nonuniform hyperbolicity techniques, it was shown by Astorg, Gauthier, Mihalache, and Vigny [6] that in the space $\mathcal{M}_d$ of rational maps of degree d, $\operatorname{Supp}(\mu_{\mathrm{bif}})$ *has positive volume*.

The technical core of Theorems 2.1 and 2.2 is the fact that T_{bif} and its exterior powers describe the asymptotic distribution of families of dynamically defined hypersurfaces in the parameter space, like parameters with a preperiodic critical point, or parameters with a periodic point of a given multiplier. Initiated in [7,8,51], this research theme has gradually evolved in scope and sophistication, notably through its connections with arithmetic equidistribution (see [54]).

A striking and unexpected consequence of this technology is an asymptotic estimate for the number of hyperbolic components in $\mathcal{M}_d$, which is so far not accessible by other means. Recall that a *hyperbolic component* is a connected component of the stability locus in which the dynamics is uniformly expanding on the Julia set. We say that a hyperbolic component Ω is *of disjoint type* $(n_1, \dots, n_{2d-2})$ if the critical points are attracted by distinct attracting cycles of respective exact period n_i.

Theorem 2.3 (Gauthier, Okuyama, and Vigny [60]). *The number $N(n)$ of hyperbolic components of disjoint type $(n, \dots, n)$ in $\mathcal{M}_d$ satisfies*

$$N(n) \underset{n\to\infty}{\sim} \frac{d^{(2d-2)n}}{(2d-2)!} \int_{\mathcal{M}_d} \mu_{\mathrm{bif}}.$$

(An analogous formula holds for arbitrary disjoint type $(n_1, \dots, n_{2d-2})$.) Note that the corresponding result in $\mathcal{P}_d$ is much easier and follows essentially from Bézout's theorem (together with a transversality argument). The value of $\int_{\mathcal{M}_d} \mu_{\mathrm{bif}}$ is known only for $d = 2$ [60].

Once the bifurcation measure is constructed on $\mathcal{P}_d$ or $\mathcal{M}_d$, it is natural to inquire about the dynamics of a μ_{bif}-typical parameter. In $\mathcal{M}_d$ this question is completely open so far. For the family of quadratic (and more generally unicritical) polynomials, it was shown by Graczyk–Swiatek [62] and Smirnov [81] in the late 1990s that a μ_{bif}-typical parameter satisfies the Collet–Eckmann condition; in particular, the local geometry of its Julia set is well understood. These results are based on combinatorial techniques and the landing of external and parameter rays, and the method carries over for degree d polynomials (see [51, THM. 10]). Interestingly, a completely new approach to the results of [62,81] was recently found, which applies to arbitrary families of rational maps.

Theorem 2.4 (De Thélin, Gauthier, and Vigny [36]). *Let $(f_\lambda)_{\lambda\in\Lambda}$ be an algebraic family of rational maps of degree d with a marked critical point $c(\lambda)$. Let T_c be the bifurcation current associated to c and $\|T_c\|$ be the associated total variation measure. Then for $\|T_c\|$-a.e. λ,*

$$\liminf_{n\to\infty} \left| Df_\lambda^n\big(c(\lambda)\big)\right| \geq \frac{1}{2}\log d > 0. \tag{2.2}$$

For the unicritical family $z^d + \lambda$, this statement is precisely the typicality of the Collet–Eckmann expansion property.

Sketch of proof. This is an application of the techniques of Section 1.4. We may assume that Λ is of dimension 1, so that T_c is just a positive measure on Λ. Consider the sequence of iterated graphs $\Gamma_{f^n(c)}$, parameterized by $\gamma_n : \lambda \mapsto (\lambda, f_\lambda^n(c(\lambda)))$. Then, as explained above, $T_c = \pi_*(\hat{T} \wedge [\Gamma_c])$, where $\pi : \Lambda \times \mathbb{P}^1 \to \Lambda$ is the first projection and $\hat{T}$ is the natural $\hat{f}$-invariant current in $\Lambda \times \mathbb{P}^1$. Using the $\hat{f}$-invariance of $\hat{T}$, we infer that

$$T_c = \pi_*\big(d^{-n}[\Gamma_{f^n(c)}] \wedge \hat{T}\big), \text{ and conversely } (\gamma_n)_*(T_c) = d^{-n}[\Gamma_{f^n(c)}] \wedge \hat{T}.$$

Since the $\Gamma_{f^n(c)}$ are algebraic curves of uniformly bounded genus, by the results of Section 1.2, the part $(d^{-n}[\Gamma_{f^n(c)}])^r$ of these curves made of disks of size r has mass $1 - O(r^2)$, and since $\hat{T}$ has continuous potential, by Theorem 1.1 the intersection $d^{-n}[\Gamma_{f^n(c)}] \wedge \hat{T}$ is carried by $(d^{-n}[\Gamma_{f^n(c)}])^r$, up to a small error $\eta(r)$. But to fill up a set of measure $1 - \eta(r)$ of $d^{-n}[\Gamma_{f^n(c)}] \wedge \hat{T}$, at least $c(r)d^n$ disjoint such disks are required, and, pulling them back by γ_n, we get a set of $c(r)d^n$ disjoint disks in Λ, covering a set of measure $1 - \eta(r)$ for T_c, each of which mapped under γ_n to a disk of size r. Being disjoint, most of the pulled-back disks in Λ have area at most Cd^{-n}, so the derivative of γ_n there must typically be larger than $Cd^{n/2}$. Analyzing how the derivative of γ_n is expressed in terms of the $Df_\lambda^k(c(\lambda))$, for $0 \leq k \leq n$, finally leads to (2.2). ■

As already mentioned, the theory of bifurcation currents has deep connections with arithmetic dynamics, and related rigidity problems in moduli spaces. A typical problem in this context is the classification of families with a marked point $(f_\lambda, a(\lambda))$ for which the bifurcation current T_a is "abnormally regular." The reader is referred to the recent monograph [55] by Favre and Gauthier for more on this topic.

2.2. Stability/bifurcation theory in higher dimension

Moving to higher dimension, it is tempting to imitate the definition of J-stability by coining a definition of stability from the noncollision of periodic points. An obvious difficulty is that in this context the automatic extension of holomorphic motions fails and the relevance of this definition needs to be justified, for instance, by proving its equivalence with other natural ones. Due to the variety of possible situations, in higher dimension the details depend on the category of maps under study. So far, this program has been fulfilled in two cases: polynomial automorphisms of $\mathbb{C}^2$ (by Lyubich and the author), and holomorphic maps on $\mathbb{P}^k$ (by Berteloot, Bianchi, and Dupont).

2.2.1. Polynomial automorphisms of $\mathbb{C}^2$

For a polynomial automorphism f of $\mathbb{C}^2$, we can define Julia sets J^+ and J^- respectively associated to forward and backward iteration, as well as the "small Julia set" $J = J^+ \cap J^-$, and $J^* \subset J$ the closure of the set of saddle periodic points, which is also the support of the maximal entropy measure [9]. Following [53], we say that a holomorphic family $(f_\lambda)_{\lambda\in\Lambda}$ of polynomial automorphisms of fixed dynamical degree d is *weakly* J^*-*stable* if (i) its saddle points do not bifurcate, hence (under mild assumptions) so do all periodic points. (Here the numbering of properties corresponds to that of the 1-dimensional case at the beginning Section 2.) Then the holomorphic motion of saddle points extends to a *branched holomorphic motion* of J^* and the condition is equivalent to (ii) $\lambda \mapsto J^*(f_\lambda)$ is continuous. Furthermore, the branched holomorphic motion extends to the "big Julia set" $J^+ \cup J^-$. It remains an open question whether weak J^*-stability yields a conjugacy on J^* or J (that is, whether an analogue of (iii) holds). It is proved in [13] that weak J^*-stability implies a probabilistic form of structural stability, that is, a conjugacy can be defined on a full measure subset for any hyperbolic measure. Also, weak J^*-stability preserves uniform hyperbolicity [13, 50], so the familiar concept of hyperbolic component makes sense in this setting.

Even if strictly speaking polynomial automorphisms have no critical points, the main issue in [53] is about condition (iv) (stability of critical points). Indeed, it a popular analogue of a prerepelling critical point for a 2-dimensional diffeomorphism is a heteroclinic tangency, so we are looking for a characterization of stability in terms of (absence of) tangencies. It is well known that in dissipative dynamics, homoclinic tangencies yield bifurcations from saddles to sources, and the main point of [53] is to find a mechanism for the converse implication. The key is the phenomenon of *semiparabolic implosion*.

Before moving on to this topic, let us point out that so far there is no theory of bifurcation currents for automorphisms of $\mathbb{C}^2$.

Question 2.5. For polynomial automorphisms of $\mathbb{C}^2$, is stability characterized by the harmonicity of the Lyapunov exponents of the maximal entropy measure? In other words, does an analogue of condition (v) above hold?

2.2.2. Semiparabolic implosion and tangencies

Parabolic implosion refers to a set of phenomena, discovered by Douady and Lavaurs, occurring when unfolding a periodic point with a rational indifferent multiplier. To be specific, consider a family of the form

$$f_\lambda(z) = (1+\lambda)z + z^2 + \text{h.o.t.}$$

in a neighborhood of the origin, for small λ. For $\lambda = 0$, the fixed point 0 admits a basin of attraction $\mathcal{B}$. Now If λ approaches the origin tangentially to the imaginary axis, we can track precisely how the parabolic basin $\mathcal{B}$ "implodes" by "passing through the eggbeater" created between two slightly repelling fixed points $p_\lambda = 0$ and $q_\lambda \approx -\lambda$. More precisely, for well-chosen λ_n, $f^n_{\lambda_n}$ converges locally uniformly in $\mathcal{B}$ to a nonconstant *Lavaurs map* $\psi : \mathcal{B} \to \mathbb{C}$, depending on (λ_n). Of course, for $\lambda_n \equiv 0$, $\psi = 0$: in this sense the limiting dynamics of f_λ as $\lambda \to 0$ is richer than that of f_0. This gives rise to a wealth of dynamical phenomena at a such a parabolic bifurcation, like the discontinuity of the Julia set or the birth of hyperbolic set of large Hausdorff dimension, which are instrumental in Shishikura's theorem that the boundary of the Mandelbrot set has dimension 2.

Bedford, Smillie and Ueda [11] extended this analysis to the unfolding of a semiparabolic fixed point of multiplicity 2 in $\mathbb{C}^2$, that is, of the form

$$f_\lambda(z,w) = \big((1+\lambda)z + z^2 + \text{h.o.t.}, b_\lambda w + \text{h.o.t.}\big), \quad \text{with } |b_0| < 1. \tag{2.3}$$

In this dissipative situation, as before the Lavaurs map is a limit of iterates of the form $f^n_{\lambda_n}$, its domain is the attracting basin $\mathcal{B}$ of the origin, but its values are contained a curve: the *repelling petal* of the semiparabolic point. For polynomial automorphisms, this leads to a precise description of the discontinuity of the Julia sets J and J^+ at $\lambda = 0$. (See also Bianchi [15] for some results about the implosion of general parabolic germs.)

If (f_λ) is an arbitrary family of dissipative polynomial automorphisms, semiparabolic bifurcations (of possibly arbitrary multiplicity) occur densely in the bifurcation locus by definition. A mechanism producing homoclinic tangencies from semiparabolic implosion was designed in [53]. Besides the analysis of Lavaurs maps (which is not as precise as in the multiplicity 2 case (2.3)), this involves a construction of "critical points" in semiparabolic basins, which by definition are tangencies between unstable manifolds (associated to some given saddle point) and the foliation of the basin by strong stable manifolds. Surprisingly, this construction is based on Wiman's classical theorem on entire functions of slow growth, and requires a stronger dissipativity condition: $|\text{Jac}(f_\lambda)| < d^{-2}$ (*substantially dissipative* regime). Altogether we obtain the following theorem, which confirms a classical conjecture of Palis in this setting:

Theorem 2.6 (Dujardin and Lyubich [53]). *In a substantially dissipative family of polynomial automorphisms of $\mathbb{C}^2$, parameters with homoclinic tangencies are dense in the bifurcation locus.*

It is expected that this result holds without the substantial dissipativity assumption. Also, it is an open question whether quadratic tangencies are always created in this process. A positive answer would yield an interesting link with the quadratic family, and add further evidence to the universality of the Mandelbrot set.

2.2.3. Holomorphic maps on $\mathbb{P}^k$

The case of families of holomorphic maps on $\mathbb{P}^k$ was studied by Berteloot, Bianchi, and Dupont in [14]. Here, as in the one-dimensional case, one starts with the stability of repelling periodic points. More precisely, one has to restrict to repelling points contained in the "small Julia set" J^* (which by definition is the support of the maximal entropy mesure μ), since there can be a number of "spurious" repelling points outside J^*. Then Berteloot, Bianchi, and Dupont obtain an almost complete generalization of the results of Mañé–Sad–Sullivan, Lyubich, and DeMarco (that is, of the above equivalent conditions (i) to (v)). As before, a remaining issue is whether this notion of weak J^*-stability implies structural stability on J^*. A main difference with the 1-dimensional case is that the characterization of bifurcation in terms of currents is now essential to establish the equivalence between the remaining conditions. More precisely, the link between the instability of critical orbits and that of periodic points is provided by a formula à la Manning–Przytycki for the Lyapunov exponent of the maximal entropy measure.

We saw in Theorems 2.1 and 2.2 that the higher bifurcation currents T_{bif}^k describe accurately certain higher-codimensional phenomena in the parameter space. It seems that the distinction between T_{bif} and its powers is not as clear in higher-dimensional dynamics: in a recent work, Astorg and Bianchi [3] showed that in a large portion of the family of polynomial skew products of $\mathbb{C}^2$, the supports of all currents T_{bif}^k coincide with the bifurcation locus. So the significance of these higher bifurcation currents in this context is yet to be explored.

2.3. Robust bifurcations

As said before, due to the finiteness of the critical locus, one-dimensional polynomial and rational maps are generically stable. Intuition from real dynamics suggests that this is not anymore the case in higher dimension. As in the previous paragraph, we discuss separately the cases of polynomial automorphisms and of holomorphic maps on $\mathbb{P}^k$.

2.3.1. Polynomial automorphisms

Given the characterization of weak J^*-stability in [53], a straightforward adaptation of the one-dimensional argument for the density of stability shows that in any holomorphic family (f_λ) of polynomial automorphisms of $\mathbb{C}^2$, the union of (weakly J^*-)stable parameters together with parameters with infinitely many sinks is dense. Prior to [53], it was actually already known that stability is not a dense phenomenon in this context, due to the following remarkable result:

Theorem 2.7 (Buzzard [23]). *There exist $d > 1$ and an open subset $\Omega \subset \mathrm{Aut}_d(\mathbb{C}^2)$ contained in the bifurcation locus. In particular, maps with infinitely many sinks are dense in Ω.*

Here $\mathrm{Aut}_d(\mathbb{C}^2)$ is the space of polynomial automorphisms of $\mathbb{C}^2$ of degree d. This deep theorem is nothing but the adaptation to the complex setting of Newhouse's theorem (see [76]) on the existence of surface diffeomorphisms with persistent homoclinic tangencies. It is obtained by first constructing transcendental examples and then approximating them by polynomial ones, hence the degree d is unknown and presumably very large. The existence of this complex Newhouse phenomenon in arbitrary degree is a major open problem.

Question 2.8. Is the bifurcation locus of nonempty interior in $\mathrm{Aut}_d(\mathbb{C}^2)$ for any $d \geq 2$?

As in the real case (cf. [76]), one may even expect that robust bifurcations (that is, interior points of the bifurcation locus) are dense in the bifurcation locus, at least in the dissipative regime. For this, it is tempting to imitate the approach of Shishikura's theorem on the Hausdorff dimension of ∂M and use semiparabolic implosion to construct large bifurcation sets from a single parabolic bifurcation: in this sense the density of robust bifurcations would be the optimal generalization of Shishikura's theorem to automorphisms of $\mathbb{C}^2$. An interesting first step would be to show that the bifurcation locus has maximal Hausdorff dimension at every point. More advanced techniques will certainly be needed to get open subsets: an ambitious research program on the intersection of complex Cantor sets was initiated by Araujo, Moreira, and Zamudio towards this perspective (see [1,2]).

Biebler observed in [18] that the existence of robust bifurcations is actually more tractable in higher dimensions and showed that: *for every $d \geq 2$, the bifurcation locus has nonempty interior in* $\mathrm{Aut}_d(\mathbb{C}^3)$. This is based on a distinct mechanism for robust bifurcation, namely the *blenders* of Bonatti and Diaz [19]. These are dynamically defined Cantor sets which are so fat in a certain "direction" that they intersect an open set of curves. The point of [18] is to use this feature as a building block for persistent tangencies.

Finally, let us point out a recent beautiful result by Yampolsky and Yang [85]: *the one-dimensional family of degree 2 Hénon maps with a golden mean Siegel disk*

$$f_a(x,y) = (x^2 + c_a - ay, x),$$
$$\text{with } c_a = (1+a)\left(\frac{\mu}{2} + \frac{a}{2\mu}\right) - \left(\frac{\mu}{2} + \frac{a}{2\mu}\right)^2 \text{ and } \mu = e^{\pi(1+\sqrt{5})i},$$

is structurally unstable at every parameter with small enough Jacobian $|a|$. This relies on a completely different approach to persistent tangencies, based on Siegel renormalization.

2.3.2. Holomorphic maps on $\mathbb{P}^k$

From the work of Berteloot, Bianchi, and Dupont, we know that the basic phenomenon responsible for bifurcations for holomorphic maps on $\mathbb{P}^k$ is when the postcritical set intersects the small Julia set J^*. Thus, to obtain robust bifurcations, it is enough to find a mechanism ensuring a robust intersection between the postcritical set and J^*. A convenient tool for this is the Bonatti–Diaz blender, which leads to:

Theorem 2.9 (Dujardin [49]). *For every $k \geq 2$ and $d \geq 2$, the bifurcation locus has nonempty interior in* $\mathrm{Hol}_d(\mathbb{P}^k)$.

Here, $\mathrm{Hol}_d(\mathbb{P}^k)$ is the space of holomorphic maps on $\mathbb{P}^k$ of degree d. A specific one-dimensional family of holomorphic maps of $\mathbb{P}^2$ with a full bifurcation locus was found independently by Bianchi and Taflin [16]. After this result, a natural question is that of the abundance of robust bifurcations in $\mathrm{Hol}_d(\mathbb{P}^k)$. Taflin [83] showed that robust bifurcations are abundant near product polynomial maps of $\mathbb{C}^2$, and Biebler [17] showed that Lattès maps of sufficiently large degree are accumulated by robust bifurcations. Blenders are involved directly or indirectly in both cases, and seem to appear quite naturally when a repelling periodic point bifurcates to a saddle. Still, the general picture remains elusive.

Question 2.10. Is the bifurcation locus in $\mathrm{Hol}_d(\mathbb{P}^k)$ the closure of its interior?

Lastly, a celebrated theorem of McMullen asserts that any stable algebraic families of rational maps on $\mathbb{P}^1$ is either isotrivial or a family of flexible Lattès examples [74]. Extending this result to higher dimensions is a promising research problem; one main obstacle is that part of the argument relies on Thurston's topological characterization of rational functions. Related preliminary results have been obtained by Gauthier and Vigny [61].

3. (NON-)WANDERING FATOU COMPONENTS

The classification of Fatou components is a basic chapter of holomorphic dynamics. For rational maps in dimension 1, periodic Fatou components can be classified into attracting basins, parabolic basins, and rotation domains (Siegel disks and Herman rings). The crowning achievement of this classification is the celebrated nonwandering domain theorem of Sullivan [82]: *for a one-dimensional rational map, any Fatou component is preperiodic*.

In higher dimensions, techniques from geometric function theory may be applied to classify periodic Fatou components. It is convenient to distinguish between recurrent and nonrecurrent periodic components: a fixed Fatou component Ω is *recurrent* if for some $x \in \Omega$, the ω-limit set $\omega(x)$ is not completely contained in $\partial\Omega$. Recurrent Fatou components were classified in various classes of rational maps in [10, 56, 57, 84]. The upshot is that in such a component either there is an transversely attracting submanifold (possibly a point) or the dynamics is of rotation type. The situation is far less understood in the nonrecurrent case. A notable exception is that of substantially dissipative automorphisms of $\mathbb{C}^2$, for which it was shown by Lyubich and Peters [72] that any nonrecurrent Fatou component is the basin of a semiparabolic periodic point.

On the other hand, it is immediately clear that the quasiconformal techniques used in Sullivan's proof are not generalizable to higher dimension. As it turns out, wandering components do exist in 2-dimensional polynomial dynamics:

Theorem 3.1 (Astorg, Buff, Dujardin, Peters, and Raissy [5]). *If $0 < a < 1$ is sufficiently close to 1, the polynomial mapping of $\mathbb{C}^2$ defined by*

$$f : (z, w) \mapsto \big(p(z, w), q(w)\big) = \left(z + z^2 + az^3 + \frac{\pi^2}{4}w, w - w^2\right)$$

admits a wandering Fatou component.

The proof is based on an original idea of M. Lyubich, and relies on a skew product version of parabolic implosion. It was further implemented in other situations in [4, 63].

Sketch of proof. Write $p(z, w) = p_0(z) + \varepsilon(z, w)$, with $p_0(z) = z + z^2$ and $\varepsilon(z, w)$ being thought of as a perturbative term. Start with an initial point (z_0, w_0) such that z_0 belongs to the parabolic basin of attraction of 0 for p_0 and w_0 a small positive number, and let as usual $(z_n, w_n) = f^n(z_0, w_0)$. Then $w_n = q^n(w)$ converges to 0 along the positive real axis, and $p_0^n(z_0)$ converges to 0 along the negative real axis. Therefore $z_n = p_0^n(z_0) + \varepsilon_n$ is pushed a little faster towards the origin by the term ε_n. The terms in $\varepsilon(z, w)$ are crafted so that if z_0 is chosen carefully in some open set of initial conditions, the iterates z_n indeed pass the origin by going "through the eggbeater" and come back close to their initial position. So we can repeat this process and conclude that (z_0, w_0) belongs to some Fatou component. But since the returning time increases with the number of iterations, this Fatou component is not periodic, and we are done. ■

At this stage the following natural questions arise:

Question 3.2. (1) Are there other dynamical mechanisms leading to wandering Fatou components?

(2) Find substantial families of higher-dimensional rational mappings without wandering domains.

Regarding the first question, a mechanism for constructing wandering domains in 2-dimensional smooth dynamics, based on the Newhouse phenomenon, was devised by Colli and Vargas [27]. Berger and Biebler recently proved that this mechanism can be implemented in certain 5-dimensional families of Hénon maps, leading to the following stunning theorem:

Theorem 3.3 (Berger and Biebler [12]). *There exists a polynomial automorphism of $\mathbb{C}^2$ of degree 6 with a wandering Fatou component.*

This solves the existence problem for wandering Fatou components for plane polynomial automorphisms, which does not seem to be amenable to the techniques of [5].

For the second question, it is a classical fact that hyperbolic dynamics prevents the existence of wandering domains. Besides this observation, not much is known. In view of Theorem 3.1, it is natural to investigate the case of skew products with a fixed attracting fiber, that is, of the form

$$f(z,w) = \big(p(z), q(z,w)\big), \quad \text{with } p(0) = 0 \text{ and } \big|p'(0)\big| < 1. \tag{3.1}$$

In this case it could be expected that Sullivan's theorem, together with the attracting nature of the invariant fiber, should be enough to prevent the existence of wandering domains. Embarrassingly enough, even in such a simple situation, there is no definitive answer so far, and furthermore it was shown by Peters and Vivas [79] that the above naive intuition does not lead to a proof. Here is the current status of the problem:

Theorem 3.4 (Lilov, Peters-Smit, Ji). *If f is an attracting skew product as in* (3.1)*, then there are no wandering components near the attracting fiber, whenever:*

- $p'(0) = 0$ [68] *or, more generally, if $|p'(0)|$ is small enough (with respect to p and q)* [65]*;*
- $|p'(0)| < 1$ *and $q(0,\cdot)$ satisfies some nonuniform hyperbolicity properties* [64,78].

There is currently no hope for a general understanding of the problem of wandering Fatou components in several dimensions, and even going beyond skew products seems to be a serious challenge. An interesting first case to be considered is that of Fatou components in the neighborhood of an invariant superattracting line, which would cover, for instance, the case of regular polynomial mappings of $\mathbb{C}^2$ near the line at infinity.

ACKNOWLEDGMENTS

It is a great pleasure to thank my collaborators Eric Bedford, Pierre Berger, Serge Cantat, Bertrand Deroin, Jeff Diller, Charles Favre, Vincent Guedj, and Misha Lyubich for so many discussions about the mathematical themes presented here, and more generally my colleagues from the holomorphic dynamics community for maintaining such a friendly atmosphere over the years. Special thanks to Charles Favre and Thomas Gauthier for their helpful comments on this paper. Nessim Sibony, who introduced me to higher dimensional holomorphic dynamics, tragically passed away while this paper was in final revision. The importance of his vision for the shaping of the field can hardly be overestimated, and his ideas will remain a source of inspiration for many of us.

REFERENCES

[1] H. Araújo and C. G. Moreira, Stable intersections of conformal Cantor sets. 2019, arXiv:1910.03715.

[2] H. Araújo, C. G. Moreira, and A. Z. Espinosa, Stable intersections of regular conformal Cantor sets with large Hausdorff dimensions. 2021, arXiv:2102.07283.

[3] M. Astorg and F. Bianchi, Higher bifurcations for polynomial skew products. 2020, arXiv:2007.00770.

[4] M. Astorg, L. Boc-Thaler, and H. Peters, Wandering domains arising from Lavaurs maps with Siegel disks. *Anal. PDE*, to appear.

[5] M. Astorg, X. Buff, R. Dujardin, H. Peters, and J. Raissy, A two-dimensional polynomial mapping with a wandering Fatou component. *Ann. of Math. (2)* **184** (2016), no. 1, 263–313.

[6] M. Astorg, T. Gauthier, N. Mihalache, and G. Vigny Collet, Eckmann and the bifurcation measure. *Invent. Math.* **217** (2019), no. 3, 749–797.

[7] G. Bassanelli and F. Berteloot, Bifurcation currents in holomorphic dynamics on $\mathbb{P}^k$. *J. Reine Angew. Math.* **608** (2007), 201–235.

[8] G. Bassanelli and F. Berteloot, Lyapunov exponents, bifurcation currents and laminations in bifurcation loci. *Math. Ann.* **345** (2009), no. 1, 1–23.

[9] E. Bedford, M. Lyubich, and J. Smillie, Polynomial diffeomorphisms of $\mathbf{C}^2$. IV. The measure of maximal entropy and laminar currents. *Invent. Math.* **112** (1993), no. 1, 77–125.

[10] E. Bedford and J. Smillie, Polynomial diffeomorphisms of $\mathbf{C}^2$. II. Stable manifolds and recurrence. *J. Amer. Math. Soc.* **4** (1991), no. 4, 657–679.

[11] E. Bedford, J. Smillie, and T. Ueda, Semi-parabolic bifurcations in complex dimension two. *Comm. Math. Phys.* **350** (2017), no. 1, 1–29.

[12] P. Berger and S. Biebler, Emergence of wandering stable components. 2020, arXiv:2001.08649.

[13] P. Berger and R. Dujardin, On stability and hyperbolicity for polynomial automorphisms of $\mathbb{C}^2$. *Ann. Sci. Éc. Norm. Supér. (4)* **50** (2017), no. 2, 449–477.

[14] F. Berteloot, F. Bianchi, and C. Dupont, Dynamical stability and Lyapunov exponents for holomorphic endomorphisms of $\mathbb{P}^k$. *Ann. Sci. Éc. Norm. Supér. (4)* **51** (2018), no. 1, 215–262.

[15] F. Bianchi, Parabolic implosion for endomorphisms of $\mathbb{C}^2$. *J. Eur. Math. Soc. (JEMS)* **21** (2019), no. 12, 3709–3737.

[16] F. Bianchi and J. Taflin, Bifurcations in the elementary Desboves family. *Proc. Amer. Math. Soc.* **145** (2017), no. 10, 4337–4343.

[17] S. Biebler, Lattès maps and the interior of the bifurcation locus. *J. Mod. Dyn.* **15** (2019), 95–130.

[18] S. Biebler, Newhouse phenomenon for automorphisms of low degree in $\mathbb{C}^3$. *Adv. Math.* **361** (2020), 106952, 39.

[19] C. Bonatti and L. J. Díaz, Persistent nonhyperbolic transitive diffeomorphisms. *Ann. of Math. (2)* **143** (1996), no. 2, 357–396.

[20] M. Brunella, Courbes entières et feuilletages holomorphes. *Enseign. Math. (2)* **45** (1999), no. 1–2, 195–216.

[21] X. Buff and A. Chéritat, Quadratic Julia sets with positive area. In *Proceedings of the International Congress of Mathematicians. III*, pp. 1701–1713, Hindustan Book Agency, New Delhi, 2010.

[22] X. Buff and A. Epstein, Bifurcation measure and postcritically finite rational maps. In *Complex dynamics*, pp. 491–512, A K Peters, Wellesley, MA, 2009.

[23] G. T. Buzzard, Infinitely many periodic attractors for holomorphic maps of 2 variables. *Ann. of Math. (2)* **145** (1997), no. 2, 389–417.

[24] S. Cantat, Dynamique des automorphismes des surfaces $K3$. *Acta Math.* **187** (2001), no. 1, 1–57.

[25] S. Cantat, Automorphisms and dynamics: a list of open problems. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. II. Invited lectures*, pp. 619–634, World Sci. Publ., Hackensack, NJ, 2018.

[26] S. Cantat and R. Dujardin, Random dynamics on real and complex projective surfaces. 2020, arXiv:2006.04394.

[27] E. Colli and E. Vargas, Non-trivial wandering domains and homoclinic bifurcations. *Ergodic Theory Dynam. Systems* **21** (2001), no. 6, 1657–1681.

[28] H. de Thélin, Sur la laminarité de certains courants. *Ann. Sci. Éc. Norm. Supér. (4)* **37** (2004), no. 2, 304–311.

[29] H. de Thélin, Sur la construction de mesures selles. *Ann. Inst. Fourier (Grenoble)* **56** (2006), no. 2, 337–372.

[30] H. de Thélin, Un critère de laminarité locale en dimension quelconque. *Amer. J. Math.* **130** (2008), no. 1, 187–205.

[31] J.-P. Demailly, Complex analytic and differential geometry. https://www-fourier.ujf-grenoble.fr/~demailly/manuscripts/agbook.pdf.

[32] L. DeMarco, Dynamics of rational maps: a current on the bifurcation locus. *Math. Res. Lett.* **8** (2001), no. 1–2, 57–66.

[33] L. DeMarco, Dynamics of rational maps: Lyapunov exponents, bifurcations, and capacity. *Math. Ann.* **326** (2003), no. 1, 43–73.

[34] L. DeMarco, Critical orbits and arithmetic equidistribution. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Volume III. Invited lectures*, pp. 1867–1886, World Sci. Publ., Hackensack, NJ, 2018.

[35] B. Deroin and R. Dujardin, Random walks, Kleinian groups, and bifurcation currents. *Invent. Math.* **190** (2012), no. 1, 57–118.

[36] H. De Thélin, T. Gauthier, and G. Vigny, Parametric Lyapunov exponents. *Bull. Lond. Math. Soc.* **53** (2021), no. 3, 660–672.

[37] J. Diller, R. Dujardin, and V. Guedj, Dynamics of meromorphic maps with small topological degree III: geometric currents and ergodic theory. *Ann. Sci. Éc. Norm. Supér. (4)* **43** (2010), no. 2, 235–278.

[38] J. Diller, R. Dujardin, and V. Guedj, Dynamics of meromorphic mappings with small topological degree II: energy and invariant measure. *Comment. Math. Helv.* **86** (2011), no. 2, 277–316.

[39] T.-C. Dinh, Suites d'applications méromorphes multivaluées et courants laminaires. *J. Geom. Anal.* **15** (2005), no. 2, 207–227.

[40] T.-C. Dinh, Pluripotential theory and complex dynamics in higher dimension. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Volume III. Invited lectures*, pp. 1561–1581, World Sci. Publ., Hackensack, NJ, 2018.

[41] T.-C. Dinh, V.-A. Nguyen, and N. Sibony, Unique ergodicity for foliations on compact Kähler surfaces. 2018, arXiv:1811.07450.

[42] T.-C. Dinh and N. Sibony, Green currents for holomorphic automorphisms of compact Kähler manifolds. 2003, arXiv:math/0311322.

[43] R. Dujardin, Laminar currents in $\mathbb{P}^2$. *Math. Ann.* **325** (2003), no. 4, 745–765.

[44] R. Dujardin, Sur l'intersection des courants laminaires. *Publ. Mat.* **48** (2004), no. 1, 107–125.

[45] R. Dujardin, Laminar currents and birational dynamics. *Duke Math. J.* **131** (2006), no. 2, 219–247.

[46] R. Dujardin, Fatou directions along the Julia set for endomorphisms of $\mathbb{CP}^k$. *J. Math. Pures Appl. (9)* **98** (2012), no. 6, 591–615.

[47] R. Dujardin, The supports of higher bifurcation currents. *Ann. Fac. Sci. Toulouse Math. (6)* **22** (2013), no. 3, 445–464.

[48] R. Dujardin, Bifurcation currents and equidistribution in parameter space. In *Frontiers in complex dynamics*, pp. 515–566, Princeton Math. Ser. 51, Princeton Univ. Press, Princeton, NJ, 2014.

[49] R. Dujardin, Non-density of stability for holomorphic mappings on $\mathbb{P}^k$. *J. Éc. Polytech. Math.* **4** (2017), 813–843.

[50] R. Dujardin, Saddle hyperbolicity implies hyperbolicity for polynomial automorphisms of $\mathbb{C}^2$. *Math. Res. Lett.* **27** (2020), no. 3, 693–709.

[51] R. Dujardin and C. Favre, Distribution of rational maps with a preperiodic critical point. *Amer. J. Math.* **130** (2008), no. 4, 979–1032.

[52] R. Dujardin and C. Favre, The dynamical Manin–Mumford problem for plane polynomial automorphisms. *J. Eur. Math. Soc. (JEMS)* **19** (2017), no. 11, 3421–3465.

[53] R. Dujardin and M. Lyubich, Stability and bifurcations for dissipative polynomial automorphisms of $\mathbb{C}^2$. *Invent. Math.* **200** (2015), no. 2, 439–511.

[54] C. Favre and T. Gauthier, Distribution of postcritically finite polynomials. *Israel J. Math.* **209** (2015), no. 1, 235–292.

[55] C. Favre and T. Gauthier, The arithmetic of polynomial dynamical pairs. 2020, arXiv:2004.13801.

[56] J. E. Fornæss and F. Rong, Classification of recurrent domains for holomorphic maps on complex projective spaces. *J. Geom. Anal.* **24** (2014), no. 2, 779–785.

[57] J. E. Fornæss and N. Sibony, Classification of recurrent domains for some holomorphic maps. *Math. Ann.* **301** (1995), no. 4, 813–820.

[58] J. E. Fornæss and N. Sibony, Harmonic currents of finite energy and laminations. *Geom. Funct. Anal.* **15** (2005), no. 5, 962–1003.

[59] T. Gauthier, Strong bifurcation loci of full Hausdorff dimension. *Ann. Sci. Éc. Norm. Supér. (4)* **45** (2012), no. 6, 947–984.

[60] T. Gauthier, Y. Okuyama, and G. Vigny, Hyperbolic components of rational maps: quantitative equidistribution and counting. *Comment. Math. Helv.* **94** (2019), no. 2, 347–398.

[61] T. Gauthier and G. Vigny, The geometric dynamical Northcott and Bogomolov properties. 2019, arXiv:1912.07907.

[62] J. Graczyk and G. Świątek, Harmonic measure and expansion on the boundary of the connectedness locus. *Invent. Math.* **142** (2000), no. 3, 605–629.

[63] D. Hahn and H. Peters, A polynomial automorphism with a wandering Fatou component. *Adv. Math.* **382** (2021), 107650, 46.

[64] Z. Ji, Non-uniform hyperbolicity in polynomial skew products. 2019, arXiv:1909.06084.

[65] Z. Ji, Non-wandering Fatou components for strongly attracting polynomial skew products. *J. Geom. Anal.* **30** (2020), no. 1, 124–152.

[66] L. Kaufmann, Self-intersection of foliation cycles on complex manifolds. *Internat. J. Math.* **28** (2017), no. 8, 1750054, 18.

[67] T. Lei, Hausdorff dimension of subsets of the parameter space for families of rational maps. (A generalization of Shishikura's result). *Nonlinearity* **11** (1998), no. 2, 233–246.

[68] K. Lilov, *Fatou theory in two complex dimensions*. Ph.D. thesis, University of Michigan, 2004.

[69] M. Lyubich, Some typical properties of the dynamics of rational mappings. *Uspekhi Mat. Nauk* **38** (1983), no. 5(233), 197–198.

[70] M. Lyubich, Investigation of the stability of the dynamics of rational functions. *Teor. Funkc. Funkc. Anal. Ih Prilozh.* **42** (1984), 72–91.

[71] M. Lyubich, Analytic low-dimensional dynamics: from dimension one to two. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Volume 1*, pp. 443–474, Kyung Moon Sa, Seoul, 2014.

[72] M. Lyubich and H. Peters, Classification of invariant Fatou components for dissipative Hénon maps. *Geom. Funct. Anal.* **24** (2014), no. 3, 887–915.

[73] R. Mañé, P. Sad, and D. Sullivan, On the dynamics of rational maps. *Ann. Sci. Éc. Norm. Supér. (4)* **16** (1983), no. 2, 193–217.

[74] C. T. McMullen, Families of rational maps and iterative root-finding algorithms. *Ann. of Math. (2)* **125** (1987), no. 3, 467–493.

[75] C. T. McMullen, The Mandelbrot set is universal. In *The Mandelbrot set, theme and variations*, pp. 1–17, London Math. Soc. Lecture Note Ser. 274, Cambridge Univ. Press, Cambridge, 2000.

[76] S. E. Newhouse, The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms. *Publ. Math. Inst. Hautes Études Sci.* **50** (1979), 101–151.

[77] K. Oguiso, Some aspects of explicit birational geometry inspired by complex dynamics. In *Proceedings of the International Congress of Mathematicians—Seoul 2014. Volume II*, pp. 695–721, Kyung Moon Sa, Seoul, 2014.

[78] H. Peters and I. M. Smit, Fatou components of attracting skew-products. *J. Geom. Anal.* **28** (2018), no. 1, 84–110.

[79] H. Peters and L. R. Vivas, Polynomial skew-products with wandering Fatou-disks. *Math. Z.* **283** (2016), no. 1–2, 349–366.

[80] M. Shishikura, The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets. *Ann. of Math. (2)* **147** (1998), no. 2, 225–267.

[81] S. Smirnov, Symbolic dynamics and Collet–Eckmann conditions. *Int. Math. Res. Not.* **7** (2000), 333–351.

[82] D. Sullivan, Quasiconformal homeomorphisms and dynamics. I. Solution of the Fatou–Julia problem on wandering domains. *Ann. of Math. (2)* **122** (1985), no. 3, 401–418.

[83] J. Taflin, Blenders near polynomial product maps of $\mathbb{C}^2$. *J. Eur. Math. Soc. (JEMS)* (2021).

[84] T. Ueda, Holomorphic maps on projective spaces and continuations of Fatou maps. *Michigan Math. J.* **56** (2008), no. 1, 145–153.

[85] M. Yampolsky and J. Yang, Structural instability of semi-Siegel Hénon maps. *Adv. Math.* **389** (2021), 107900.

ROMAIN DUJARDIN

Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation, 4 place Jussieu, 75005 Paris, France, romain.dujardin@sorbonne-universite.fr

RIGIDITY, LATTICES, AND INVARIANT MEASURES BEYOND HOMOGENEOUS DYNAMICS

DAVID FISHER

ABSTRACT

This article discusses two recent works by the author, one with Brown and Hurtado on Zimmer's conjecture and one with Bader, Miller, and Stover on totally geodesic submanifolds of real and complex hyperbolic manifolds. The main purpose of juxtaposing these two very disparate sets of results in one article is to emphasize a common aspect: that the study of invariant and partially invariant measures outside the homogeneous setting is important to questions about rigidity in geometry and dynamics. I will also discuss some open questions including some that seem particularly compelling in light of this juxtaposition.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 22E40; Secondary 37C75, 37A17, 53C35

KEYWORDS

Rigitidy, invariant measures, lattices in Lie groups

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3484–3507
DOI 10.4171/ICM2022/147

1. INTRODUCTION

This article focuses on some recent developments concerning the rigidity of discrete subgroups of Lie groups. This area has long had deep connections with ergodic theory and dynamical systems going back to seminal work of Furstenberg, Margulis, Mostow, and Zimmer [44,62,63,68,91]. Here we emphasize the role of invariant measures for groups and their subgroups. From one point of view, a key step in Margulis' proof of his superrigidity theorem can be written as finding an invariant measure for a subgroup in a group action. Invariant measures have also long been a key object of study in homogeneous dynamics. Rigidity of discrete groups and homogeneous dynamics have long been allied and interacting fields, but the developments here are part of a strengthening of those connections, particularly in terms of proving rigidity results directly through the study of invariant measures in inhomogeneous settings, using techniques, ideas, and results from homogeneous dynamics. In particular, both results study dynamical systems with homogeneous factors and the dynamics on the homogeneous factor help control the dynamics on the total system.

Let H be a Lie group, $\Lambda < H$ a discrete subgroup, and $S < H$ a subgroup. Homogeneous dynamics is most often the study of invariant measures, orbit closures, and equidistribution for the S action on H/Λ. This area has been quite fruitful with applications to areas as diverse as number theory, geometry, and physics. Here we are more often concerned with dynamical systems where the space H/Λ is replaced by a space that is not homogeneous. Perhaps the most famous example of this is the action of $\mathrm{SL}(2,\mathbb{R})$ on the moduli space of quadratic differentials on a surface. This area is not our topic here, but the fruitful importation of ideas from homogeneous dynamics to this area has a long history, starting with work of Veech and currently culminating in the work of Eskin, Mirzakhani, and Mohammadi [32,33,84]. Some ideas arising in this setting have been pushed even further into the inhomogeneous world by work of Brown–Rodriguez Hertz and ongoing work of Brown–Eskin–Filip [21]. More closely allied with the developments here is work of Katok, Kalinin, and Rodriguez Hertz on invariant measures for actions of higher rank abelian groups on compact manifolds [54].

I want to point to one other principle that has played a key role in many developments, which is the notion of stiffness, first formalized by Furstenberg [45]. When studying actions of amenable groups, it suffices to consider invariant measures. When the acting group S is not amenable, to understand the dynamics it is important to study the broader class of stationary measures. Furstenberg called an action of a group S on a space X *stiff* if every stationary measure for S was in fact invariant. Even before Furstenberg defined the term, Nevo and Zimmer had considered the case where S is a higher rank simple Lie group acting on an arbitrary measure space X and given criteria for stiffness in terms of measurable projective quotients [71]. More recently, Benoist and Quint proved dramatic results on stiffness in the homogeneous setting that were inspirational for the work of Eskin–Mirzakhani mentioned above [12].

In this article we point to directions where homogeneous dynamics techniques are applied outside the homogeneous setting to prove rigidity results about discrete groups and their actions, and in particular questions where we need to move beyond the question of stiffness. The first work I will describe, joint with Brown and Hurtado, resolves important cases of Zimmer's conjecture. A key ingredient in our work is to move beyond stiffness and consider an even larger class of measures than the stationary ones. In this context, the stationary measures can be made to correspond to the invariant ones for some subgroup $P < S$ but our proof requires understanding something about the invariant measures for a much smaller subgroup $A < P$. The other results I will focus on concern totally geodesic submanifolds of real and complex hyperbolic manifolds of finite volume and are joint works with Bader, Lafont, Miller, and Stover. In this setting, once again a key object is to construct certain measures that are invariant under subgroups of the acting group. In this work, the class of measures studied is simply different than the stationary ones, not more or less general.

Overall, I think the results mentioned above, as well as numerous results by other authors, point to the development of a broad area of research in which the study of invariant measures beyond homogeneous dynamics has broad implications for rigidity theory, and that many of these developments will spring from broadening the efficacy of ideas that originate in homogeneous dynamics. It feels too early to attempt a survey of these developments, so I will restrict myself to an account of these developments in my own work. Along the way I will point to open questions inspired by that work, some of which fits this introductory framework and some of which do not.

2. ZIMMER'S CONJECTURE AND THE ZIMMER PROGRAM

In this section I will discuss some aspects of recent work with Brown and Hurtado on Zimmer's conjecture and also discuss the implications for further work. In the course of this, I will point to some work in progress with Melnick concerning examples and also point to an old paper of Uchida which has important implications for the Zimmer program and which seems too little known. For a different take on some results and questions discussed here, see Brown's contribution in these proceedings [17].

2.1. Zimmer's conjecture

Throughout this subsection G will be a simple Lie group with real rank at least 2 and $\Gamma < G$ will be a lattice. The reader will lose little by considering the case of $G = \mathrm{SL}(n, \mathbb{R})$ and $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ with $n > 2$. In [92, 93], Zimmer laid out a program for understanding Γ actions on a compact manifold M preserving a volume form ω. The base case of this program is to show that for any homomorphism $\rho : \Gamma \to \mathrm{Diff}(M, \omega)$, the image $\rho(\Gamma)$ always preserves a smooth Riemannian metric if $\dim(M)$ is less than the dimension of the minimal real G representation. This conjecture was motivated by Zimmer's cocycle superrigidity theorem, which in this context produced a measurable Riemannian metric whose associated volume form was ω.

Already in the 1990s, this conjecture about low-dimensional actions had been transported out of the volume-preserving setting by numerous works, particularly concerning actions on the circle, see, e.g., [31,49]. The work with Brown and Hurtado proved this more general conjecture in many cases, here I state only the following special case.

Theorem 2.1 (Brown, Fisher, Hurtado). *Let Γ be a lattice in $\mathrm{SL}(n,\mathbb{R})$, let M be a compact manifold and let $\rho : \Gamma \to \mathrm{Diff}(M)$ be a homomorphism. Then*

(1) *if $\dim(M) < n-1$, the image of ρ is finite;*

(2) *if $\dim(M) < n$ and $\rho(\Gamma)$ preserves a volume form on M, then the image of ρ is finite.*

While I include a very rough sketch of ideas in the proof, more detailed outlines can be found in [16,39,40]. A key step is showing that Γ acts with *subexponential growth of derivatives*. We let l be any word length on the group Γ.

Definition 2.2. Let $\rho : \Gamma \to \mathrm{Diff}(M)$ be an action of a finitely generated group Γ on a compact manifold M. We say ρ has *subexponential growth of derivatives* if for every $\epsilon > 0$ there exists $C > 0$ such that

$$\sup_{x\in M} \|D\rho(\gamma)_x\| < Ce^{\epsilon l(\gamma)}.$$

The idea that controlling the growth of derivatives is pivotal was long known, see the discussion in [39, END OF SECTION 7]. However, a particular novelty introduced in [18] is that we use the strong property (T) of Lafforgue to convert subexponential growth of derivatives to an invariant Riemannian metric without any further hypothesis. For much more on strong property (T), see de la Salle's contribution in these proceedings [24].

The approach to proving subexponential growth of derivatives in [18–20] was also distinct from previous ideas on Zimmer's conjecture but did have some classical inspirations as well as a more closely related one in [52]. We give a somewhat ad hoc definition of zero Lyaponov exponents for a group action here, other definitions are possible and most are somewhat weaker than this, but this one suffices for current purposes.

Definition 2.3. Let $\rho : \Gamma \to \mathrm{Diff}(M)$ be an action of a finitely generated group on a compact manifold. Let μ be a measure on M. We say ρ has *zero first Lyapunov exponent* for μ if

$$\lim_{l(\gamma)\to\infty} \frac{\ln \|D\rho(\gamma)_x\|}{l(\gamma)} = 0$$

for μ almost every x in M.

Essentially, the proof of Theorem 2.1 consists of showing that exponential growth of derivatives must be witnessed by a positive Lyapunov exponent for some Γ invariant measure and then seeing that this contradicts Zimmer's cocycle superrigidity theorem. The motivation is probably most easily encapsulated by this classical proposition.

Proposition 2.4. *Let M be a compact manifold and let $\rho : \mathbb{Z} \to \mathrm{Diff}(M)$ be the action generated by a single diffeomorphism f. Then ρ has subexponential growth of derivatives if and only if ρ has zero first Lyapunov exponent for every f-invariant measure μ.*

The proposition is not easily adapted to group actions partly because if the group is not amenable, there may be no invariant measures at all. One attempt to remedy this would be to consider stationary measures and random Lyapunov exponents, but there is no useful analogue of the proposition in that context. The issue that arises is that the random Lyapunov exponent might vanish while exponential growth of derivatives still occurs along a very thin set of trajectories in the acting group. There is a weaker statement that is true and even useful in some contexts that is implicit in [52, SECTION 3]. That argument does produce a measure on a skew product over a shift space. However, the measure produced when projected to the shift space is not independent and identically distributed but quite arbitrary and does not fall into the usual context of random dynamics, stationary measures, and stiffness.

For the work on Zimmer's conjecture, we take a long detour which I will only describe a small part of in the next subsection, to make clear the connections to homogeneous dynamics and also to make clear why it is not enough to prove stiffness of the action.

2.2. Measures in the proof of Zimmer's conjecture

The first step in the proof of Theorem 2.1 uses the notion of an induced action, a variant of induced representations due to Mackey. This notion is also similar to the construction of flat bundles. If Γ acts on a manifold M via a homomorphism $\rho : \Gamma \to \mathrm{Diff}(M)$, then we can build a G action on a manifold $(G \times M)/\Gamma$. This can be specified just by specifying commuting G and Γ actions on $G \times M$. We do this by the formula

$$g(g_0, m)\gamma = (gg_0\gamma^{-1}, \rho(\gamma)m).$$

Note that there is a G-equivariant map $\pi : (G \times M)/\Gamma \to G/\Gamma$ and this map exhibits $(G \times M)/\Gamma$ as a fiber bundle over G/Γ with fiber M. Also note that the tangent bundle to $(G \times M)/\Gamma$ admits a G-invariant subbundle consisting of directions tangent to fibers of the projection π, i.e., $(G \times TM)/\Gamma \subset T(G \times M)/\Gamma$.

We then define *fiberwise zero first Lyapunov exponent* and *fiberwise subexponential growth of derivatives* for the G action on $(G \times M)/\Gamma$ by restricting all derivatives in the definitions to the invariant subbundle $(G \times TM)/\Gamma \subset T(G \times M)/\Gamma$. It is a relatively easy exercise to see that if $\Gamma < G$ is cocompact, then the subexponential growth of derivatives for the Γ action is equivalent to the fiberwise subexponential growth of derivatives for the G action on $(G \times M)/\Gamma$. The situation when G/Γ has finite volume but is not compact is considerably more complicated, and we do not discuss it here. At this point, the structure of Lie groups begins to play an important role. It turns out that G can always be written as a product KAK where K is compact and A is abelian. For $\mathrm{SL}(n, \mathbb{R})$, these groups are $K = \mathrm{SO}(n)$ and A the group of diagonal matrices of determinant one. Since K is compact, we can average any Riemannian metric on $(G \times M)/\Gamma$ over the K action and obtain a K-invariant metric. This means that for the action of G any growth of derivatives that we see comes entirely from the action of A. Modifying the proof of Proposition 2.4 and retaining the notation and terminology above, we prove

Lemma 2.5. *Given $\rho : \Gamma \to \mathrm{Diff}(M)$ then either ρ has subexponential growth of derivatives or there is an A-invariant measure μ on $(G \times M)/\Gamma$ with nonzero fiberwise first Lyapunov exponent for some element a in A.*

If μ were in fact G-invariant, then this can be seen to contradict Zimmer's cocycle superrigidity theorem. A key point that makes it possible to prove Lemma 2.5 is that A is abelian and so amenable.

We proceed by proving that μ can be replaced by a measure that is in fact G-invariant. This is done in two steps. First, we average the measure over certain subgroups of G to produce a measure μ' whose projection $\pi_*\mu'$ to G/Γ is Haar measure. The difficulty here is to do the averaging while retaining that μ' is A-invariant and that some a in A has positive first Lyapunov exponent for μ'. After this step, we can use a result of Brown, Rodriguez Hertz, and Wang, together with some algebraic computations, to show that μ' is in fact G-invariant [22]. This contradiction shows that ρ does in fact have subexponential growth of derivatives.

The step of averaging the measure μ to produce μ' makes extensive use of homogeneous dynamics and, in particular, work of Ratner and Shah [75,76,80]. The work of Brown, Rodriguez Hertz, and Wang pivots on relations between invariant measures and entropy, and in particular on an extension of the important work of Ledrappier and Young [57]. A key ingredient in both parts is the theory of Lyapunov exponents and, particularly, the fact that for actions of an abelian group A, Lyapunov exponents give rise to linear functionals on A.

In this context, stationary measures can be thought of roughly as just measures invariant under P and not all of G. It turns out not to be possible to start a proof by considering only P-invariant measures, instead of the wider class of A-invariant measures. A priori, exponential growth of derivatives is only witnessed on some sequence g_n of elements in G and a naive rewriting of the proof of Proposition 2.4 only gives an exponent along that particular sequence. One can choose this sequence to be in P by essentially the same reasoning by which we choose it to be in A. But at that point, it is then very hard to proceed since Lyapunov exponents on P do not a priori have any structure analogous to their structure as linear functionals on A. A main observation is that one can in fact choose g_n to be in A and use this to produce Lyapunov exponent first for a measure invariant under a 1-parameter subgroup of A and then by averaging under all of A. While our argument has been rewritten by An, Brown, and Zhang as then producing a P-invariant measure from this A-invariant measure, at this step one has to argue using a great deal of homogeneous dynamics and using a rather intricate averaging procedure [4].

2.3. Other results, future directions

The joint work with Brown and Hurtado classifies actions of lattices in $\mathrm{SL}(n,\mathbb{R})$ in dimension at most $n-2$ and volume-preserving actions in dimension $n-1$. A more recent result of Brown, Rodriguez Hertz, and Wang completes the picture through dimension $n-1$, showing that in general an action of a lattice in $\mathrm{SL}(n,\mathbb{R})$ on an $(n-1)$-dimensional manifold either factors through a finite group or extends to an $\mathrm{SL}(n,\mathbb{R})$ action. It is easy to see that there are exactly two actions of $\mathrm{SL}(n,\mathbb{R})$ on $(n-1)$-manifolds, namely the action on $\mathbb{P}(\mathbb{R}^n)$ and lift of that action to S^{n-1}. Clearly, a natural question would be to also classify actions in dimension n, but this becomes surprisingly harder, mainly because there are many more examples.

Already actions of $\mathrm{SL}(n,\mathbb{R})$ on n-manifolds are quite complicated and not fully classified. A remarkable and little known paper of Uchida classifies analytic actions of $\mathrm{SL}(n,\mathbb{R})$ on S^n, and the parameter space turns out to be infinite-dimensional. In a work in progress with Melnick, we are extending this to a classification of all analytic actions of $\mathrm{SL}(n,\mathbb{R})$ on n-manifolds and may also produce a smooth classification, though there are missing ingredients at the moment. For lattice actions, there are more examples known, constructed first by Katok and Lewis [56] and studied by the author with various coauthors [13, 43]. In addition, some ideas from the work of Uchida allow us to adapt a continuous construction described by Farb and Shalen and show that it can be done analytically [35, 83].

The current conjectural picture of actions of $\mathrm{SL}(n,\mathbb{R})$ lattices on n-manifolds that results from these developments is quite complicated, and we will not attempt to describe it here. Instead, we state a conjecture in dimension n about stiff actions.

Conjecture 2.6. *Let $\Gamma < \mathrm{SL}(n,\mathbb{R})$ be a lattice and assume Γ acts on an n-manifold stiffly. Then either the action factors through a finite group or the action lifts to a finite cover where it is smoothly conjugate to an affine action of a finite index subgroup of $\mathrm{SL}(n,\mathbb{Z})$ on $\mathbb{T}^n$.*

For nonstiff actions, one can formulate a conjecture about actions of lattices in $\mathrm{SL}(n,\mathbb{R})$ on n-manifolds but the statement becomes quite involved. The pivotal fact that one would need to even begin classifying actions is summarized by

Conjecture 2.7. *Let $\Gamma < \mathrm{SL}(n,\mathbb{R})$ be a lattice and assume Γ acts on an n manifold M. Assume that the induced action on $(G \times M)/\Gamma$ admits a P-invariant measure than is not G-invariant. Then the support of this measure is of the form $(G \times N)/\Gamma$ where N is an embedded $(n-1)$-sphere or $\mathbb{P}(\mathbb{R}^n)$ in M.*

One can even weaken the hypotheses to consider A-invariant measures that are not G-invariant, and it seems that the only additional possibility will be Haar measure on closed A orbits in $(G \times M)/\Gamma$. While I do not state the full conjecture here, I do hope to include it in a future work. An added difficulty is that it seems one can obtain analytic actions with an open subset where the action is analytically conjugate to the one that extends to the action of $\mathrm{SL}(n,\mathbb{R})$ on $\mathbb{R}^n \backslash \{0\}$. These subsets do not support any invariant probability measures and new ideas are definitely needed to capture this behavior in a classification. Another key step in completing a classification would involve designing an equivariant surgery that cuts along the invariant N from Conjecture 2.7 and simplifies the resulting manifold and action. A full classification may be considerably easier if one assumes the action is volume-preserving. This rules out the "bad" open sets just described and would also considerably simplify the required surgery operations.

In fact, in the current state of knowledge, one expects that a variant of Conjecture 2.6 might hold much more generally. For this, we require a definition from [42].

Definition 2.8.

1. Let A and D be topological groups, and $B < A$ a closed subgroup. Let $\rho : D \times A/B \to A/B$ be a continuous action. We call ρ *affine*, if, for every $d \in D$ there is a continuous automorphism L_d of A and an element $t_d \in A$ such that $\rho(d)[a] = [t_d \cdot L_d(a)]$.

2. Let A and B be as above. Let C and D be two commuting groups of affine diffeomorphisms of A/B, with C compact. We call the action of D on $C\backslash A/B$ a *generalized affine action*.

3. Let A, B, D, and ρ be as in case 1 above. Let M be a compact Riemannian manifold and $\iota : D \times A/B \to \mathrm{Isom}(M)$ a C^1 cocycle. We call the resulting skew product D action on $A/B \times M$ a *quasiaffine action*. If C and D are as in case 2, and $\alpha : D \times C\backslash A/B \to \mathrm{Isom}(M)$ is a C^1 cocycle, then we call the resulting skew product D action on $C\backslash A/B \times M$ a *generalized quasiaffine action*.

We note that generalized quasiaffine actions for D a higher-rank simple group or a lattice in such a group might be more constrained than it first appears. It seems at least possible that there are considerable restrictions on the cocycle into $\mathrm{Isom}(M)$ defining the skew product. Very partial results in this direction are obtained by Witte Morris and Zimmer in [67].

Question 2.9. Assume G is a higher-rank simple Lie group and $\Gamma < G$ is a lattice and M is a compact manifold. Let $\rho : \Gamma \to \mathrm{Diff}(M)$ be a stiff action. Is the action generalized quasiaffine?

A negative answer to Question 2.9 would require a genuinely new idea. All existing constructions of actions which are not generalized quasiaffine involve cutting and pasting along certain singular divisors and these singular divisors always carry stationary measures that are not invariant.

All of our understanding of these examples suggest the following

Conjecture 2.10. *Assume the setup of Question 2.9. Instead of assuming stiffness, assume that the action preserves a rigid geometric structure or that a single element admits a dominated splitting, then the action is generalized quasiaffine.*

For all of the conjectures and questions just mentioned, one might try to start with P-invariant measures in the induced action, but the proof of Zimmer's conjecture does indicate that it is perhaps more fruitful to start by studying A-invariant measures. It also indicates that one intermediate goal might be producing some uniform hyperbolicity, such as a dominated splitting. Subexponential growth of derivatives is exactly the uniform absence of hyperbolicity. The best evidence for the dynamical form of this conjecture is in dimension n in a recent paper of my student Homin Lee [58]. There is more plentiful evidence for the geometric form, going back to results of Zimmer, but the question remains mostly open see [37, SECTION 6].

In addition, one might ask for some analogue of Conjecture 2.7 in higher dimensions. The analogue might well have the identical statement but it is not as clear what N should occur here. Examples first constructed by Benveniste show that the cutting and pasting may occur along much more complicated submanifolds in general, see, e.g., [13, 36]. In these examples, the P-invariant but not G-invariant measures are in fact supported on more complicated sets and not on manifolds of the form G/Q where Q is a parabolic of G. Instead, one gets, for example, sets that are G/Q bundles over G/Γ for some lattice. One can also build examples where the cutting and pasting occurs along a high-dimensional sphere or projective space and the P-invariant measures are supported on a very low-dimensional submanifold of that space simply by doing the Katok–Lewis example for some large value of N and restricting the action to $\mathrm{SL}(n,\mathbb{Z}) < \mathrm{SL}(N,\mathbb{Z})$ for some $3 \leq n \ll N$. In these examples one will also end up with invariant open subsets where the induced action admits no stationary probability measure.

Because of this complexity, it quickly becomes hard to state a general conjecture in high dimensions precisely. A formulation favored by Zimmer and later Labourie is that the action is homogeneous on an open dense set or perhaps even built of locally homogeneous pieces. For a different conjecture, concerning ways in which one might expect the Γ action to extend locally to a G action, see [38, CONJECTURE 5.6].

Before ending this section, I will recall that while the results we can prove for Zimmer's conjecture are sharp for lattices $\mathrm{SL}(n,\mathbb{R})$ and $\mathrm{Sp}(2n,\mathbb{R})$ they are not sharp for lattices in other simple and semisimple Lie groups. There are two obstructions that arise, each of which is serious. For nonsplit groups, including even $\mathrm{SL}(n,\mathbb{C})$ and $\mathrm{SL}(n,\mathbb{H})$, there is an issue arising where we employ the work of Brown, Rodriguez Hertz, and Wang [22]. That work only "sees" the number of roots of a simple Lie group and not the dimensions of the root subspaces. To see the impact on dimensions where we can prove a result, recall that any simple Lie group G contains a maximal $\mathbb{R}$-split subgroup G' of the same $\mathbb{R}$-rank. If we let Q be a maximal parabolic of highest dimension G and Q' a maximal parabolic of highest dimension in G', then one expects that for all lattices $\Gamma < G$ that all smooth actions on compact manifolds are isometric below $\dim(G/Q)$, but our methods only prove this for $\dim(G'/Q')$. To see the effect in practice, one should consider, say, $\mathrm{SO}(m,n)$ for $m < n$ and note that the maximal $\mathbb{R}$-split subgroup is $\mathrm{SO}(m, m+1)$. This shows that this gap between expected results and what we can prove can be arbitrarily large. For $\mathbb{C}$-split groups, An, Brown, and Zhang have announced a remedy to this issue, but it appears their solution for that case is not sufficiently robust to overcome the problem in general [4].

There is another gap that arises from the fact that our proofs seem, in most cases, to be much better in the context where one does not preserve a volume form. Our techniques always only manage to constrain volume preserving actions in one dimension more than they preserve non-volume-preserving ones. This is because all we are using about volume-preserving actions is the single linear condition an invariant volume form imposes on Lyapunov exponents. Conjecturally, all volume-preserving actions should be isometric below the dimension d of the minimal G representation, while all actions should only be isometric below the dimension of G/Q where Q is a maximal parabolic of largest dimension. For the real split form of E_8, these numbers are 248 and 57. Surprisingly, this is the largest gap that occurs between the conjectured minimal dimensional nonisometric volume-preserving action and the conjectured minimal dimensional nonisometric action.

Problem 2.11. Complete the proof of Zimmer's conjecture in general by overcoming the two problems just discussed.

If I had to guess, I would say that the second problem is considerably harder than the first to overcome. There is no clear robust dynamical behavior to exploit to resolve the problem.

3. TOTALLY GEODESIC MANIFOLDS AND RANK ONE SYMMETRIC SPACES

This section concerns recent results by Bader, the author, Miller, and Stover, motivated by questions of McMullen and Reid in the case of real hyperbolic manifolds. Throughout this section a geodesic submanifold will mean a closed immersed, totally geodesic submanifold. (In fact, all results can be stated also for orbifolds, but we ignore this technicality here.) A geodesic submanifold is *maximal* if it is not contained in a proper geodesic submanifold of smaller codimension.

For arithmetic manifolds, the presence of one maximal geodesic submanifold can be seen to imply the existence of infinitely many. The argument involves lifting the submanifold S to a finite cover $\tilde{M}$ where an element λ of the commensurator acts as an isometry. It is easy to check that for most choices of λ, the submanifold $\lambda(S)$ can be pushed back down to a geodesic submanifold of M that is distinct from S. This was perhaps first made precise in dimension 3 by Maclachlan–Reid and Reid [60,78], who also exhibited the first hyperbolic 3-manifolds with no totally geodesic surfaces.

In the real hyperbolic setting the main result from [6] is

Theorem 3.1 (Bader, Fisher, Miller, Stover). *Let Γ be a lattice in $\mathrm{SO}_0(n,1)$. If the associated locally symmetric space contains infinitely many maximal geodesic submanifolds of dimension at least 2, then Γ is arithmetic.*

Remark 3.2.

(1) The proof of this result involves proving a superrigidity theorem for *certain* representations of the lattice in $\mathrm{SO}_0(n,1)$. As the conditions required become a bit technical, we refer the interested reader to [6]. The superrigidity is proven using ideas and methods introduced in [7].

(2) At about the same time, Margulis and Mohammadi gave a different proof for the case $n = 3$ and Γ cocompact [64]. They also proved a superrigidity theorem, but both the statement and the proof are quite different than in [6].

(3) A special case of this result was obtained a year earlier by the author, Lafont, Miller, and Stover [41]. There we prove that a large class of nonarithmetic manifolds have only finitely many maximal totally geodesic submanifolds. This

includes all the manifolds constructed by Gromov and Piatetski-Shapiro but not the examples constructed by Agol and Belolipetsky-Thomson.

Theorem 3.1 has a reformulation entirely in terms of homogeneous dynamics, and homogenous dynamics play a key role in the proof. It is also interesting that a key role is also played by dynamics that are not quite homogeneous but that take place on a projective bundle over the homogeneous space G/Γ. In fact, the work can be used to give a classification of invariant measures for certain subgroups $W < \mathrm{SO}_0(n,1)$ on these projective bundles.

Even more recently the same authors have extended this result to cover the case of complex hyperbolic manifolds.

Theorem 3.3 (Bader, Fisher, Miller, Stover). *Let $n \geq 2$ and $\Gamma < \mathrm{SU}(n,1)$ be a lattice, and $M = \mathbb{C}\mathbb{H}^n/\Gamma$. Suppose that M contains infinitely many maximal totally geodesic submanifolds of dimension at least 2, then Γ is arithmetic.*

As before, this is proven using homogeneous dynamics, dynamics on a projective bundle over G/Γ, and a superrigidity theorem. Here the superrigidity theorem is even more complicated than before and depends also on results of Simpson and Pozzetti [74,81]. A very different proof for the case where the totally geodesic submanifolds are all assumed to be complex submanifolds was given very shortly after ours by Baldi and Ullmo [8]. There is almost no overlap of ideas between the two proofs, theirs characterizes the totally geodesic submanifolds in terms of special intersections and then studies them using Hodge theory and o-minimality.

The results in this section provide new evidence that totally geodesic manifolds play a very special role in nonarithmetic lattices and perhaps provide some evidence that the conventional wisdom on Questions 3.10 and 3.4 below should be reconsidered.

3.1. Other results and open questions

It is important to preface this section by saying that for all semisimple Lie groups G other than $\mathrm{SO}(n,1)$ for $n > 2$ and $\mathrm{SU}(n,1)$ for $n > 1$, we have an essentially complete classification of lattices in G. For $\mathrm{SO}(2,1) = \mathrm{SU}(1,1)$, the lattices are exactly the fundamental groups of hyperbolic surfaces of finite volume, which have been understood for quite some time. For all the remaining groups, all lattices are arithmetic. I will discuss the known construction of nonarithmetic lattices in $\mathrm{SO}(n,1)$ and $\mathrm{SU}(n,1)$. To begin slightly out of order, I emphasize one of the most important open problems in the area.

Question 3.4. For what values of n does there exist a nonarithmetic lattice in $\mathrm{SU}(n,1)$?

The answer is known to include 2 and 3. The first examples were constructed by Mostow in [69] using reflection group techniques. The list was slightly expanded by Mostow and Deligne using monodromy of hypergeometric functions [25,70]. The exact same list of examples was rediscovered/reinterpreted by Thurston in terms of conical flat structures on the 2-sphere [82], see also [79]. There is an additional approach via algebraic geometry suggested by Hirzebruch and developed by him in collaboration with Barthels and Höfer [9]. More examples have been discovered recently by Couwenberg, Heckman, and Looijenga using the Hirzebruch-style techniques and by Deraux, Parker, and Paupert using complex reflection group techniques [23,28–30]. But as of this writing there are only 22 commensurability classes of nonarithmetic lattices known in $\mathrm{SU}(2,1)$ and only 2 known in $\mathrm{SU}(3,1)$. An obvious refinement of Question 3.4 is the following:

Question 3.5. For what values of n do there exist infinitely many commensurability classes of a nonarithmetic lattice in $\mathrm{SU}(n,1)$?

We remark here that the approach via conical flat structures was extended by Veech and studied further by Ghazouani and Pirio [48,85]. Regrettably this approach does not yield more nonarithmetic examples. It seems that the reach of this approach might be extended to be roughly equivalent to the reach of the approach via monodromy of hypergeometric functions, see [47]. There appears to be some consensus among the experts that the answer to both Questions 3.4 and 3.5 should be "for all n", see, e.g., [55, CONJECTURE 10.8]. We point to a recent result of Esnault and Groechenig that indicates that complex hyperbolic lattices are in fact much more constrained than their real hyperbolic analogues [34].

Theorem 3.6. *Let Γ be a lattice in $G = \mathrm{SU}(n,1)$ for $n > 1$. Then Γ is integral, i.e., $\Gamma < G(k)$ for some number field k, and if v is any finite place of k then $\Gamma < G(k_v)$ is precompact.*

The earliest nonarithmetic lattices in $\mathrm{SO}(n,1)$ for $n > 2$ were constructed by Makarov and Vinberg by reflection group methods [61,86]. It is known by work of Vinberg that these methods will only produce nonarithmetic lattices in dimension less than 30 [87]. The largest known nonarithmetic lattice produced by these methods is in dimension 18 by Vinberg and the full limit of reflection group constructions is not well understood [88]. We refer the reader to [10] for a detailed survey. The following question seems natural:

Question 3.7. In what dimensions do there exist lattices in $\mathrm{SO}(n,1)$ or $\mathrm{SU}(n,1)$ that are commensurable to nonarithmetic reflection groups? In what dimensions do there exist lattices in $\mathrm{SO}(n,1)$ or $\mathrm{SU}(n,1)$ that are commensurable to arithmetic reflection groups?

For the real hyperbolic setting, there are known upper bounds of 30 for arithmetic lattices and 997 for any lattices. The upper bound of 30 also applies for nonarithmetic uniform hyperbolic lattices [10,87]. In the complex hyperbolic setting, there seem to be no known upper bounds, but a similar question recently appeared in, e.g., [55, QUESTION 10.10]. For a much more detailed survey of reflection groups in hyperbolic spaces, see [10].

A dramatic result of Gromov and Piatetski-Shapiro vastly increased our stock of nonarithmetic lattices in $\mathrm{SO}(n,1)$ by an entirely new technique:

Theorem 3.8 (Gromov and Piatetski-Shapiro). *For each n, there exist infinitely many commensurability classes of nonarithmetic uniform and nonuniform lattices in* $\mathrm{SO}(n,1)$.

The construction in [50] involves building hybrids of two arithmetic manifolds by cutting and pasting along totally geodesic codimension-one submanifolds. The key observation is that noncommensurable arithmetic manifolds can contain isometric totally geodesic codimension-one submanifolds. This method has been extended and explored by many authors for a variety of purposes, see, for example, [1, 2, 11, 46]. It has also been proposed that one might build nonarithmetic complex hyperbolic lattices using a variant of this method, though that proposal has largely been stymied by the lack of codimension-one totally geodesic codimension one submanifolds. The absence of codimension-1 submanifolds makes it difficult to show that attempted "hybrid" constructions yield discrete groups. For more information, see, e.g., [72, 73, 90] and [55, CONJECTURE 10.9]. We point out here that the results of Esnault and Groechenig discussed above implies that the "inbreeding" variant of Agol and Belolipetsky-Thomson [2, 11] cannot be adapted to produce nonarithmetic manifolds in the complex hyperbolic setting even if the original method of Gromov and Piatetski-Shapiro can be. To me personally, this seems a very strong negative indication on the possibility of such an adaptation. While the key observation of Agol was made long after the paper of Gromov and Piateksi-Shapiro, the constructions are very similar.

In [50], Gromov and Piatetski-Shapiro ask the following intriguing question:

Question 3.9. Is it true that, in high enough dimensions, all lattices in $\mathrm{SO}(n, 1)$ are built from subarithmetic pieces?

The question is somewhat vague and [50] also contains a group-theoretic variant that is easily seen to be false for the examples of Agol and Belolipetsky-Thomson, but a more precise starting point is:

Question 3.10. For $n > 3$, is it true that any nonarithmetic lattice in $\Gamma < \mathrm{SO}(n, 1)$ intersects some conjugate of $\mathrm{SO}(n-1, 1)$ in a lattice?

This is equivalent to asking if every finite-volume nonarithmetic hyperbolic manifold in dimension at least 4 contains a closed codimension-one totally geodesic submanifold. Both reflection group constructions and all known variants of hybrid constructions contain such submanifolds. The consensus in the field seems to be that the answer to this question should be no, but I know of no solid evidence for that belief. In particular, starting in dimension 4, Wang finiteness shows that the quantitative structure of hyperbolic manifolds is very different than in dimension 3 [89]. And recent work of Gelander–Levit makes it at least plausible that variants of the hybrid and inbreeding constructions construct "enough" hyperbolic manifolds to capture all examples [46]. This is very different than the situation in dimension 3 where hyperbolic Dehn surgery constructs many "more" hyperbolic manifolds and concretely exhibits the failure of Wang finiteness. One can think of Questions 3.9 and 3.10 as asking for particular qualitative reasons behind this difference in quantitative behavior.

In the next subsection, I will discuss some approaches to giving a positive answer to this question or perhaps criterion for a positive answer for some examples building on ideas in [6]. It is also not known to what extent the hybrid and reflection group constructions build distinct examples. Some first results, indicating that the classes are different, are contained in [41, THEOREM 1.7] and [65, THEOREM 1.5].

Given Theorem 3.1, it is reasonable to ask more detailed questions about the finite collection of totally geodesic submanifolds in a nonarithmetic hyperbolic manifold. A very reasonable question, on which first results have been obtained by Lindenstrauss and Mohammadi, is whether there is any bound on the finite number in terms of geometric invariants of the hyperbolic manifold. Their theorem, stated in Mohammadi's contribution to these proceedings as [66, THEOREM 7.4], gives a result in the class of constructions of the style of Gromov and Piatetski-Shapiro in the case of three-dimensional hyperbolic manifolds. A key ingredient is the Angle Rigidity Theorem found in a joint work of the author with Lafont, Miller, and Stover [41]. While the proof of finiteness in [41] is definitely superseded by that in [6], it is more useful for proving bounds because it gives an explicit open set of "impossible configurations" for the finite set of maximal totally geodesic submanifolds in many cut and paste constructions of hyperbolic manifolds. Roughly speaking, the key result in [41] says that, given a nonarithmetic manifold built in the manner of Gromov and Piatetski-Shapiro, any closed totally geodesic submanifold intersecting the "cut and paste" hypersurface must do so at a right angle. All other angles of intersection are forbidden. It seems highly unlikely that these are the only "impossible configurations."

Question 3.11. Find other restrictions on the possible configurations of totally geodesic submanifolds in a nonarithmetic hyperbolic manifold.

It is worth mentioning that our understanding of lattices in $\mathrm{SO}(2,1)$ and $\mathrm{SO}(3,1)$ is both more developed and very different. Lattices in $\mathrm{SO}(2,1)$ are completely classified, but there are many of them, with the typical isomorphism class of lattices having many nonconjugate realizations as lattices, parameterized by moduli space. In $\mathrm{SO}(3,1)$, Mostow rigidity means there are no moduli spaces. But Thurston–Jorgensen hyperbolic Dehn surgery still allows one to construct many "more" examples of lattices, including many that yield a negative answer to Question 3.10. There remains an interesting sense in which the answer to Question 3.9 could still be yes even for dimension 3.

Question 3.12. Can every finite-volume hyperbolic 3-manifold be obtained as Dehn surgery on an arithmetic manifold?

To clarify the question, it is known that every finite-volume hyperbolic 3-manifold is obtained as a topological manifold by Dehn surgery on some cover of the figure-8 knot complement, which is known to be the only arithmetic knot complement [51,77]. What is not known is whether one can obtain the geometric structure on the resulting three-manifold as geometric deformation of the complete geometric structure on the arithmetic manifold on which one performs Dehn surgery.

We end this section by discussing an additional question that illustrates the difference in our knowledge of real hyperbolic manifolds in dimension 3 and dimensions at least 4 and then discuss an intriguing variant in complex hyperbolic geometry. We call a group Γ *virtually large* if it has a finite index subgroup that surjects onto a nonabelian free group. The following theorem of Lubotzky shows a strong connection between totally geodesic submanifolds and this property [59].

Theorem 3.13. *Let M be a finite-volume hyperbolic manifold. Then if M admits a closed codimension-*1 *totally geodesic submanifold, the fundamental group of M is virtually large.*

It follows from work of Agol that in dimension 3 all finite-volume hyperbolic manifolds have a virtually large fundamental group [3]. But starting in dimension 5, there are explicit examples where we do not know if the fundamental group is virtually large. These are the so-called *second-type* arithmetic groups constructed using quaternion algebras. In fact, effectively the only way we know to show that a hyperbolic manifold of dimension 4 or higher has a virtually large fundamental group is to apply Lubotzky's theorem. We ask the following question in an intentionally provocative manner:

Question 3.14. If M is a finite-volume hyperbolic manifold of dimension at least 4, then is having a virtually large fundamental group equivalent to having a closed totally geodesic submanifold of codimension 1?

There seems to be something close to a consensus that the answer to this question should be "no" in general and all lattices in $\mathrm{SO}(n, 1)$ should be virtually large. However, I know of no strong evidence for this belief other than the results in dimension 3. There is even one potential strategy for proving a negative answer to Question 3.14, arising from Agol's work on the virtual Haken conjecture [3].

Question 3.15. If M is a finite-volume hyperbolic manifold, can we cubulate M?

The answer is yes in dimension 3 by the work of Kahn–Markovic [53]. Agol's work then implies that this cubulation is special, which is more than enough to prove largeness. The concrete challenge in this setting is to prove that *second-type* arithmetic hyperbolic manifolds can be cubulated. That the first-type can be is shown by Bergeron, Haglund, and Wise in [14]. It seems well known to experts that all inbreeding and hybrid examples can also be cubulated, but there does not appear to be an explicit statement in the literature. However, all of these approaches to cubulating hyperbolic manifolds of dimension at least 4 depend on the existence of totally geodesic submanifolds of codimension 1. To cubulate in cases where codimension-1 totally geodesic submanifolds do not exist, the work in dimension 3 suggests that one should find some other convex or quasiconvex submanifolds or subspaces of codimension 1. Due to density of the commensurator, in the arithmetic setting, it suffices to find one such manifold.

It is interesting in the context of this article to also summarize what is known about the largeness for lattices in $\mathrm{SU}(n, 1)$. In this case all known constructions do not produce largeness directly. Instead, one constructs a holomorphic retract onto a totally geodesic Riemann subsurface as a forgetful map in terms of the Deligne–Mostow hypergeometric monodromy construction [27]. It is possible that only these particular Deligne–Mostow constructions yield virtually large lattices in $\mathrm{SU}(n, 1)$, but here we only ask a weaker question.

Question 3.16. Let M be a finite-volume complex hyperbolic manifold. If the fundamental group of M is large, does M admit a holomorphic retract to a totally geodesic Riemann subsurface?

We remark here that by a result of Delzant and Py, complex hyperbolic manifolds of (complex) dimension at least 2 are not cubulated, so the approach to a negative answer to Question 3.14 is not available in this context [26].

3.2. Dynamics and (non)arithmeticity

The work in [5, 6] gives a broader set of tools for determining when a manifold is arithmetic. We discuss some further repercussions of these ideas here, though they have less decisive corollaries than the main results of those papers. We restrict here to the real hyperbolic setting for simplicity, so throughout this section $G = \mathrm{SO}(n,1)$ with $n > 2$. We will write $W = \mathrm{SO}(n-1,1)$, though for some technical results below, there are analogous statements when $W = \mathrm{SO}(k,1)$ for any $1 < k < n-1$.

It is known that, given a lattice $\Gamma < G$, there exists a number field ℓ and an ℓ structure on G such that $\Gamma < G(\ell)$. We note that there is a collection of valuations on ℓ, and we write ν_0 for the valuation for which $G(\ell_{\nu_0}) = \mathrm{SO}(n,1)$ in such a way that we get the given lattice embedding of Γ into $\mathrm{SO}(n,1)$.

Definition 3.17. Given a valuation ν on ℓ not equivalent to ν_0, the *arithmeticity obstruction* defined by ν is the embedding $\rho_\nu : \Gamma \to G(\ell_\nu) = H$. We say that the arithmeticity obstruction vanishes if Γ is precompact in H.

It was observed by Margulis that if all arithmeticity obstructions vanish, then Γ is arithmetic. If all nonarchimedean arithmeticity obstructions vanish, then Γ is called *integral*. Note that Γ is automatically Zariski dense in any $G(\ell_\nu)$. The standard and surprising technique for showing that arithmeticity obstructions vanish is to assume they do not and use a superrigidity theorem to see that this implies that ρ_ν extends to a continuous representation from G to H. This is easily seen to be impossible.

We now state the superrigidity theorem from [6] that is used there to show that arithmeticity obstructions vanish. We restrict attention to the case of $\mathrm{SO}(n,1)$ for simplicity. This theorem is stated in terms of the existence of a certain W-invariant measure on a certain flat bundle over G/Γ. The goal of this section is to explain that it is possible to use other dynamical results to weaken that hypothesis. Finding optimal hypothesis and more applications seems potentially important to understanding the questions discussed in the last subsection.

In this theorem we consider a local field k and a k-algebraic group $\mathbf{H}$ satisfying one additional condition. Let P be a minimal parabolic subgroup of G and U its unipotent radical. A pair consisting of a local field k and a k-algebraic group $\mathbf{H}$ is said to be *compatible* with G if for every nontrivial k-subgroup $\mathbf{J} < \mathbf{H}$ and any continuous homomorphism $\tau : P \to N_{\mathbf{H}}(\mathbf{J})/\mathbf{J}(k)$, where $N_{\mathbf{H}}(\mathbf{J})$ is the normalizer of $\mathbf{J}$ in $\mathbf{H}$, we have that the Zariski closure of $\tau(U')$ coincides with the Zariski closure of $\tau(U)$ for every nontrivial subgroup $U' < U$. That this condition is satisfied by any group arising in an arithmeticity obstruction for a lattice in $\mathrm{SO}(n,1)$ is a key point in [6]. The fact that this is already no longer true for $\mathrm{SU}(n,1)$ introduces new difficulties to overcome in [5] that we do not discuss here.

Theorem 3.18. *Let G be $\mathrm{SO}_0(n,1)$ for $n \geq 3$, $W < G$ be a noncompact simple subgroup, and $\Gamma < G$ be a lattice. Suppose that k is a local field and $\mathbf{H}$ is a connected k-algebraic group such that the pair consisting of k and $\mathbf{H}$ is compatible with G. Finally, let $\rho : \Gamma \to \mathbf{H}(k)$ be a homomorphism with unbounded, Zariski dense image. If there exist a k-rational faithful irreducible representation $\mathbf{H} \to \mathrm{SL}(V)$ on a k-vector space V and a W-invariant measure ν on $(G \times \mathbb{P}(V))/\Gamma$ that projects to Haar measure on G/Γ, then ρ extends to a continuous homomorphism from G to $\mathbf{H}(k)$.*

We retain the assumptions made on k and $\mathbf{H}$ and discuss how one can weaken the assumption in Theorem 3.18 using results of Eskin–Bonatti–Wilkinson. We first motivate the connection by stating a variant that follows easily.

Theorem 3.19 (BFMS reformulated). *Take the assumptions of Theorem 3.18 and replace the assumption of the existence of the invariant measure with the assumption that the W action on $(G \times V)/\Gamma$ is not irreducible. Then ρ extends to a continuous homomorphism from G to $\mathbf{H}(k)$.*

The two statements are equivalent by standard techniques. The point is that either the W-invariant measure or the W-invariant subspace are used to produce a Γ-equivariant measurable map $\phi : W\backslash G \to (\mathbf{H}/\mathbf{L})(k)$ for some proper algebraic subgroup $\mathbf{L} < \mathbf{H}$. The existence of this map is in fact equivalent to the existence of either such a subspace for some choice of V or such a measure for some (a priori different) choice of V. In this language, the first step of the work in [6] is using the existence of infinitely many maximal W orbit closures to produce an invariant measure for the W action on $(G \times \mathbb{P}(V))/\Gamma$ for some well chosen V. It is worth pointing out that there are other criteria for irreducibility in the literature, here we focus on one due to Bonatti, Eskin, and Wilkinson [15]. To obtain a simple statement, we let $A < P$ be the Cartan subgroup and choose the H representation on V such that the first Lyapunov exponent for A on V-bundle $(G \times V)/\Gamma$ is simple. It is easy to verify that one can choose such a V using tensor constructions. We let $P < G$ be the parabolic and $P_W = P \cap W$ the parabolic in W. Combining the main result of [15] with those of [6], we obtain the following:

Theorem 3.20. *With the common assumptions of the last two theorems, if there is more than one P_W-invariant measure on $(G \times \mathbb{P}(V))/\Gamma$ projecting to Haar measure on G/Γ, then ρ extends to a continuous homomorphism from G to $\mathbf{H}(k)$.*

This is a straightforward concatenation of [15, THEOREM 1.1] and Theorem 3.19. Observe first that the V-bundle $(G \times V)/\Gamma$ is G irreducible by hypothesis. The assumption of the second invariant measure in Theorem 3.20 and [15, THEOREM 1.1] imply that this bundle is not irreducible for W. We then apply Theorem 3.19.

Let $E_1(x)$ be the subspace of V corresponding to the first Lyapunov exponent for A and write $(G \times E_1)/\Gamma$ for the corresponding subbundle of $(G \times V)/\Gamma$. It is observed in [15] that $(G \times E_1)/\Gamma$ is P-invariant. Using that E_1 is one-dimensional, this yields a P-invariant section s of the bundle $(G \times \mathbb{P}(V))/\Gamma$, and we can push the Haar measure on

G/Γ forward along this section to build a P-invariant measure on $(G \times \mathbb{P}(V))/\Gamma$. We note that this measure cannot be W-invariant since P and W together generate G and it is easy to see that there is no G-invariant measure on $(G \times \mathbb{P}(V))/\Gamma$. Since $P_W < P$, this shows the existence of a W-invariant measure on $(G \times \mathbb{P}(V))/\Gamma$ projecting to Haar measure and implies the existence of at least two P_W-invariant measures on $(G \times \mathbb{P}(V))/\Gamma$ projecting to Haar measure. It seems a priori easier to produce P_W-invariant measures since P_W is amenable and therefore one can average. Producing a measure that is not also P-invariant becomes the challenge.

We also note here that this entire discussion on varying hypotheses applies in any case where G is a rank-one group and $W < G$ is a simple subgroup, we only require $G = \mathrm{SO}(n, 1)$ to use the superrigidity theorem from [6] to extend the representation. So the discussion adapts easily to the case of $\mathrm{SU}(n, 1)$ by the results of [5]. There is one additional fact that is special to $\mathrm{SO}(n, 1)$ and relates to Questions 3.10 and 3.14, namely if the hyperbolic manifold $K\backslash G/\Gamma$ has no totally geodesic manifolds of codimension-one and $W = \mathrm{SO}(n-1, 1)$ then the P_W action on G/Γ is uniquely ergodic. So in this setting we have the following.

Corollary 3.21. *Adding the assumption on totally geodesic submanifolds above to Theorem 3.20, if P_W is not uniquely ergodic on $(G \times \mathbb{P}(V))/\Gamma$ then ρ extends.*

It is somewhat surprising that arithmeticity is in this context associated with the existence of additional ergodic measures. Since Γ is finitely generated, it is easy to see that only finitely many valuations of ℓ are relevant to arithmeticity of Γ. So arithmeticity of Γ follows from the failure of unique ergodicity for P_W in finitely many dynamical systems for G. We note that the paper of Bonatti–Eskin–Wilkinson gives somewhat more explicit information than just about the number of measures. There is one "obvious" ergodic measure supported on the line in $P(V)$ corresponding to the first Lyapunov exponent. Forcing the vanishing of the arithmeticity obstruction simply requires finding any P_W-invariant measure where disintegration along fibers is not supported on that line.

In the context of Question 3.14, the possibilities for applying Corollary 3.21 are much broader. Given a lattice Γ that is large, one has a homomorphism $\Gamma \to F_2$ which one can compose with any (irreducible) representation of F_2 on a vector space V to obtain a G space $(G \times \mathbb{P}(V))/\Gamma$ to which one can attempt to apply Corollary 3.21. Since a representation factoring through F_2 cannot extend, the corollary actually asserts that largeness implies unique ergodicity of P_W in many dynamical systems. Of course, all lattices in $\mathrm{PSL}(2, \mathbb{C})$ are known to be virtually large and there are many where the locally symmetric space has no closed totally geodesic surfaces, so for that choice of G all of these constructions build many dynamical systems with a unique invariant measures for the corresponding P_W. In this case P_W is just the $ax + b$ group. It is not at this moment clear what structure one would like to exploit to show a difference between dimension 3 and higher dimensions.

One also might attempt to use Theorem 3.20 to either reprove the arithmeticity of lattices in $\mathrm{Sp}(n, 1)$ and F_4^{-20} by ergodic-theoretic methods or to reprove Theorem 3.6 by ergodic-theoretic methods. In addition, one could attempt a proof of Mostow rigidity using this circle of ideas. Certainly, the ideas are close to those that allowed Margulis to prove Mostow rigidity as a consequence of superrigidity in higher rank.

We note that while Margulis and Mohammadi suggested an alternate approach to the theorem in [6] and showed that their approach could succeed in the case where $G = \mathrm{SO}(3,1)$, their approach does not yield any of the observations in this subsection, but only produces results in the presence of infinitely many maximal closed totally geodesic submanifolds.

ACKNOWLEDGMENTS

I am happy to thank all of my collaborators, without whom the works discussed here would not exist. Particular thanks to Nick Miller for a detailed reading of an earlier draft and to Ian Agol and Matthew Stover for interesting conversations about largeness.

FUNDING

This work was partially supported by a Fellowship from the Simons Foundation, by a visiting professorship at the Miller Research Institute and by NSF DMS-1906107.

REFERENCES

[1] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, and I. Samet, On the growth of L^2-invariants of locally symmetric spaces, II: exotic invariant random subgroups in rank one. *Int. Math. Res. Not. IMRN* **9** (2020), 2588–2625.

[2] I. Agol, Systoles of hyperbolic 4-manifolds. 2007, arXiv:math/0612290.

[3] I. Agol, The virtual Haken conjecture. *Doc. Math.* **18** (2013), 1045–1087.

[4] J. An, A. Brown, and Z. Zhang, Zimmer's conjecture for non-split semisimple lie groups. In preparation.

[5] U. Bader, D. Fisher, N. Miller, and M. Stover, Arithmeticity, superrigidity and totally geodesic submanifolds of complex hyperbolic manifolds. 2020.

[6] U. Bader, D. Fisher, N. Miller, and M. Stover, Arithmeticity, superrigidity, and totally geodesic submanifolds. *Ann. of Math. (2)* **193** (2021), no. 3, 837–861.

[7] U. Bader and A. Furman, An extension of Margulis' Super-Rigidity Theorem. In *Dynamics, geometry and number theory: the impact of Margulis on modern mathematics*. University of Chicago Press, Chicago.

[8] G. Baldi and E. Ullmo, Special subvarieties of non-arithmetic ball quotients and Hodge Theory. 2021, arXiv:2005.03524.

[9] G. Barthel, F. Hirzebruch, and T. Höfer, *Geradenkonfigurationen und Algebraische Flächen*. Aspects Math. D4. Friedr. Vieweg & Sohn, Braunschweig, 1987.

[10] M. Belolipetsky, Arithmetic hyperbolic reflection groups. *Bull. Amer. Math. Soc. (N.S.)* **53** (2016), no. 3, 437–475.

[11] M. V. Belolipetsky and S. A. Thomson, Systoles of hyperbolic manifolds. *Algebr. Geom. Topol.* **11** (2011), no. 3, 1455–1469.

[12] Y. Benoist and J.-F. Quint, Stationary measures and invariant subsets of homogeneous spaces (III). *Ann. of Math. (2)* **178** (2013), no. 3, 1017–1059.

[13] E. J. Benveniste and D. Fisher, Nonexistence of invariant rigid structures and invariant almost rigid structures. *Comm. Anal. Geom.* **13** (2005), no. 1, 89–111.

[14] N. Bergeron, F. Haglund, and D. T. Wise, Hyperplane sections in arithmetic hyperbolic manifolds. *J. Lond. Math. Soc. (2)* **83** (2011), no. 2, 431–448.

[15] C. Bonatti, A. Eskin, and A. Wilkinson, Projective cocycles over $\mathrm{SL}(2, \mathbb{R})$ actions: measures invariant under the upper triangular group. In *Quelques aspects de la théorie des systèmes dynamiques: un hommage à Jean-Christophe Yoccoz. I*, pp. 157–180, Astérisque 415, SMF, 2020.

[16] A. Brown, Entropy, smooth ergodic theory and rigidity of group actions. Preprint, 2017.

[17] A. Brown, Lattice subgroups acting on manifolds. In *ICM 2022 Proceedings, Vol. 5*, pp. 3388–3411, EMS Press, 2022.

[18] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture: Subexponential growth, measure rigidity, and strong property (T). Preprint, 2016.

[19] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture for non-uniform lattices and escape of mass. Preprint, 2019.

[20] A. Brown, D. Fisher, and S. Hurtado, Zimmer's conjecture for actions of $\mathrm{SL}(m,\mathbb{Z})$. *Invent. Math.* **221** (2020), no. 3, 1001–1060.

[21] A. Brown and F. Rodriguez Hertz, Measure rigidity for random dynamics on surfaces and related skew products. *J. Amer. Math. Soc.* **30** (2017), no. 4, 1055–1132.

[22] A. Brown, F. Rodriguez Hertz, and Z. Wang, Invariant measures and measurable projective factors for actions of higher-rank lattices on manifolds. 2016, arXiv:1609.05565.

[23] W. Couwenberg, G. Heckman, and E. Looijenga, Geometric structures on the complement of a projective arrangement. *Publ. Math. Inst. Hautes Études Sci.* **101** (2005), 69–161.

[24] M. de la Salle, Analysis on simple Lie groups and lattices. In *ICM 2022 Proceedings, Vol. 4*, pp. 3166–3188, EMS Press, 2022.

[25] P. Deligne and G. D. Mostow, Monodromy of hypergeometric functions and non-lattice integral monodromy. *Publ. Math. Inst. Hautes Études Sci.* **63** (1986), 5–89.

[26] T. Delzant and P. Py, Cubulable Kähler groups. *Geom. Topol.* **23** (2019), no. 4, 2125–2164.

[27] M. Deraux, Forgetful maps between Deligne–Mostow ball quotients. *Geom. Dedicata* **150** (2011), 377–389.

[28] M. Deraux, A new non-arithmetic lattice in $\mathrm{PU}(3, 1)$. 2019, arXiv:1710.04463.

[29] M. Deraux, J. R. Parker, and J. Paupert, New non-arithmetic complex hyperbolic lattices. *Invent. Math.* **203** (2016), no. 3, 681–771.

[30] M. Deraux, J. R. Parker, and J. Paupert, New non-arithmetic complex hyperbolic lattices II. 2020, arXiv:1611.00330.

[31] B. Deroin and S. Hurtado, Non left-orderability of lattices in higher rank semisimple lie groups. Preprint, 2020.

[32] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the SL(2, $\mathbb{R}$) action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127** (2018), 95–324.

[33] A. Eskin, M. Mirzakhani, and A. Mohammadi, Isolation, equidistribution, and orbit closures for the SL(2, $\mathbb{R}$) action on moduli space. *Ann. of Math. (2)* **182** (2015), no. 2, 673–721.

[34] H. Esnault and M. Groechenig, Cohomologically rigid local systems and integrality. *Selecta Math. (N.S.)* **24** (2018), no. 5, 4279–4292.

[35] B. Farb and P. Shalen, Lattice actions, 3-manifolds and homology. *Topology* **39** (2000), no. 3, 573–587.

[36] D. Fisher, Deformations of group actions. *Trans. Amer. Math. Soc.* **360** (2008), no. 1, 491–505.

[37] D. Fisher, Groups acting on manifolds: around the Zimmer program. In *Geometry, rigidity, and group actions*, pp. 72–157, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 2011.

[38] D. Fisher, Superrigidity, arithmeticity, normal subgroups: results, ramifications and directions. Preprint, 2016.

[39] D. Fisher, Recent progress in the Zimmer program. In *Group actions in ergodic theory, geometry, and topology*, University of Chicago Press, Chicago, 2019.

[40] D. Fisher, Recent developments in the Zimmer program. *Notices Amer. Math. Soc.* **67** (2020), no. 4, 492–499.

[41] D. Fisher, J.-F. Lafont, N. Miller, and M. Stover, Finiteness of maximal geodesic submanifolds in hyperbolic hybrids. *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 11, 3591–3623.

[42] D. Fisher and G. A. Margulis, Local rigidity for cocycles. In *Surveys in differential geometry, Vol. VIII (Boston, MA, 2002)*, pp. 191–234, Surv. Differ. Geom. 8, Int. Press, Somerville, MA, 2003.

[43] D. Fisher and K. Whyte, Continuous quotients for lattice actions on compact spaces. *Geom. Dedicata* **87** (2001), no. 1–3, 181–189.

[44] H. Furstenberg, Random walks and discrete subgroups of Lie groups. In *Advances in probability and related topics, vol. 1*, pp. 1–63, Dekker, New York, 1971.

[45] H. Furstenberg, Stiffness of group actions. In *Lie groups and ergodic theory (Mumbai, 1996)*, pp. 105–117 Tata Inst. Fund. Res. Stud. Math. 14, Tata Inst. Fund. Res, Bombay, 1998.

[46] T. Gelander and A. Levit, Counting commensurability classes of hyperbolic manifolds. *Geom. Funct. Anal.* **24** (2014), no. 5, 1431–1447.

[47] S. Ghazouani and L. Pirio, Moduli spaces of flat tori and elliptic hypergeometric functions. 2016.

[48] S. Ghazouani and L. Pirio, Moduli spaces of flat tori with prescribed holonomy. *Geom. Funct. Anal.* **27** (2017), no. 6, 1289–1366.

[49] E. Ghys, Groups acting on the circle. *Enseign. Math. (2)* **47** (2001), no. 3–4, 329–407.

[50] M. Gromov and I. Piatetski-Shapiro, Nonarithmetic groups in Lobachevsky spaces. *Publ. Math. Inst. Hautes Études Sci.* **66** (1988), 93–103.

[51] H. M. Hilden, M. T. Lozano, and J. M. Montesinos, On knots that are universal. *Topology* **24** (1985), no. 4, 499–504.

[52] S. Hurtado, A. Kocsard, and F. Rodríguez-Hertz, The Burnside problem for $\mathrm{Diff}_\omega(\mathbb{S}^2)$. *Duke Math. J.* **169** (2020), no. 17, 3261–3290.

[53] J. Kahn and V. Markovic, The surface subgroup and the Ehrenepreis conjectures. In *Proceedings of the international congress of mathematicians—Seoul 2014. Vol. II*, pp. 897–909, Kyung Moon Sa, Seoul, 2014.

[54] B. Kalinin, A. Katok, and F. Rodriguez Hertz, Nonuniform measure rigidity. *Ann. of Math. (2)* **174** (2011), no. 1, 361–400.

[55] M. Kapovich, Lectures on complex hyperbolic Kleinian groups. 2019, arXiv:1911.12806.

[56] A. Katok and J. Lewis, Global rigidity results for lattice actions on tori and new examples of volume-preserving actions. *Israel J. Math.* **93** (1996), 253–280.

[57] F. Ledrappier and L.-S. Young, The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin's entropy formula. *Ann. of Math. (2)* **122** (1985), no. 3, 509–539.

[58] H. Lee, Remarks on global rigidity of higher rank lattice actions. 2020, arXiv:2010.11874.

[59] A. Lubotzky, Free quotients and the first Betti number of some hyperbolic manifolds. *Transform. Groups* **1** (1996), no. 1–2, 71–82.

[60] C. Maclachlan and A. W. Reid, Commensurability classes of arithmetic Kleinian groups and their Fuchsian subgroups. *Math. Proc. Cambridge Philos. Soc.* **102** (1987), no. 2, 251–257.

[61] V. S. Makarov, On a certain class of discrete groups of Lobačevskiĭ space having an infinite fundamental region of finite measure. *Dokl. Akad. Nauk SSSR* **167** (1966), 30–33.

[62] G. A. Margulis, Non-uniform lattices in semisimple algebraic groups. In *Lie groups and their representations*, pp. 371–553, Halted, New York, 1975.

[63] G. A. Margulis, Discrete groups of motions of manifolds of nonpositive curvature. In *Proceedings of the international congress of mathematicians (Vancouver, BC, 1974), Vol. 2*, pp. 21–34, 1975.

[64] G. Margulis and A. Mohammadi, Arithmeticity of hyperbolic 3-manifolds containing infinitely many totally geodesic surfaces. 2019, arXiv:1902.07267.

[65] O. Mila, The trace field of hyperbolic gluings. *Int. Math. Res. Not. IMRN* **6** (2021), 4392–4412.

[66] A. Mohammadi, Finitary analysis in homogeneous spaces. In *ICM 2022 Proceedings, Vol. 5*, pp. 3530–3551, EMS Press, 2022.

[67] D. W. Morris and R. J. Zimmer, Ergodic actions of semisimple Lie groups on compact principal bundles. *Geom. Dedicata* **106** (2004), 11–27.

[68] G. D. Mostow, *Strong rigidity of locally symmetric spaces*. Princeton University Press, Princeton, NJ, 1973.

[69] G. D. Mostow, On a remarkable class of polyhedra in complex hyperbolic space. *Pacific J. Math.* **86** (1980), no. 1, 171–276.

[70] G. D. Mostow, Generalized Picard lattices arising from half-integral conditions. *Publ. Math. Inst. Hautes Études Sci.* **63** (1986), 91–106.

[71] A. Nevo and R. J. Zimmer, A structure theorem for actions of semisimple Lie groups. *Ann. of Math. (2)* **156** (2002), no. 2, 565–594.

[72] J. Paupert, Non-discrete hybrids in $\mathrm{SU}(2,1)$. *Geom. Dedicata* **157** (2012), 259–268.

[73] J. Paupert and J. Wells, Hybrid lattices and thin subgroups of Picard modular groups. *Topology Appl.* **269** (2020), 106918.

[74] M. B. Pozzetti, Maximal representations of complex hyperbolic lattices into $\mathrm{SU}(M,N)$. *Geom. Funct. Anal.* **25** (2015), no. 4, 1290–1332.

[75] M. Ratner, On Raghunathan's measure conjecture. *Ann. of Math. (2)* **134** (1991), no. 3, 545–607.

[76] M. Ratner, Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.* **63** (1991), no. 1, 235–280.

[77] A. W. Reid, Arithmeticity of knot complements. *J. Lond. Math. Soc. (2)* **43** (1991), no. 1, 171–184.

[78] A. W. Reid, Totally geodesic surfaces in hyperbolic 3-manifolds. *Proc. Edinb. Math. Soc. (2)* **34** (1991), no. 1, 77–88.

[79] R. Schwartz, Notes on shapes of polyhedra. 2015.

[80] N. A. Shah, Limit distributions of polynomial trajectories on homogeneous spaces. *Duke Math. J.* **75** (1994), no. 3, 711–732.

[81] C. T. Simpson, Higgs bundles and local systems. *Publ. Math. Inst. Hautes Études Sci.* **75** (1992), 5–95.

[82] W. P. Thurston, Shapes of polyhedra and triangulations of the sphere. In *The Epstein birthday schrift*, pp. 511–549, Geom. Topol. Monogr. 1, Geom. Topol. Publ., Coventry, 1998.

[83] F. Uchida, Classification of real analytic $\mathrm{SL}(n,\mathbb{R})$ actions on n-sphere. *Osaka J. Math.* **16** (1979), no. 3, 561–579.

[84] W. A. Veech, The Teichmüller geodesic flow. *Ann. of Math. (2)* **124** (1986), no. 3, 441–530.

[85] W. A. Veech, Flat surfaces. *Amer. J. Math.* **115** (1993), no. 3, 589–689.

[86] E. B. Vinberg, Discrete groups generated by reflections in Lobačevskiĭ spaces. *Mat. Sb.* **72** (1967), no. 114, 471–488. Correction, ibid. **73** (1967), no. 115, 303.

[87] E. B. Vinberg, The nonexistence of crystallographic reflection groups in Lobachevskiĭ spaces of large dimension. *Funktsional. Anal. i Prilozhen.* **15** (1981), no. 2, 67–68.

[88] E. B. Vinberg, Non-arithmetic hyperbolic reflection groups in higher dimensions. *Mosc. Math. J.* **15** (2015), no. 3, 593–602, 606.

[89] H. C. Wang, Topics on totally discontinuous groups. In *Symmetric spaces (Short Courses, Washington Univ., St. Louis, MO, 1969–1970)*, pp. 459–487, Pure Appl. Math. 8, 1972.

[90] J. Wells, Non-arithmetic hybrid lattices in $\mathrm{PU}(2,1)$. 2019.

[91] R. J. Zimmer, Strong rigidity for ergodic actions of semisimple Lie groups. *Ann. of Math. (2)* **112** (1980), no. 3, 511–529.

[92] R. J. Zimmer, Arithmetic groups acting on compact manifolds. *Bull. Amer. Math. Soc. (N.S.)* **8** (1983), no. 1, 90–92.

[93] R. J. Zimmer, Actions of semisimple groups and discrete subgroups. In *Proceedings of the international congress of mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)*, pp. 1247–1258, Amer. Math. Soc., Providence, RI, 1987.

DAVID FISHER

Math Department – MS 136, Rice University, P.O. Box 1892, Houston, TX 77005-1892, USA, davidfisher@rice.edu

FURSTENBERG DISJOINTNESS, RATNER PROPERTIES, AND SARNAK'S CONJECTURE

MARIUSZ LEMAŃCZYK

ABSTRACT

A recent progress on Sarnak's conjecture on Möbius orthogonality is discussed with the main focus on the proof of Veech's conjecture and its consequences.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37A44; Secondary 11N37, 37A05, 37A17, 37A30, 37B10, 37B40

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3508–3528
DOI 10.4171/ICM2022/51

1. INTRODUCTION. STATE-OF-THE-ART

This article deals with Sarnak's conjecture [34] from 2010, also called the Möbius Orthogonality Conjecture, MOC for short, see (1.2) below. More precisely, the aim of this article is to give an account on the progress concerning MOC caused by the proof of Veech's conjecture [38] from 2016 in the recent article [24]. In order to do it, we will present a panorama of earlier concepts and results concerning the new interactions between ergodic theory and analytic number theory caused by MOC, especially relating MOC to the celebrated Chowla conjecture [5] from 1965. We will be concentrated on the directions of research on the ergodic theory side, with a special focus on projects in which the author of the article took part. In this extended introduction, we present the state-of-the-art of the subject. To keep this presentation reasonably short, some elementary definitions and facts from dynamics are postponed to Sections 2 and 3, while some facts (especially around entropy) are treated as "commonly" known and can be found in the ergodic theory literature, see, e.g., [9,16,39]. On the analytic number theory side, basic facts can be found, e.g., in [19,23].

Möbius function. Each natural number $n \in \mathbb{N} := \{1, 2, \ldots\}$ has its (unique) decomposition into a product of primes which is the basic fact about the not finitely generated *multiplicative structure* of $\mathbb{N}$. The set $\mathbb{P}$ of primes is believed to behave like a "random" subset of natural numbers. This randomness should be reflected in the properties of arithmetic functions $\boldsymbol{u} : \mathbb{N} \to \mathbb{C}$ that preserve the multiplicative structure of $\mathbb{N}$, that is, they are themselves *multiplicative*, $\boldsymbol{u}(mn) = \boldsymbol{u}(m)\boldsymbol{u}(n)$ whenever $(m, n) = 1$ (in other words, they are determined by their values on the powers of the primes). One of most prominent multiplicative functions is the *Möbius function* $\boldsymbol{\mu} : \mathbb{N} \to \{-1, 0, 1\}$ for which $\boldsymbol{\mu}(1) = 1$, $\boldsymbol{\mu}(n) = (-1)^k$, with n being the product of k distinct primes, and $\boldsymbol{\mu}(n) = 0$ for the remaining $n \in \mathbb{N}$. Is this function "random"? If so, this should be reflected in the phenomenon of cancelations of ±1s like for a sample from an independent process. It is then classical that the Prime Number Theorem (PNT), i.e., $|\{p \le N : p \in \mathbb{P}\}| \sim \frac{N}{\log N}$, is equivalent to $\lim_{N\to\infty} \frac{1}{N} \sum_{n \le N} \boldsymbol{\mu}(n) = 0$, and the Riemann Hypothesis is equivalent to a quantitative version of the above, namely $\sum_{n \le N} \boldsymbol{\mu}(n) = O(N^{\frac{1}{2}+\varepsilon})$ for each $\varepsilon > 0$.

The Chowla conjecture. Another way to express the randomness of $\boldsymbol{\mu}$ is the Chowla conjecture [5] from 1965 claiming that the autocorrelations of the Möbius function, unless they are correlations of $\boldsymbol{\mu}^2$, vanish:[1]

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n \le N} \boldsymbol{\mu}^{s_1}(n + a_1) \cdots \boldsymbol{\mu}^{s_k}(n + a_k) = 0 \tag{1.1}$$

[1] Originally, the Chowla conjecture was formulated for the Liouville function $\boldsymbol{\lambda} : \mathbb{N} \to \{-1, 1\}$ and concerned *all* correlations. Liouville function depends on the parity of all prime factors (counted with multiplicities) and clearly satisfies $\boldsymbol{\mu} = \boldsymbol{\lambda} \cdot \boldsymbol{\mu}^2$.

for each $k \geq 1$, $0 \leq a_1 < \cdots < a_k$, and $s_j \in \{1,2\}$, not all s_j being 2.[2] We will see later (see Section 3) that the Chowla conjecture is precisely the fact that $\boldsymbol{\mu}$ is a generic point for a kind of Bernoulli measure over the so-called Mirsky measure, for which $\boldsymbol{\mu}^2$ is generic, i.e., intuitively, relative to the positions of zeros, in $\boldsymbol{\mu}$ we observe a replacement of 1s in $\boldsymbol{\mu}^2$ by ± 1s with equal probability.

Sarnak's conjecture (MOC). Another way of talking about the randomness of $\boldsymbol{\mu}$ could be in terms of correlations with other sequences. In [23], this is expressed by the Möbius Randomness Law: $\boldsymbol{\mu}$ is so random that it does not correlate with any "reasonable" bounded sequence. In 2010, P. Sarnak [34] formulated a much more precise form of this vague randomness principle, namely

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n \leq N} f(T^n x)\boldsymbol{\mu}(n) = 0, \tag{1.2}$$

for each *zero (topological) entropy* homeomorphism T of a compact metric space X, all $f \in C(X)$, and all $x \in X$.[3] In other words, $\boldsymbol{\mu}$ is random because $\boldsymbol{\mu}$ does not correlate with *any* deterministic sequence,[4] we also say that $\boldsymbol{\mu}$ *is orthogonal* to all (bounded) deterministic sequences or to all deterministic systems. Of course, in (1.2), we can consider other (always bounded though) arithmetic functions $\boldsymbol{u} : \mathbb{N} \to \mathbb{C}$ and consider an analogous problem of orthogonality to a selected class of topological systems (for example, it is well known that if we replace the Möbius function $\boldsymbol{\mu}$ with the Liouville function $\boldsymbol{\lambda}$, then the corresponding Sarnak's conjectures are equivalent; see, e.g., [10]).[5]

Proposition 1.1 (Sarnak [34], for proofs see [1, 35]). *Chowla conjecture implies Sarnak's conjecture.*

Whether the converse is true remains open, but we have the following:

Proposition 1.2 ([17]). *Sarnak's conjecture implies the Chowla conjecture along a subsequence (i.e., in (1.1) we need to consider (N_s) instead of N[6]).*

[2] If all $s_j = 2$ then the limit exists but need not be zero: $\boldsymbol{\mu}^2$ is the characteristic function of the set of square-free numbers whose natural density is $6/\pi^2$. In fact, the frequencies of all blocks of 0, 1s on $\boldsymbol{\mu}^2$ exist – $\boldsymbol{\mu}^2$ is a generic point for a shift-invariant measure $\nu_{\boldsymbol{\mu}^2}$ called the *Mirsky measure* – see Section 3 for details.

[3] It is "for all $x \in X$" which is the core of Sarnak's conjecture: the "x-almost every" version of (1.2) holds for *every* dynamical system (in particular, regardless of the entropy [34], see also [1]).

[4] In the theory of dynamical systems, zero-entropy systems are also called *deterministic* and the corresponding continuous observables $(f(T^n x))_{n\in\mathbb{Z}}$ are precisely deterministic sequences.

[5] The reader can notice that if $\boldsymbol{u}$ is orthogonal to all deterministic sequences, then for *any* bounded deterministic sequence $(a(n))$, the arithmetic function $\boldsymbol{v}(n) := \boldsymbol{u}(n)a(n)$ is also orthogonal to all deterministic sequences. Of course, such an operation in general "kills" the multiplicativity of $\boldsymbol{u}$.

[6] Tao in [37] strengthened this result by showing that (N_s) can be selected to have full logarithmic density.

We will detail more on that at the end of the introduction when we consider the logarithmic versions of the Chowla and Sarnak's conjectures.

Two strategies to "attack" Sarnak's conjecture. Returning to the original Sarnak's conjecture, we can view (1.2) as a classical Cesàro (ergodic) sum with **m**ultiplicative **w**eights (this point of view leads to the MW-strategy) or we can reverse the roles and consider Cesàro sums of $\boldsymbol{\mu}$[7] with **e**rgodic **w**eights (this point of view leads to the EW-strategy). Both strategies lead sooner or later to an interplay between analytic number theory and the theory of joinings in dynamics, in particular, the disjointness theory of Furstenberg in ergodic theory. Let us now say a few words on these strategies; more details, especially on the EW-strategy, will be provided later.

MW-strategy. DDKBSZ criterion. The core of the MW-strategy (in which we only use the fact that $\boldsymbol{\mu}$ is multiplicative) is the following numerical DDKBSZ criterion:[8]

Theorem 1.3 ([4, 27]). *Assume that $(f_n)_{n\in\mathbb{N}} \subset \mathbb{C}$ is bounded. If*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f_{pn}\overline{f}_{qn} = 0 \tag{1.3}$$

for all distinct, sufficiently large primes $p, q \in \mathbb{P}$, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f_n \boldsymbol{u}(n) = 0$$

for each bounded multiplicative *function $\boldsymbol{u} : \mathbb{N} \to \mathbb{C}$.*

Then, the DDKBSZ criterion is used in the following manner. Take *any* dynamical system (X, T) with a unique invariant measure μ (which must also be unique for all nonzero powers of T) and let $x \in X$. In the space $M(X \times X)$ of probability measures on $X \times X$ consider the sequence of *empiric measures* $(\frac{1}{N}\sum_{n\leq N} \delta_{(T^{pn}x, T^{qn}x)})$ (see Section 3 for more details). By the compactness of the weak-$*$-topology, there exists a subsequence (N_k) such that

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n\leq N_k} \delta_{(T^{pn}x, T^{qn}x)} = \rho,$$

where necessarily the measure ρ is $T^p \times T^q$-invariant and if $f \in C(X)$ then

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n\leq N_k} f(T^{pn}x)\overline{f(T^{qn}x)} = \int_{X\times X} f \otimes \overline{f}\, d\rho.$$

By our assumption, the two projections of ρ on X are μ. *If* we are able to show that ρ is the product measure $\mu \otimes \mu$, then we can easily apply the DDKBSZ criterion for the continuous zero-mean (for the measure μ) functions f and the sequences $(f(T^n x))$ (i.e., $f_n = f(T^n x)$ in Theorem 1.3). A spectacular case when this approach works was first demonstrated in the

[7] In analytic number theory, the problem of studying means of multiplicative functions is classical, cf. Halász theorem.

[8] DDKBSZ stands for Daboussi, Delange, Kátai, Bourgain, Sarnak, and Ziegler.

case of horocycle flows in [4] using Ratner's theory. However, a "typical" playground for the MW-strategy is when the automorphisms T^p and T^q considered with the same T-invariant measure μ are disjoint (see Section 2 for definition) in the Furstenberg sense, as in this case $\mu \otimes \mu$ is simply the *only* $T^p \times T^q$-invariant measure. (Note that this is not applicable to the horocycle flows themselves as all positive time automorphisms are isomorphic, so there are many $T^p \times T^q$-invariant measures.) As we can see, the MW-strategy leads to pure ergodic theory problems. While a list of papers in which the disjointness of powers has been proved can be found in the surveys [10,28], in Section 4, we will detail more on the answer to Ratner's question about the validity of MOC for smooth time changes of horocyclic flows. Namely, while for the algebraic actions the underlying configuration space is homogeneous, once the time is changed (in a nontrivial way), the configuration space becomes nonhomogeneous for the action of the time-changed flow which surprisingly leads to another extreme: the new flows enjoy formidable internal disjointness properties which allows us to use Theorem 1.3 to answer positively Ratner's question and go beyond.

MW-Strategy. The AOP property. There is a pure ergodic theory counterpart of DDKBSZ criterion, namely the notion of AOP (Asymptotic Orthogonality of Powers) introduced in [3]: a measure-theoretic (ergodic) dynamical system $(X, \mathcal{B}, \mu, T)$ has the AOP property ($J^e(T^p, T^q)$ below stands for the set of ergodic joinings between T^p and T^q) if

$$\limsup_{p\neq q, \mathbb{P}\ni p,q\to\infty} \sup_{\rho\in J^e(T^p,T^q)} \left| \int_{X\times X} f \otimes g \, d\rho \right| = 0$$

for each $f, g \in L^2_0(X, \mu)$. Obviously, AOP takes place if the prime powers of an automorphisms are disjoint but an AOP automorphism can have all nonzero powers isomorphic. AOP implies zero entropy and total ergodicity, i.e., all nonzero powers are ergodic. All ergodic quasidiscrete spectrum automorphisms [3], and also ergodic nil-automorphisms enjoy this property [12]. Let us see why AOP is useful when proving Möbius orthogonality. Suppose that we want to prove Möbius orthogonality for *all* (uniquely ergodic) models of totally ergodic rotations. Well, the Möbius orthogonality can first be easily established in *some* models of such rotations.

Namely, let X stand for any compact Abelian monothetic group with Haar measure λ_X. Then λ_X is the only (ergodic) invariant measure for any rotation $Tx = x + x_0$, with $\{nx_0 : n \in \mathbb{Z}\}$ being dense in X. If $\chi : X \to \mathbb{S}^1$ is any nontrivial character of X then

$$\frac{1}{N}\sum_{n\le N} \chi(T^{pn}x)\overline{\chi(T^{qn}x)} = \frac{1}{N}\sum_{n\le N} \left(\chi(x_0)^{p-q}\right)^n \to 0$$

whenever $p \neq q$ (remembering that, by total ergodicity, $\chi(x_0)$ is not a root of unity). Since the dual group $\hat{X}$ is linearly dense in $C(X)$, all totally ergodic rotations are Möbius orthogonal by virtue of Theorem 1.3. But now take *any* topological system (Z, R) which is uniquely ergodic (with a unique invariant measure κ) and suppose that (Z, κ, R) and (X, λ_X, T) are measure-theoretically isomorphic. Since the eigenfunctions in $L^2(Z, \kappa)$ need not be continuous,[9] using DDKBSZ criterion does not seem to be possible. We can prove, however, (see [3]) that the AOP property holds. Moreover, once a uniquely ergodic system satisfies AOP, it must be orthogonal to any (bounded) multiplicative function [3]. Now, AOP being a measure-theoretic invariant must be satisfied in all uniquely ergodic models of totally ergodic rotations. In fact, AOP implies something which looks much stronger.

MW-strategy. The strong MOMO property. A dynamical system (X, T) is said to satisfy the strong MOMO[10] property (see [2]) if for each increasing sequence (b_k) of natural numbers, $b_{k+1} - b_k \to \infty$, and each $f \in C(X)$, we have

$$\lim_{K\to\infty} \frac{1}{b_K} \sum_{k<K} \Big\| \sum_{b_k \le n < b_{k+1}} \mu(n) f \circ T^n \Big\|_{C(X)} = 0.$$

We have the following:

Theorem 1.4 ([2]). *The following holds:*

(i) *The strong MOMO property of a topological systems (X, T) implies its Möbius orthogonality.*

(ii) *The strong MOMO property of a topological system (X, T) implies uniformity (in $x \in X$) in the definition of Möbius orthogonality property.*

(iii) *Sarnak's conjecture is equivalent to the fact that all zero-entropy systems satisfy the strong MOMO property.*

(iv) *No system with positive entropy satisfies the strong MOMO property.*[11]

Moreover, we have the following:

Proposition 1.5 ([2]). *Let $(Z, \mathcal{D}, \kappa, R)$ be any totally ergodic measure-theoretic system. If it satisfies the AOP property then each of its uniquely ergodic models satisfies the strong MOMO property. In particular, all such uniquely ergodic models are Möbius orthogonal.*

Short interval behavior. The reader certainly noticed that by putting $f = 1$ in the definition of the strong MOMO property, we obtain that, whenever $b_{k+1} - b_k \to \infty$,

$$\lim_{K\to\infty} \frac{1}{b_K} \sum_{k<K} \Big| \sum_{b_k \le n < b_{k+1}} \mu(n) \Big| = 0.$$

It is not hard to see that this property is equivalent to the following:

$$\frac{1}{M} \sum_{M \le m < 2M} \frac{1}{H} \Big| \sum_{m \le n < m+H} \mu(n) \Big| \to 0,$$

[9] Topological system (Z, R) can be even topologically mixing, which excludes the possibility of continuous eigenfunctions.

[10] The acronym comes from *Möbius Orthogonality of Moving Orbits*.

[11] In [8] there are examples of positive entropy systems which are Möbius orthogonal. As Theorem 1.4 (iv) shows, this cannot happen for the strong MOMO property (assuming the Chowla conjecture).

whenever $H \to \infty$ and $H = o(M)$. This tells us that on a "typical" short interval (i.e., of length H) we have cancelations of 1s and -1s. This is a special property of the Möbius function proved in the breakthrough paper [30] by Matomäki and Radziwiłł in 2015. Together with the subsequent paper [31], it allowed in [3] to prove Sarnak's conjecture for all uniquely ergodic models of *finite* rotations[12] and all totally ergodic rotations. The fact, that all dynamical systems whose all invariant measures yield automorphisms with discrete spectrum satisfy Sarnak's conjecture was first proved in [20, 21].

EW-strategy. Let us now pass to the second strategy which consists in the following. Using combinatorial properties of $\boldsymbol{\mu}$, we count on deriving special ergodic properties of the Furstenberg systems (see Section 3.1 for this crucial definition) of $\boldsymbol{\mu}$ (we recall that the Chowla conjecture predicts that there is only one Furstenberg system of $\boldsymbol{\mu}$ given by $\hat{\nu}_{\boldsymbol{\mu}^2}$, i.e., by the relatively independent extension of the Mirsky measure of $\boldsymbol{\mu}^2$). We then expect that Furstenberg systems will display "enough" of disjointness with at least a subclass of zero entropy systems to advance on MOC or else, expressing MOC as some ergodic property of them, when translated it back to $\boldsymbol{\mu}$, will tell us which new combinatorial properties of $\boldsymbol{\mu}$ are needed to prove Sarnak's conjecture. So, of course, the crucial question is whether Sarnak's conjecture can be expressed in the language of Furstenberg systems of $\boldsymbol{\mu}$. This was conjectured by Veech [38] and finally proved in [24] ($\Pi(\kappa)$ below stands for the Pinsker σ-algebra of $(X_{\boldsymbol{\mu}}, \mathcal{B}(X_{\boldsymbol{\mu}}), \kappa, S)$, i.e., the largest zero-entropy factor of the system, while $\pi_0 : \{-1,0,1\}^{\mathbb{Z}} \to \mathbb{R}$, $\pi_0(y) = y_0$).

Theorem 1.6 ([24]). *Sarnak's conjecture holds if and only if, for all Furstenberg systems κ of $\boldsymbol{\mu}$, we have $\pi_0 \perp L^2(\Pi(\kappa))$.*

(This theorem also holds in the logarithmic case.) The above put across the intuition that the Chowla conjecture in ergodic theory corresponds to the Bernoulli property (maximal chaos), while Sarnak's conjecture is rather related to the weaker property, namely the Kolmogorov property (K-property) of a measure-preserving system, meant "locally", i.e., for the single function π_0. It is classical in ergodic theory that the K-property is equivalent to K-mixing (called also uniform mixing). We will see in Section 5.4 that K-mixing property applied "locally" to π_0 yields a combinatorial condition on $\boldsymbol{\mu}$ equivalent to MOC. Roughly, this condition is about cancelations of $+1$s and -1s along larger and larger shifts of the sets of return times of blocks which can also be interpreted as the intuition that the multiplicative and additive structures of $\mathbb{N}$ are independent. While due to the MW-strategy many examples of classes of zero-entropy systems for which the MOC holds were given, to apply the DDKBSZ criterion, the arguments were provided ad hoc, depending on the class under consideration which shows its certain weakness.[13] In contrast, the EW-strategy aims at general results and some spectacular successes were obtained for the logarithmic versions of the Chowla and Sarnak's conjectures which we now present.

[12] In other words, before 2015, we had no chances to prove Sarnak's conjecture, as we were already stuck in a relatively simple class of dynamical systems with zero entropy.

[13] On the other hand, this strategy leads to the study of internal disjointness properties of measure-preserving systems which is of independent interest in ergodic theory.

The logarithmic versions of the Chowla and Sarnak's conjectures. When we replace Cesàro sums in (1.2) and (1.1) by their logarithmic versions,

$$\lim_{N\to\infty}\frac{1}{L_N}\sum_{n\le N}\frac{1}{n}f(T^nx)\boldsymbol{\mu}(n)=0$$

and

$$\lim_{N\to\infty}\frac{1}{L_N}\sum_{n\le N}\frac{1}{n}\boldsymbol{\mu}^{s_1}(n+a_1)\cdots\boldsymbol{\mu}^{s_k}(n+a_k)=0,$$

where $L_N=\sum_{n\le N}\frac{1}{n}$, we obtain *logarithmic* Sarnak's and the Chowla conjectures, respectively. The first striking result was obtained by Tao in 2015 (cf. the corresponding knowledge about MOC, i.e., Proposition 1.2):

Theorem 1.7 ([36]). *The logarithmic Chowla conjecture and the logarithmic Sarnak's conjecture are equivalent.*

Hence Sarnak's conjecture implies the logarithmic Chowla conjecture, and in [17] it is proved that the logarithmic Chowla conjecture implies the Chowla conjecture along a subsequence, hence Proposition 1.2 follows. The logarithmic Sarnak's conjecture is still open, but a significant progress has been achieved by Frantzikinakis and Host in [14] (in 2018). In that paper, the authors were able to relate logarithmic Furstenberg systems of the Möbius function (and many other strongly aperiodic multiplicative functions) to the theory of strongly stationary processes. They basically observed the following principle: either such a system is ergodic and then it must be $\hat{\nu}_{\boldsymbol{\mu}^2}$ or the corresponding Furstenberg system is disjoint from all ergodic systems. By proving new disjointness theorems in ergodic theory, this led them to the following remarkable result:

Theorem 1.8 ([14]). *All zero-entropy systems (X,T) for which the set $M^e(X,T)$ of ergodic invariant measures is countable are logarithmically Möbius orthogonal. In particular, all zero-entropy uniquely ergodic systems are logarithmically Möbius orthogonal.*

In Tao's proof of Theorem 1.7 an important step was to show the equivalence with the third condition which resembles the strong MOMO property (which we discussed above) uniformly with respect to all nil-rotations of a fixed nil-manifold. In fact, one of surprising consequences of Theorem 1.6, which also uses previous results by Tao [35] and Frantzikinakis [13] reduces the logarithmic Sarnak's conjecture to "merely" algebraic situation.

Theorem 1.9 ([24]). *The logarithmic Sarnak's conjecture holds if and only if all systems (X,T) for which each member of $M^e(X,T)$ yields a nil-system are logarithmically Möbius orthogonal.*

Sarnak's conjecture – where are we stuck? Returning to the original MOC, we would like first to notice that there is no result comparable to Frantizkinakis–Host's theorem (Theorem 1.8). In fact, only few general results concerning large classes of zero entropy systems which are Möbius orthogonal are known, namely, besides the already mentioned discrete spectrum case, MOC holds for systems whose all invariant measures yield rigid systems (with some arithmetic limitations on the arithmetics of rigidity sequences) [25] (the polynomial mean complexity characterization from [22] is for the logarithmic case).

A quick look at [24] shows that, at the moment, we are stuck with MOC since (surprisingly) we are not able to prove the strong MOMO property for zero-entropy algebraic automorphisms of the tori. We speak about very special unipotent systems, the simplest (nontrivial) one being $X = \mathbb{T}^2$ and $T(x, y) = (x, x + y)$.[14] It is almost obvious that such systems are Möbius orthogonal (apply, for example, the DDKBSZ criterion) but the situation changes dramatically if we try to prove the strong MOMO property for T (cf. Theorem 1.4 (iii)). In fact, the (potential) strong MOMO property applied to the function $(x, y) \mapsto e^{2\pi i y}$ gives the following:

$$\lim_{K\to\infty} \frac{1}{b_K} \sum_{k<K} \sup_{x\in\mathbb{T}} \left| \sum_{b_k \le n < b_{k+1}} \mu(n) e^{2\pi i n x} \right| = 0,$$

for each sequence (b_k) satisfying $b_{k+1} - b_k \to \infty$. This reminds of a version of the classical Davenport's estimate[15] [7] but the sup inside makes it completely open (see also the discussion on the averaged form of Chowla conjecture in [31]).

2. ERGODIC THEORY – BASIC CONCEPTS

Given a standard Borel probability space $(X, \mathcal{B}, \mu)$, we consider *automorphisms* T of it and the quadruple $(X, \mathcal{B}, \mu, T)$ is often called a *dynamical system*. That is, $T : X \to X$ is invertible, bi-measurable,[16] and $\mu(A) = \mu(T^{-1}A) = \mu(TA)$ for each $A \in \mathcal{B}$. If S is another automorphism (acting on $(Y, \mathcal{C}, \nu)$) then S is a *factor* of T if there exists a measurable $\phi : X \to Y$ which is equivariant, i.e., $\phi \circ T = S \circ \phi$, pushing forward μ onto ν, i.e., $\phi_*(\mu) = \nu$.[17] Then, we obtain a (unique) disintegration of μ over ν:

$$\mu = \int_Y \mu_y \, d\nu(y) \tag{2.1}$$

with μ_y being probability measures on $(X, \mathcal{B})$ concentrated on $\phi^{-1}(y)$ (it is not hard to see that $T_*(\mu_y) = \mu_{Sy}$ for ν-a.e. $y \in Y$). If ϕ is invertible, then T and S are *isomorphic*.

An automorphism T is called *ergodic* if whenever $T^{-1}A = A$ (a.e.) then $\mu(A)$ equals zero or one. But in general, of course, T is nonergodic. In this situation, we consider its ergodic decomposition, which is simply the distintegration (2.1) of μ over the factor $(X/\mathcal{I}, \mathcal{I}, \mu|_{\mathcal{I}}, \mathrm{Id})$, where $\mathcal{I}$ stands for the σ-algebra of invariant sets.

[14] The reader can notice that the ergodic measures for T yield either irrational rotations or finite (cyclic) rotations. There are uncountably many ergodic measures.

[15] The estimate is $\sup_{t\in\mathbb{T}} |\sum_{n\le N} \mu(n)e^{2\pi i n t}| = O(N/\log^A N)$ for each $A > 0$.

[16] More precisely, if needed, we complete $\mathcal{B}$, and we can also assume that T is well defined only on a T-invariant subset $X_0 \subset X$ of full measure. Generally, in what follows we do not distinguish between sets, functions, etc., if they differ on a subset of measure zero.

[17] Note that, setting $\mathcal{A} = \phi^{-1}(\mathcal{C})$, we can represent S as T acting on $(X/\mathcal{A}, \mathcal{A}, \mu|_{\mathcal{A}})$, where "points" in $X/\mathcal{A}$ are cosets of the relation on X of being indistinguishable by the sets of $\mathcal{A}$. By that reason, factors of T are identified with T-invariant sub-σ-fields of $\mathcal{B}$.

With T we can associate a unitary operator U_T, called Koopman operator, acting on $L^2(X,\mathcal{B},\mu)$ by the formula $U_T(f)=f\circ T$. Studying the properties of Koopman operators is the *spectral theory* of dynamical systems. It is not hard to see that ergodicity means precisely that the only invariant functions of U_T are the constants. An automorphism T is called *weakly mixing* if its Cartesian square $T\times T$ acting on $(X\times X,\mathcal{B}\otimes\mathcal{B},\mu\otimes\mu)$ is ergodic. This is equivalent to the fact that the *spectral measure*[18] of each zero mean f, i.e., of $f\in L^2_0(X,\mathcal{B},\mu)$, is atomless. If the Fourier transforms of elements from L^2_0 vanish at infinity, we speak about *mixing* of T.

Given two automorphisms T and S acting on $(X,\mathcal{B},\mu)$, $(Y,\mathcal{C},\nu)$, respectively, by a *joining* between them we mean any measure ρ on $(X\times Y,\mathcal{B}\otimes\mathcal{C})$ with the coordinate projections μ, ν, respectively, and being $T\times S$-invariant. Denote the set of joinings by $J(T,S)$ which is always nonempty as $\mu\otimes\nu\in J(T,S)$. If T and S are additionally ergodic, we can ask about the subset $J^e(T,S)$ of ergodic joinings. This set is nonempty as the ergodic decomposition of any joining consists (a.e.) of joinings. A crucial concept here is that of disjointness introduced by Furstenberg [15] in 1967: we say that T and S are *disjoint*, $T\perp S$, if $J(T,S)=\{\mu\otimes\nu\}$. One should stress that to have disjointness of T and S, at least one of these automorphisms must be ergodic. Note also that if T and S are disjoint then they cannot have a nontrivial common factor (the converse to this implication does not hold). It is not hard to see that if T and S are *spectrally* disjoint, that is, if their maximal spectral types on the corresponding L^2_0-spaces are mutually singular, then $T\perp S$. This yields, in particular, classical disjointness results: identity Id is disjoint with all ergodic automorphisms, discrete spectrum automorphisms (i.e., those whose Koopman operators possess an orthonormal basis consisting of eigenvectors) are disjoint from weakly mixing automorphisms. For more classical examples of automorphisms and instances of disjointness, see [16].

We can repeat the above concepts almost word for word in case of actions of the group $\mathbb{R}$ (or other locally compact Abelian groups) on $(X,\mathcal{B},\mu)$, remembering that we consider only *measurable* actions of $\mathbb{R}$, called *flows* $\mathcal{T}=(T_t)_{t\in\mathbb{R}}$: the map

$$X\times\mathbb{R}\ni(x,t)\mapsto T_tx\in X$$

is measurable. This assumption yields that $t\mapsto U_{T_t}f$ is continuous in the strong topology for each $f\in L^2(X,\mathcal{B},\mu)$. We also recall that the spectral measures of the corresponding Koopman representations are defined on the dual of the acting group, hence on $\mathbb{R}$ in case of flows.

Given $p\in\mathbb{R}^+$, the flow $\mathcal{T}_p:=(T_{pt})_{t\in\mathbb{R}}$ is a *rescaling* of the original flow $\mathcal{T}$. It is not hard to see that the disjointness of the rescaling flows $\mathcal{T}_p$ and $\mathcal{T}_q$ $(0<p<q)$ is equivalent to the disjointness of the time-p and time-q automorphisms, i.e., of T_p and T_q.

[18] A spectral measure σ_f is a finite (nonnegative) Borel measure on the circle $\mathbb{S}^1$ whose Fourier transform is given by

$$\hat{\sigma}_f(n):=\int z^n\,d\sigma_f(z)=\langle U_T^n f,f\rangle=\int_X f\left(T^nx\right)\overline{f(x)}\,d\mu(x)$$

for each $n\in\mathbb{Z}$. Among the spectral measures there are the *maximal* ones (in the sense of absolute continuity of measures); each of them is a measure of *maximal spectral type*.

3. MEASURE-THEORETIC DYNAMICAL SYSTEMS – CONSTRUCTIONS AND EXAMPLES

3.1. Topological dynamics. Subshifts. Invariant measures for homeomorphisms

In topological dynamics we study homeomorphisms T acting on compact metric spaces X; (X,T) is a *topological dynamical system.* Such a system is called *transitive* if there is a point $x_0 \in X$ whose orbit $\{T^n x_0 : n \in \mathbb{Z}\}$ is dense. When all orbits are dense, the system is called *minimal*. The latter is equivalent to the fact that (X,T) has no proper subsystems. If A is a compact metric space, then $A^{\mathbb{Z}}$ considered with the product metric is also compact and $(A^{\mathbb{Z}}, S)$ with S the left *shift*, $S((x_n)_{n\in\mathbb{Z}}) = (x_{n+1})_{n\in\mathbb{Z}}$, is a topological dynamical system called the *full shift* (with the set of states A). Then every closed subset $X \subset A^{\mathbb{Z}}$ which is S-invariant yields a subsystem (X,S) of the full shift, so called *subshift*. By taking first $y = (y_n)_{n\in\mathbb{Z}} \in A^{\mathbb{Z}}$ and setting

$$X_y := \overline{\{S^n y : n \in \mathbb{Z}\}},$$

we obtain (X_y, S) a transitive subshift. In particular, we obtain $(X_{\boldsymbol{\mu}^2}, S)$, called the *square-free system*, where $X_{\boldsymbol{\mu}^2} \subset \{0,1\}^{\mathbb{Z}}$, and $(X_{\boldsymbol{\mu}}, S)$, called the *Möbius subshift*, where $X_{\boldsymbol{\mu}} \subset \{-1,0,1\}^{\mathbb{Z}}$.

The notions of a (topological) factor and isomorphism (conjugacy) are defined similarly to the measure-theoretic category remaining in the class of continuous maps. An important invariant of topological conjugacy is that of entropy $h(T) = h(X,T)$. We refer the reader to [39] for general definitions, however, if A is finite and (X,S) is a subshift, then

$$h(X,S) = \lim_{N\to\infty} \frac{1}{N} \log\left|\mathcal{L}(X) \cap A^N\right|,$$

where $\mathcal{L}(X)$ is the *language* of X, i.e., the set of all words (blocks) appearing in $x \in X$. Clearly, if $X = X_y$, it is enough to compute only words appearing in y.

A topological dynamical system (X,T) yields measure-theoretic dynamical systems through Borel T-invariant measures: if $M(X,T)$ stands for the set of Borel T-invariant measures and $\nu \in M(X,T)$, then it yields a measure-theoretic dynamical system $(X, \mathcal{B}, \nu, T)$, where $\mathcal{B} = \mathcal{B}(X)$ denotes the σ-algebra of Borel subsets of X. We will detail slightly on that. Let $M(X)$ denote the space of probability measures on X. With the weak-$*$-topology, it becomes a metrizable compact space: $\mu_n \to \mu$ if and only $\lim_{n\to\infty} \int_X f\,d\mu_n = \int_X f\,d\mu$ for each $f \in C(X)$. If $x \in X$ then the measures of the form $\frac{1}{N}\sum_{n<N} \delta_{T^n x}$ are called *empiric* measures. Note that any limit ν of a convergent subsequence of empiric measures,

$$\frac{1}{N_k} \sum_{n<N_k} \delta_{T^n x} \to \nu, \tag{3.1}$$

must be a T-invariant measure (one says also that x is *quasi-generic* for ν). This is the classical Krylov–Bogoljubov theorem which tells us that the set $M(X,T)$ is nonempty. It automatically yields that the set $M^e(X,T)$ of ergodic measures is also nonempty. Note that another way to obtain an invariant measure is to change the Cesàro way of summation into the logarithmic one,

$$\frac{1}{L_{N_k}}\sum_{n<N_k}\frac{1}{n}\delta_{T^n x}\to\nu,$$

where the limit measure is also T-invariant. We say that $x\in X$ is *generic for* $\nu\in M(X,T)$ *along* (N_k) if (3.1) holds. For example, $\boldsymbol{\mu}^2$ is generic (along the whole sequence of natural numbers) for the so-called Mirsky measure $\nu_{\boldsymbol{\mu}^2}$ and if the Chowla conjecture holds then $\boldsymbol{\mu}$ is generic for the relatively independent extension $\hat{\nu}_{\boldsymbol{\mu}^2}$ of the Mirsky measure, where

$$\hat{\nu}_{\boldsymbol{\mu}^2}(C)=\frac{1}{2^{\mathrm{supp}(C)}}\nu_{\boldsymbol{\mu}^2}(C^2)$$

for each block C of -1, 0, 1s. We recall that each ergodic measure has a generic point. The set of measures (called also "visible") for which x is quasi-generic is denoted by $V(x)$ and it is compact. Note that either $|V(x)|=1$ (we say then that x is generic for this unique measure) or $V(x)$ is uncountable as it is also connected. Note also that

$$M^e(X,T)\subset V(X,T):=\bigcup_{x\in X}V(x).$$

If $y\in A^{\mathbb{Z}}$ and $\nu\in V(y)$ then the measure-theoretic system $(X_y,\mathcal{B}(X_y),\nu,S)$ is called a *Furstenberg system of* y.

We can introduce similar notions (and prove similar facts) for the logarithmic way of averaging. In general, there is no relation between $V(x)$ and $V^{\log}(x)$ unless x is generic for $\nu\in V(x)$ (it is then logarithmically generic for the same measure).

A topological system (X,T) is *uniquely ergodic* if $|M(X,T)|=1$. The unique invariant measure is then necessarily ergodic. Uniquely ergodic and minimal systems are called *strictly ergodic*. The classical Jewett–Krieger theorem tells us that each ergodic system has a strictly ergodic model.

3.2. Flows, special flows, change of time

Let us first see how, given a flow, to produce new flows with the same orbits but (potentially) representing completely different (even disjoint!) dynamics. Assume that $\mathcal{R}=(R_t)$ is a flow on $(Z,\mathcal{D},\kappa)$ and let $v:Z\to\mathbb{R}$, $v\geq\varepsilon_0>0$ and $v\in L^1(Z,\mathcal{D},\kappa)$. Then, for κ-a.e. $z\in Z$ and all $t\in\mathbb{R}$, there is a unique solution $u=u(t,z)$ of

$$\int_0^u v(R_s z)\,ds=t.$$

Then we set $\tilde{R}^v_t(z)=R_{u(t,z)}(z)$ and obtain a new flow $\widetilde{\mathcal{R}^v}=(\tilde{R}^v_t)$ which preserves the measure $(\frac{v}{\int v\,d\kappa})d\kappa$ for which $u(t,z)$ is a cocycle. On the other hand, $(u,x)\mapsto\int_0^u v(R_s x)\,ds$ defines a cocycle for $\mathcal{R}$. If $v':Z\to\mathbb{R}$ is another time change and, for some measurable $\xi:Z\to\mathbb{R}$,

$$\int_0^u v(R_s z)\,ds=\int_0^u v'(R_s z)\,ds-\xi(z)+\xi(R_u z)$$

for κ-a.e $z\in Z$ and all $u\in\mathbb{R}$ (that is, the two cocycles for $\mathcal{R}$ are cohomologous), then the two time changes $\widetilde{\mathcal{R}^v}$, $\widetilde{\mathcal{R}^{v'}}$ are isomorphic. If $v'=c$ is additionally a constant (that is, the cocycle given by v' is a quasi-coboundary), then $\widetilde{\mathcal{R}^{v'}} = \widetilde{\mathcal{R}^{c}} = (R_{t/c})_{t\in\mathbb{R}}$, so this time change is isomorphic to a rescaling of the original flow $\mathcal{R}$.

We now invoke a construction transforming $\mathbb{Z}$-actions (automorphisms) into $\mathbb{R}$-actions (flows) which is a kind of inducing representation.[19] Assume that $(X, \mathcal{B}, \mu, T)$ is a dynamical system and let $f : X \to \mathbb{R}^+$ be an L^1-function. Consider the probability space $(X^f, \mathcal{B}^f, \mu^f)$, where

$$X^f = \{(x,r) \in X \times \mathbb{R} : x \in X, 0 \le r < f(x)\}$$

with $\mathcal{B}^f$ being the restriction of the product σ-algebra, and $\mu^f := (\mu \otimes \lambda_{\mathbb{R}})|_{X^f} / \int_X f\, d\mu$. We now define the *special flow over T under the roof function f* by setting

$$T_t^f(x,r) = \left(T^n x, r + t - f^{(n)}(x)\right),$$

where $n \in \mathbb{Z}$ is unique such that

$$f^{(n)}(x) \le r + t < f^{(n+1)}(x)$$

and $f^{(n)}(x) = f(x) + f(Tx) + \cdots + f(T^{n-1}x)$ if $n > 0$, $f^{(0)}(x) = 0$ and $f^{(m+n)}(x) = f^{(m)}(x) + f^{(n)}(T^m x)$ for $m, n \in \mathbb{Z}$.

If $f = 1$ then we speak about the *suspension flow $\hat{T}$* over T,

$$\hat{T}_t(x,r) = \left(T^{[t+r]}x, \{t + r\}\right),$$

for $(x,r) \in X \times [0,1)$. Note that for $k \in \mathbb{N}$,

$$\int_0^k f\left(\hat{T}_s(x,0)\right) ds = \int_0^k f\left(T^{[s]}x\right) ds = f^{(k)}(x)$$

allows us to see the special flow T^f as a time change of the suspension flow over T. It follows that, given two special flows over the same automorphism T, we can obtain one from the other by a time change.

The Kakutani–Ambrose theorem tells us that each flow has a special representation. Representing a flow as a special flow (over a “known” automorphism) is a useful operation, and finding T, especially in the smooth case, leads to seeking a good transversal to orbits of the original flow. For example, in the case of smooth flows on surfaces it often leads to the study of special flows over interval exchange transformations and interesting roof functions having “controllable” singularities.

4. RATNER’S QUESTION, MW-STRATEGY, AND MOC FOR SMOOTH TIME CHANGES OF HOROCYCLE FLOWS

4.1. Horocycle flows and MOC

One of the most important zero-entropy classes in dynamics is given by horocycle flows whose definition we now recall. Let $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ be a discrete subgroup with finite

[19] The reader can check that the Koopman representation of the special flow defined below is indeed the genuine induced representation of the Koopman operator associated to the automorphism.

covolume, in fact, we consider only the case Γ is cocompact, so that the homogeneous space $M = \Gamma\backslash \mathrm{PSL}_2(\mathbb{R})$ is compact and then the system is uniquely ergodic. Let us consider the corresponding *horocycle flow* $(h_t)_{t\in\mathbb{R}}$ and the *geodesic flow* $(g_t)_{t\in\mathbb{R}}$ on M given by

$$h_t(\Gamma x) = \Gamma \cdot \left(x \cdot \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}\right) \quad \text{and} \quad g_t(\Gamma x) = \Gamma \cdot \left(x \cdot \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{t} \end{bmatrix}\right).$$

Since

$$g_s h_t g_s^{-1} = h_{e^{-2s}t} \quad \text{for all } s, t \in \mathbb{R}, \tag{4.1}$$

the flows $(h_t)_{t\in\mathbb{R}}$ and $(h_{e^{-2s}t})_{t\in\mathbb{R}}$ are measure-theoretically isomorphic for each $s \in \mathbb{R}$ (in particular, all positive time automorphisms are isomorphic). In 2011, Bourgain, Sarnak, and Ziegler proved the following:

Theorem 4.1 ([4]). *Each time-t automorphism h_t is Möbius orthogonal.*

The main idea is to use the MW-strategy, and to show that, in fact, these time-t automorphisms are orthogonal to *any* (zero mean, bounded) multiplicative function. It works here because of the famous Ratner's theory: given $x \in \mathrm{PSL}_2(\mathbb{R})$, any point $(\Gamma x, \Gamma x)$ is generic for a measure ρ (which must be a joining by unique ergodicity: $\rho \in J(T^p, T^q)$, where $T = h_t$) and, moreover, this joining is ergodic and of algebraic nature. As shown in [4], this algebraic nature yields that, perhaps except for finitely many primes, we must obtain the product measure.[20] The proof really depends on some algebraic properties of horocycle flows and because of that M. Ratner asked in 2013 what happens if we (smoothly) change time and study MOC in this class.

4.2. Time changes of horocycle flows and MOC

In general, especially for flows which are mixing, it is difficult to decide whether or not they are disjoint. Horocycle flows are mixing and so are their smooth time changes. In 1983, M. Ratner [32] discovered a new property of horocycle flows which basically gave a quadratic way of divergence of distinct orbits of nearby points, which allows one to observe some drift of these orbits. This geometric property has surprisingly strong rigidity joining consequences. Ratner also showed that smooth time changes of horocycle flows enjoy this divergence property [33]. It took more than 20 years to understand how to variate her original property (keeping the joining consequences) to now commonly called Ratner's properties, to observe quantitatively the drift phenomenon also beyond the horocyclic world, in particular to see it in dimension 2 (e.g., for some smooth flows on surfaces). A kind of a breakthrough new disjointness criterion has been recently proved in [26]. It is tailored for flows (with a Ratner's property) having different speed of divergence (polynomial or subpolynomial) of distinct orbit of close points. It fits to nontrivial smooth time changes $(\tilde{h}_t^v)$ of horocycle flows as one of the main results of [26] shows:

[20] This might suggest that we have the AOP property, but, in fact, as noticed in [3], (4.1) applied to some compact regions implies that AOP fails for horocycle flows.

Theorem 4.2 ([26]). *Assume that the cocycle determined by a positive $v \in W^6(M)$ has a nontrivial support outside of the discrete series*[21] *and is not a quasi-coboundary. Then, for any real numbers $0 < p < q$, the rescalings $(\tilde{h}^v_{pt})_{t\in\mathbb{R}}$ and $(\tilde{h}^v_{qt})_{t\in\mathbb{R}}$ are disjoint.*

The situation looks a little bit paradoxical as, for the horocycle flows themselves, we know that they are Möbius orthogonal, but the problem of whether the convergence in (1.2) is uniform (in $x \in M$) is open, the strong MOMO property is open, and we also do not know whether the Möbius orthogonality takes place in *all* uniquely ergodic models of horocycle flows. On the other hand, when we change time (as above), for the flows whose dynamics intuitively become more complicated, the answers to these questions are simply positive due to Theorem 4.2, Proposition 1.5, and Theorem 1.4 (ii).

Using the same disjointness criterion, other disjointness results concerning some mixing locally Hamiltonian flows (on surfaces), considered most often in their special representations over irrational rotations (Arnol'd special flows) are proved in [26] to enjoy similar internal disjointness properties. Hence, the Möbius orthogonality for them is also established.

5. SARNAK'S CONJECTURE AND FURSTENBERG SYSTEMS

It is not clear at all that the MOC can be expressed in terms of Furstenberg systems of the Möbius function. In fact, following [24], we will consider a general problem in which $\boldsymbol{\mu}$ is replaced by a function $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$ (the unit disc). We want to characterize those $\boldsymbol{u}$ which are orthogonal to all zero-entropy systems,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f(T^n x)\boldsymbol{u}(n) = 0, \tag{5.1}$$

for each (X, T) of zero entropy, all $f \in C(X)$, and $x \in X$. One can now wonder what is special in the zero (topological) entropy class. For that we need to recall some classical facts, namely the variational principle, which tells us that $h(X, T) = 0$ if and only if the measure-theoretic entropy of each T-invariant measure is zero. By a convexity property of the entropy, this is still equivalent to the fact that all ergodic measures have zero entropy. In this way, we replaced the original assumption on (X, T) by an assumption on the *measure-theoretic* systems determined by invariant measures. More than that, while thinking about problem (5.1), we only care about properties of systems determined by *visible* measures $\mu \in V(X, T)$. Finally, one can wonder what is special in the class of measure-theoretic systems with zero entropy. Classical ergodic theory tells us that this is a class which is closed under taking joinings and factors and for each automorphism $(Z, \mathcal{D}, \kappa, R)$ there exists a largest factor $\Pi(\kappa) \subset \mathcal{D}$ of zero entropy, called the Pinsker factor of R. All this leads us to the concept of a characteristic class and the problem of orthogonality to such.

[21] This assumption is dropped in [11]. Flaminio and Forni used more directly Ratner's work [33] and show that the cocycles determined by v and $v \circ g_r$, where $r = -\frac{1}{2}\log(q/p)$ are not jointly cohomologous.

5.1. Characteristic classes and the problem of orthogonality

A class $\mathscr{F}$ of automorphisms (it is implicit that this class is closed under isomorphism) is called *characteristic* if it is closed under taking (countable) joinings and factors. Classical classes, like the zero-entropy class, the class of systems with discrete spectrum, and the class of automorphisms which are distal, are characteristic classes (many more classes are listed in [24]). Once $\mathscr{F}$ is fixed, we consider the class $\mathscr{C}_{\mathscr{F}}$ of those topological systems (X, T) for which $(X, \mathscr{B}(X), \nu, T) \in \mathscr{F}$ for each $\nu \in V(X, T)$. The following theorem establishes the most useful ergodic properties following the concept of a characteristic class.

Theorem 5.1 ([24]). *Assume that $\mathscr{F}$ is a characteristic class then, for each automorphism $(Z, \mathscr{D}, \kappa, R)$, there exists a largest factor $\mathscr{D}_{\mathscr{F}} \subset \mathscr{D}$ belonging to $\mathscr{F}$. Moreover, any joining of $(Z, \mathscr{D}, \kappa, R)$ with an automorphism from $\mathscr{F}$ is uniquely determined by its restriction to a joining with $\mathscr{D}_{\mathscr{F}}$.*

With each class $\mathscr{F}$, we can associate the class $\mathscr{F}_{\mathrm{ec}}$ consisting of the automorphisms whose all ergodic components are in $\mathscr{F}$.

Proposition 5.2 ([24]). *If $\mathscr{F}$ is characteristic, then also $\mathscr{F}_{ec}$ is characteristic.*

In general, we then have $\mathscr{C}_{\mathscr{F}} \subset \mathscr{C}_{\mathscr{F}_{\mathrm{ec}}}$, but the reader can check that if $\mathscr{F}$ is the zero-entropy class then we have equality. Moreover,

$$\mathscr{C}_{\mathscr{F}_{\mathrm{ec}}} = \big\{(X, T) : \big(X, \mathscr{B}(X), \nu, T\big) \in \mathscr{F} \text{ for each } \nu \in M^e(X, T)\big\}.$$

The zero-entropy class turns out to be special in the family of characteristic classes:

Proposition 5.3 ([24]). *The zero-entropy class is the largest proper characteristic class.*

It is also shown in [24] that there exists the smallest nontrivial characteristic class (it consists of all identities of standard Borel probability spaces).

5.2. Orthogonality to characteristic classes. Veech's conjecture

Given a class $\mathscr{C}$ of topological systems, we can now consider the problem of orthogonality of $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$ to $\mathscr{C}$, that is,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f(T^n x)\boldsymbol{u}(n) = 0 \tag{5.2}$$

for each $(X, T) \in \mathscr{C}$, all $f \in C(X)$ and $x \in X$. If orthogonality takes place, we write $\boldsymbol{u} \perp \mathscr{C}$. The central result of [24] is the following:

Theorem 5.4 ([24]). *Assume that $\mathscr{F}$ is a characteristic class and $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$. Then $\boldsymbol{u} \perp \mathscr{C}_{\mathscr{F}_{ec}}$ if and only if*

$$\pi_0 \perp L^2\big(\big(\mathscr{B}(X_{\boldsymbol{u}}), \kappa\big)_{\mathscr{F}_{ec}}\big) \quad \textit{for each Furstenberg system } \kappa \in V(\boldsymbol{u}). \tag{5.3}$$

Remark 5.5. Condition (5.3) will be called the *Veech condition* as Veech formulated it in [38], in a form of a conjecture, as a statement equivalent to MOC in case of $\boldsymbol{u} = \boldsymbol{\mu}$ and $\mathscr{F}$

equal to the (measure-theoretic) zero-entropy class. Theorem 5.4 proves in particular Veech's conjecture but, clearly, goes beyond it.

In the subsequent subsections we will say a few words on the tools that are employed for the proof of Theorem 5.4 and briefly indicate some consequences of it.

5.3. Proof of Theorem 5.4

The sufficiency in Theorem 5.4 follows from a more general result:

Theorem 5.6 ([24]). *If $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$ satisfies the Veech condition with respect to a characteristic class $\mathscr{F}$ then $\boldsymbol{u} \perp \mathscr{C}_{\mathscr{F}}$.*

The proof of this theorem is purely ergodic, belongs to joining theory, and is based on a fundamental non-disjointness lemma [29].[22]

The necessity requires more tools. The first relies on the existence of the so-called Hansel's models, being a counterpart of the classical Jewett–Kieger theorem in the nonergodic case. Namely, if $(Z, \mathscr{D}, \nu, R)$ is a measure-theoretic dynamical system and we fix a set of full measure of ergodic components then a Hansel model [18] of it is any topological system (X, T) for which there exists $\nu \in M(X, T)$ yielding a measure-theoretic isomorphic copy of R and such that *each* $x \in X$ is generic for one of the chosen ergodic components. The next major step is a new lifting lemma[23] (going largely beyond the context considered in [6]) on quasi-generic points for joinings, and tailored to be applicable for the strong MOMO property:

Lemma 5.7 ([24]). *Assume that (Y, S) and (X, T) are topological systems. Let $\nu \in M(X, T)$, $u \in Y$ be generic along an increasing sequence (N_m) for $\kappa \in M(Y, S)$, and $\rho \in J(\kappa, \nu)$. Then there exist a sequence $(x_n) \subset X$ and a subsequence (N_{m_ℓ}) such that $(S^n u, x_n)$ is generic along (N_{m_ℓ}) for ρ and the set $\{n \geq 0 : x_{n+1} \neq T x_n\}$ is of the form (b_k) with $b_{k+1} - b_k \to \infty$ when $k \to \infty$.*

We then use some joining techniques and Lemma 5.7 to Hansel models of the largest $\mathscr{F}_{\text{ec}}$-factors of Furstenberg systems of $\boldsymbol{u}$. Finally, the reason why we use $\mathscr{F}_{\text{ec}}$ (and not $\mathscr{F}$ itself) is that the orthogonality of $\boldsymbol{u}$ to $\mathscr{C}_{\mathscr{F}_{\text{ec}}}$ is *equivalent* to the strong MOMO property (relative to $\boldsymbol{u}$) of all systems in $\mathscr{C}_{\mathscr{F}_{\text{ec}}}$ (such a result, in full generality, is unknown for $\mathscr{F}$).

5.4. Some consequences of Theorem 5.4

We now come back to the problem of orthogonality of $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$ to the zero (topological) entropy class (MOC is a particular case of this). In this case, Theorem 5.4 describes a kind of relative Kolmogorov property which, by some ergodic considerations, can

[22] Veech in [38], only for the zero entropy class, gives a rather complicated proof based on the concept of quasi-factors by Glasner and Weiss. For this particular class, the proof is also implicit in [1].

[23] The lemma is also valid for the logarithmic way of averaging, which seems to be the first result of that type in the literature.

be replaced by so called (relative) K-mixing. We will now write a combinatorial reformulation of the latter property assuming (for sake of simplicity) that there is only one Furstenberg system of $\boldsymbol{u}$:

Corollary 5.8 ([24]). *If $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$ is generic then $\boldsymbol{u}$ is orthogonal to all zero-entropy systems if and only if*

$$\lim_{m\to\infty} \lim_{N\to\infty} \left| \frac{1}{N} \sum_{n\le N} \boldsymbol{u}(n) 1_{\boldsymbol{u}(m+n),\boldsymbol{u}(m+n+1),\ldots,\boldsymbol{u}(m+n+\ell-1)\in C} \right| = 0$$

uniformly in $\ell \ge 1$ and in C, a set of blocks of length ℓ.

The above corollary is, of course, about cancelations of $+1$s and -1s along larger and larger shifts of return times to a fixed set of blocks (of a fixed length). By a rather standard argument, it can be replaced with a *conditional* cancelation phenomenon for a single “typical” block.

Another consequence of Theorem 5.4 is a purely ergodic proof of the so-called averaged Chowla property shown first (even in the quantitative version) in [31] for the Möbius function: for each $\boldsymbol{u} : \mathbb{N} \to \mathbb{D}$, for which all circle rotations satisfy the strong MOMO (relative to $\boldsymbol{u}$) property,[24] we have

$$\lim_{H\to\infty} \frac{1}{H^k} \sum_{h_1,\ldots,h_k\le H} \lim_{k\to\infty} \frac{1}{N_k} \left| \sum_{n\le N_k} \boldsymbol{u}(n) \prod_{i=1}^{k} c_i(n+h_i) \right| = 0$$

for all sequences $c_i : \mathbb{N} \to \mathbb{D}, i = 1, \ldots, k$.

The strength of Theorem 5.4 also follows from the fact that it is valid in the logarithmic context which is better understood. As we have already mentioned in the introduction, using it together with some earlier results by Tao and Frantzikinakis yields Theorem 1.9.

FUNDING

This work was partially supported by Narodowe Centrum Nauki grant UMO-2019/33/B/ST1/00364.

REFERENCES

[1] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue, The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discrete Contin. Dyn. Syst.* **37** (2017), 2899–2944.

[2] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue, Möbius disjointness for models of an ergodic system and beyond. *Israel J. Math.* **228** (2018), 707–751.

[3] H. El Abdalaoui, M. Lemańczyk, and T. de la Rue, Automorphisms with quasi-discrete spectrum, multiplicative functions and average orthogonality along short intervals. *Int. Math. Res. Not. IMRN* **2017** (2017), no. 14, 4350–4368.

[24] We recall that this takes place for the Möbius function.

[4] J. Bourgain, P. Sarnak, and T. Ziegler, Disjointness of Möbius from horocycle flows. In *From Fourier and number theory to radon transforms and geometry, in memory of Leon Ehrenpreiss*, pp. 67–83, Dev. Math. 28, Springer, 2012.

[5] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*. Math. Appl. 4, Gordon and Breach Science Publishers, New York, 1965.

[6] J.-P. Conze, T. Downarowicz, and J. Serafin, Correlation of sequences and of measures, generic points for joinings and ergodicity of certain cocycles. *Trans. Amer. Math. Soc.* **369** (2017), 3421–3441.

[7] H. Davenport, On some infinite series involving arithmetical functions. II. *Quart. J. Math. Oxford* **8** (1937), 313–320.

[8] T. Downarowicz and J. Serafin, Almost full entropy subshifts uncorrelated to the Möbius function. *Int. Math. Res. Not. IMRN* **2019** (2019), no. 11, 3459–3472.

[9] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*. Grad. Texts in Math. 259, Springer, London, 2011.

[10] S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk, Sarnak's conjecture – what's new. In *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics, CIRM Jean-Morlet Chair, Fall 2016*, edited by S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk, p. 418, Lecture Notes in Math. 2213, Springer, 2018.

[11] L. Flaminio and G. Forni, Orthogonal powers and Möbius conjecture for smooth time-changes of horocycle flows. *Electron. Res. Announc. Math. Sci.* **26** (2019), 16–23.

[12] L. Flaminio, K. Frączek, J. Kułaga-Przymus, and M. Lemańczyk, Approximate orthogonality of powers for ergodic affine unipotent diffeomorphisms. *Studia Math.* **244** (2019), 43–97.

[13] N. Frantzikinakis, *Ergodicity of the Liouville system implies the Chowla conjecture*. Discrete Analysis 2017, paper 8, p. 23.

[14] N. Frantzikinakis and B. Host, The logarithmic Sarnak conjecture for ergodic weights. *Ann. of Math. (2)* **187** (2018), no. 3, 869–931.

[15] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Syst. Theory* **1** (1967), 1–49.

[16] E. Glasner, *Ergodic theory via joinings*. American Mathematical Society, Providence, 2003.

[17] A. Gomilko, D. Kwietniak, and M. Lemańczyk, Sarnak's conjecture implies the Chowla conjecture along a subsequence. In *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, pp. 237–247, Lecture Notes in Math. 2213, Springer, Cham, 2018.

[18] G. Hansel, Strict uniformity in ergodic theory. *Math. Z.* **135** (1974), 221–248.

[19] A. Hildebrand, *Introduction to analytic number theory*. http://www.math.uiuc.edu/~ajh/531.fall05/.

[20] W. Huang, Z. Wang, and X. Ye, Measure complexity and Möbius disjointness. *Adv. Math.* **347** (2019), 827–858.

[21] W. Huang, Z. Wang, and G. Zhang, Möbius disjointness for topological models of ergodic systems with discrete spectrum. *J. Mod. Dyn.* **14** (2019), 277–290.

[22] W. Huang, L. Xu, and X. Ye, Polynomial mean complexity and logarithmic Sarnak conjecture. 2020, arXiv:2009.02090.

[23] H. Iwaniec and E. Kowalski, *Analytic number theory*. Amer. Math. Soc. Colloq. Publ. 53, American Mathematical Society, Providence, RI, 2004.

[24] A. Kanigowski, J. Kułaga-Przymus, M. Lemańczyk, and T. de la Rue, On arithmetic functions orthogonal to deterministic sequences. 2021, arXiv:2105.11737.

[25] A. Kanigowski, M. Lemańczyk, and M. Radziwiłł, Rigidity in dynamics and Möbius disjointness. *Fund. Math.* **255** (2021), 309-336.

[26] A. Kanigowski, M. Lemańczyk, and C. Ulcigrai, On disjointness of some parabolic flows. *Invent. Math.* **221(1)** (2020), 1–111.

[27] I. Kátai, A remark on a theorem of H. Daboussi. *Acta Math. Hungar.* **47** (1986), no. 1–2, 223–225.

[28] J. Kułaga-Przymus and M. Lemańczyk, Sarnak's conjecture from the ergodic theory point of view. In *Encyclopedia for Complexity and Systems Science*, published online 2021, arXiv:2009.04757.

[29] M. Lemańczyk, F. Parreau, and J.-P. Thouvenot, Gaussian automorphisms whose ergodic self-joinings are Gaussian. *Fund. Math.* **164** (2000), 253–293.

[30] K. Matomäki and M. Radziwiłł, Multiplicative functions in short intervals. *Ann. of Math. (2)* **183** (2016), 1015–1056.

[31] K. Matomäki, M. Radziwiłł, and T. Tao, An averaged form of Chowla's conjecture. *Algebra Number Theory* **9** (2015), 2167–2196.

[32] M. Ratner, Horocycle flows, joinings and rigidity of products. *Ann. of Math. (2)* **118** (1983), 277–313.

[33] M. Ratner, Rigid reparametrizations and cohomology for horocycle flows. *Invent. Math.* **88** (1987), no. 2, 341–374.

[34] P. Sarnak, *Three lectures on the Möbius function, randomness and dynamics*. http://publications.ias.edu/sarnak/.

[35] T. Tao, *The Chowla and the Sarnak conjecture*, What's new. http://terrytao.wordpress.com/2012/10/14/the-chowla-conjecture-and-the-sarnak-conjecture/.

[36] T. Tao, Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In *Number theory – diophantine problems, uniform distribution and applications: Festschrift in honour of Robert F. Tichy's 60th birthday*, edited by C. Elsholtz and P. Grabner, pp. 391–421, Springer, Cham, 2017.

[37] T. Tao, *The logarithmically averaged and non-logarithmically averaged Chowla conjectures*. https://terrytao.wordpress.com/2017/10/20/the-logarithmically-averaged-and-non-logarithmically-averaged-chowla-conjectures/.

[38] W. A. Veech, *Möbius dynamics*. Unpublished lecture notes, Spring semester 2016.

[39] P. Walters, *An introduction to ergodic theory*. Grad. Texts in Math. 79, Springer, New York, 1982.

MARIUSZ LEMAŃCZYK

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Chopin street 12/18, 87-100 Toruń, Poland, mlem@mat.umk.pl

FINITARY ANALYSIS IN HOMOGENEOUS SPACES

AMIR MOHAMMADI

ABSTRACT

In this paper we give an overview of recent developments pertaining to the quantitative aspects of dynamics of group actions on homogeneous spaces.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37A17; Secondary 20G30, 37A45, 22E40, 11B75, 11J25

KEYWORDS

Measure and topological rigidity, unipotent flows, quantitative analysis

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3530–3551
DOI 10.4171/ICM2022/68

1. INTRODUCTION

Dynamical systems have become a major player in several unexpected areas in modern mathematics. Homogeneous spaces and the moduli spaces of compact Riemann surfaces serve as two hubs where techniques from dynamical systems and analysis duel in a nearly magical fashion with the structure provided by the rich geometric, algebraic, and arithmetic properties of the underlying space.

Investigations in these directions have resulted in several breakthrough results with striking applications in other areas of mathematics. However, most of these celebrated achievements share the lacuna that they are not quantitative. It is much anticipated and a challenging task to develop finitary arguments in these contexts; this article aims at providing an overview of some of the quantitative results in this setting.

Let us begin by recalling the general frame work of homogeneous dynamics. Let $G \subset \mathrm{SL}_d(\mathbb{R})$ be a connected linear Lie group, and let $\Gamma \subset G$ be a lattice (a discrete subgroup with finite covolume). Let $W \subset G$ be a closed connected subgroup of G. The following problem has proven to be of fundamental importance:

Describe the behavior of the orbit Wx for every *point $x \in G/\Gamma$.*

Note that we demand information about the orbit of every point in the space not merely a typical point, which is a more common theme in ergodic theory. Note also that in the above generality, one cannot expect a meaningful answer to this problem. For example, if $G = \mathrm{SL}_2(\mathbb{R})$ and W is the group of diagonal matrices in G, then individual orbits can have very complicated behavior and in particular the closure of orbits can be a fractal set, see, e.g., [64].

If W is generated by unipotent elements,[1] however, Raghunathan had conjectured that for every $x \in G/\Gamma$ there exists a connected subgroup $W \subset L \subset G$ so that Lx is *periodic* and the closure of Wx equals Lx; an orbit Lx is periodic if the stabilizer of x in L is a lattice in L, see Section 2.

Raghunaths's conjecture in its full generality was proved by Ratner [88–90]. Prior to Ratner's seminal work, some important special cases of this conjecture were established by Margulis [74], and Dani and Margulis [25,26].

As was alluded to above, these fundamental results are not quantitative, e.g., they do not provide any rate at which the orbit fills up its closure. Indeed Ratner's work relies on the pointwise ergodic theorem which is hard to effectivize. The work of Dani and Margulis uses minimal sets, which though formally ineffective, can be effectivized with some effort. However, this a rather challenging task; moreover, the rates one obtains are often poor, see Section 6 for further discussion.

We note that good effective bounds for equidistribution of unipotent orbits can have far reaching consequences. Indeed, the Riemann hypothesis is equivalent to giving an error

[1] An $d \times d$ matrix is called unipotent if all its complex eigenvalues are 1. A connected subgroup of $\mathrm{SL}_d(\mathbb{R})$ is called unipotent if all its elements are unipotent.

term of the form $O_\varepsilon(y^{\frac{3}{4}+\varepsilon})$ for equidistribution of periodic horocycles of period $1/y$ on the modular surface [91, 100]. Motivated by related but less dramatic applications, one is interested in obtaining rates which have *polynomial nature*. In the generality that will be discussed in Section 6, however, such bounds seem beyond the reach of the current technology. That said, there have been some exciting developments in this direction which will be discussed in the sequel.

We bring this introduction to a close by mentioning that there have also been groundbreaking works in a similar vein to the rigidity phenomena which lie at the heart of this paper, but in different contexts: In fact, the papers [30, 68] concern higher rank diagonalizable flows; the papers [5–7, 12] concern the classification of stationary measures; the papers [40, 41] concern the action of $\mathrm{SL}_2(\mathbb{R})$ on moduli spaces and apply also the method developed for stationary measures; and [4, 66, 79, 80] concern the case where Γ has infinite covolume. These works, with the exception of [12], are all qualitative and *any* effective account of these would be very intriguing.

2. COMPLEXITY OF PERIODIC ORBITS

Let $L \subset G$ be a closed subgroup. A point $x \in X = G/\Gamma$ is called *L-periodic* if

$$\mathrm{Stab}_L(x) = \{g \in L : gx = x\}$$

is a lattice in L. A periodic L-orbit (or simply a periodic orbit if L is clear from the context) is an orbit Lx where x is an L-periodic point. Note that a periodic L-orbit is always closed in X, see [87].

The rigidity results we will discuss here assert that the closure of an orbit Wx is a periodic orbit Lx of an intermediate subgroup $W \subset L \subset G$. It is therefore natural to expect that quantitative statements in this context will in general depend on delicate properties of the point x and the acting group W. Indeed, already for an irrational rotation of a circle, Diophantine properties of the angle of rotation dictate the rate of equidistribution. In the more general context at the heart of our discussions here, periodic orbits of intermediate subgroups will play the role of rational numbers. Consequently, it is crucial to fix a measure of complexity for the periodic orbits which are obstructions to the density of an orbit in X.

Fix some open bounded neighborhood Ω of the identity in G. For a periodic orbit $Lx \subset X$, define

$$\mathrm{vol}(Lx) = \frac{m_L(Lx)}{m_L(\Omega)}, \tag{2.1}$$

where m_L is an arbitrary Haar measure on L and $m_L(Lx)$ is the covolume of $\mathrm{Stab}_L(x)$ in L with respect to m_L. This notion of volume will serve as our measure of the complexity of the periodic orbit,

We refer the reader to [31, §2.3] for basic properties of the above definition. Here we only mention that even though this notion depends on the choice of Ω, two different choices of Ω give rise to comparable definitions of vol, in the sense that their ratio is bounded above and below. Therefore, we ignore the dependence on Ω in the notation.

Given a periodic orbit Lx, we let μ_{Lx} denote the probability L-invariant measure on Lx. The G-invariant probability measure on X will be denoted by m_X.

The general theme of a finitary statement will be a dichotomy as follows: Unless there is an explicit obstruction with *low complexity*, the orbit Wx *fills up* X with an explicit rate—as we will see, the quality of this rate varies in different situations.

3. EFFECTIVE EQUIDISTRIBUTION OF NILFLOWS

Perhaps the first natural place to seek quantitative density theorems is the case of nilflows. Let X be a nilmanifold. That is, $X = G/\Gamma$ where G is a closed connected subgroup of the group of strictly upper triangular $d \times d$ matrices and $\Gamma \subset G$ is a lattice.

Rigidity results in this setting have been known for quite some time thanks to works of Weyl, Kronecker, L. Green, and Parry [2,85], and more recently Leibman [67].

Quantitative results, *with a polynomial error rate*, have also been established in this context and beyond the abelian case, see [46,53]. The complete solution was given by B. Green and T. Tao [53]. The following is a special case of the main result in [53].

Theorem 3.1 ([53]). *Let $X = G/\Gamma$ be a nilmanifold as above. There exists some $A \geq 1$ depending on* $\dim G$ *so that the following holds. Let $x \in X$, let $\{u(t) : t \in \mathbb{R}\}$ be a one-parameter subgroup of G, let $0 < \eta < 1/2$, and let $T > 0$. Then at least one of the following holds for the partial trajectory $\{u(t)x : t \in [0,T]\}$:*

(1) *For every $f \in C^\infty(X)$, we have*

$$\left|\frac{1}{T}\int_0^T f\big(u(t)x\big)\,\mathrm{d}t - \int_X f\,\mathrm{d}m_X\right| \ll_{X,f} \eta,$$

where the dependence on f is given using a certain Lipschitz norm.

(2) *For every $0 \leq t_0 \leq T$, there exist some $g \in G$ and some $H \subsetneq G$ so that $H\Gamma/\Gamma$ is periodic with* $\mathrm{vol}(gH\Gamma/\Gamma) \ll_X \eta^{-A}$ *and for all $t \in [0,T]$ with $|t - t_0| \leq \eta^A T$, we have*

$$\mathrm{dist}_X\big(u(t)x, gH\Gamma/\Gamma\big) \ll_X \eta,$$

where dist_X *is a metric on X induced from a right invariant Riemannian metric on G.*

We refer the reader to [33] for this formulation and the deduction of it from the main result in [53]. Let us, however, highlight here the aforementioned dichotomy: either the orbit $\{u(t)x : t \in [0,T]\}$ is effectively equidistributed, part (1) in Theorem 3.1, or there is an explicit obstruction of low complexity which prevents this, part (2) in Theorem 3.1.

4. HOROSPHERICAL GROUPS

Let G be a connected semisimple Lie group. A subgroup $W \subset G$ is called horospherical if there exists an ($\mathbb{R}$-diagonalizable) element $a \in G$ so that

$$W = W^{+}(a) := \{g \in G : a^{n} g a^{-n} \to e \text{ as } n \to -\infty\}.$$

It is well known that $W \subset G$ is horospherical if and only if it is the unipotent radical of a proper parabolic subgroup of G. In particular, a horospherical subgroup is always unipotent,[2] but not vice versa; indeed, if W is horospherical, then $G/N_G(W)$ is compact where $N_G(W)$ denotes the normalizer of W in G.

The study of the action of a horospherical subgroup of G on G/Γ has a long history, and rigidity theorems á la Ratner in this case were established by Hedlund, Furstenberg, Veech, and Dani [20, 21, 24, 48, 59, 97] prior to Ratner's theorems. Indeed, thanks to the fact that the behavior of individual orbits of a horospherical subgroup can be related to the decay of matrix coefficients, effective equidistribution, with a polynomial error rate, can also be established. The first works in this direction we are aware of are [16, 64, 91], as well as the more recent [45, 93, 95], and this has now been established in much greater generality [62, 63, 78, 82]. Closely related is the case of translates of orbits of subgroups of G which are fixed by an involution [3, 29, 39].

We refer the reader to [78, THM. 3.1] for the case of $\mathrm{SL}_n(\mathbb{R})/\mathrm{SL}_n(\mathbb{Z})$ and to [62, THM. 1.11] for the general case. Let us only mention here that in this case, obstructions to effective equidistribution of Wx, where $W = W^{+}(a)$, can be described using the rate of excursion of $\{a^{-n}x : n \in \mathbb{N}\}$ to infinity. Consequently, quantitative nondivergence of unipotent flows [22, 23, 27, 65, 73] plays a crucial role in the analysis, see also the discussion in Section 6.2. In particular, when $X = G/\Gamma$ is compact, Wx is equidistributed in (X, m_X) with a polynomial rate for *every* point $x \in X$.

Another class of examples where one may attempt to bring properties of horospherical subgroups to bear are provided by semidirect product constructions. Let $G = H \ltimes V$ where H is a noncompact semisimple Lie group and V is an irreducible representation of H. One then investigates the action of a horospherical subgroup $W \subset H$ on G/Γ. This case is significantly more complicated that the case of horospherical subgroups, and only partial progress has been made in this direction. Indeed, Strömbergsson [96] used analytic methods to settle the case of $G = \mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2$ with the standard action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$, $\Gamma = \mathrm{SL}_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$, and W the group of unipotent upper triangular matrices in $\mathrm{SL}_2(\mathbb{R})$; his method has also been used to tackle some other cases.

We end this section by mentioning that ideas developed in the homogeneous setting have also found applications in the study of horospherical foliation (strong unstable foliation) in the space of translation surfaces, see, e.g., [42, 70].

[2] The fact that a horospherical subgroup is unipotent follows readily from the definition.

5. PERIODIC ORBITS OF SEMISIMPLE GROUPS

Until roughly 15 years ago, the source of quantitative treatments in this context could essentially be traced back to the settings discussed in Sections 3 and 4. However, the situation has recently improved. In the remaining parts of this article, we discuss some of these advances.

One of the earliest works in this *new wave* was the landmark paper of Einsiedler, Margulis, and Venkatesh [32] concerning the periodic orbits of semisimple groups. Let $\boldsymbol{G}$ be a connected, semisimple algebraic $\mathbb{Q}$-group, and let G be the connected component of the identity in the Lie group $\boldsymbol{G}(\mathbb{R})$. Let $\Gamma \subset G$ be a congruence subgroup of $\boldsymbol{G}(\mathbb{Q})$, and put $X = G/\Gamma$. Let $H \subset G$ be a semisimple subgroup without any compact factors which has a finite centralizer in G.

The following is the main equidistribution theorem proved in [32].

Theorem 5.1 ([32]). *There exists some $\delta = \delta(G, H)$ so that the following holds. Let Hx be a periodic orbit. For every $V > 1$, there exists a subgroup $H \subset S \subset G$ so that Sx is periodic,* $\operatorname{vol}(Sx) \le V$, *and*

$$\left| \int_X f \, \mathrm{d}\mu_{Hx} - \int_X f \, \mathrm{d}\mu_{Sx} \right| \ll_{G,\Gamma,H} \mathcal{S}(f) V^{-\delta} \quad \textit{for all } f \in C_c^\infty(X),$$

where $\mathcal{S}(f)$ denotes a certain Sobolev norm.

Theorem 5.1 is an effective version (of a special case) of a theorem by Mozes and Shah [84]. The polynomial nature of the error term, i.e., a (negative) power of V, in Theorem 5.1 is quite remarkable—effectivizations of dynamical arguments often yield much worse rates, see Section 6. The source of this polynomial rate is the uniform spectral gap for congruence quotients, which is used as a crucial input in [32].

As it was alluded to already, the fact that one deals with periodic orbits of semisimple groups in arithmetic quotients is an indispensable features of the ideas developed in [32], namely the uniform spectral gap for congruence quotients. However, some of the other assumptions made in Theorem 5.1 may be relaxed. Indeed, in a subsequent work, Einsiedler, Margulis, Mohammadi, and Venkatesh [31] proved an adelic statement which lifts two of the restrictions imposed in Theorem 5.1: the fact that H is assumed fixed (the estimates in Theorem 5.1 depend on H) and splitting assumption on H at the archimedean place (H has no compact factors).

Let $\boldsymbol{G}$ be a connected, semisimple, algebraic $\mathbb{Q}$-group[3] and set $X = \boldsymbol{G}(\mathbb{A})/\boldsymbol{G}(\mathbb{Q})$ where $\mathbb{A}$ denotes the ring of adeles. Then X admits an action of the locally compact group $\boldsymbol{G}(\mathbb{A})$ preserving the probability measure m_X. Let $\mathbf{H}$ be a semisimple, simply connected, algebraic $\mathbb{Q}$-group, and let $g \in \boldsymbol{G}(\mathbb{A})$. Fix also an algebraic homomorphism $\iota : \mathbf{H} \to \mathbf{G}$ defined over $\mathbb{Q}$ with finite central kernel. For example, let $\boldsymbol{G} = \mathrm{SL}_d$ and $\boldsymbol{H} = \mathrm{Spin}(Q)$ for an integral quadratic form Q in d variables.

[3] The paper [31] allows for any number field, F, but unless X is compact, δ in Theorem 5.2 will depend on $\dim \boldsymbol{G}$ and $[F : \mathbb{Q}]$.

To this algebraic data and any $g \in \boldsymbol{G}(\mathbb{A})$, one associates a homogeneous set

$$Y := g\iota\big(\boldsymbol{H}(\mathbb{A})/\mathbf{H}(\mathbb{Q})\big) \subset X$$

and a homogeneous probability measure μ.

The following is a special case of the main theorem in [31].

Theorem 5.2 ([31]). *Assume further that* $\mathbf{G}$ *is simply connected. There exists some* $\delta > 0$, *depending only on* $\dim \boldsymbol{G}$, *so that the following holds. Let* Y *be a homogeneous set and assume that* $\iota(\boldsymbol{H}) \subset \boldsymbol{G}$ *is maximal. Then*

$$\left| \int_X f \, \mathrm{d}\mu - \int_X f \, \mathrm{d}m_X \right| \ll_{\boldsymbol{G}} \mathcal{S}(f) \operatorname{vol}(Y)^{-\delta} \quad \textit{for all } f \in C_c^\infty(X),$$

where $\mathcal{S}(f)$ *is a certain adelic Sobolev norm.*

The flexibility that Theorem 5.2 provides has interesting number theoretic applications. Indeed, the following generalization of Duke's theorem is proved in [31].

Let $\mathcal{Q}_d = \mathrm{PO}_d(\mathbb{R})\backslash \mathrm{PGL}_d(\mathbb{R})/\mathrm{PGL}_d(\mathbb{Z})$ be the space of positive definite quadratic forms on $\mathbb{R}^d$ up to the equivalence relation defined by scaling and equivalence over $\mathbb{Z}$. Equip $\mathcal{Q}_d$ with the push-forward of the normalized Haar measure on $\mathrm{PGL}_d(\mathbb{R})/\mathrm{PGL}_d(\mathbb{Z})$.

Let Q be a positive definite integral quadratic form on $\mathbb{Z}^d$, and let $\operatorname{genus}(Q)$ (resp. spin $\operatorname{genus}(Q)$) be its genus (resp. spin genus).

Theorem 5.3 ([31]). *Suppose* $\{Q_n\}$ *varies through any sequence of pairwise inequivalent, integral, positive definite quadratic forms. Then the genus (and also the spin genus) of* Q_n, *considered as a subset of* $\mathcal{Q}_d$, *equidistributes as* $n \to \infty$ *(with speed determined by a power of* $|\operatorname{genus}(Q_n)|$*).*

It is worth mentioning that when $d = 3, 4$, this theorem *even in its qualitative form* is new. When $d > 5$, the qualitative version of this theorem follows from an equidistribution theorem proved in [52], see also [43] for related analysis in the presence of a splitting condition at the archimedean place

Another application of Theorem 5.2 is an independent proof of property (τ) except for groups of type A_1. In particular, the paper [31] provides an alternative proof of the main result of Clozel in [19], albeit with weaker exponents, see [31, §4].

In addition to the ingredients involved in [32], the proof of Theorem 5.2 relies on Prasad's volume formula [86] and the work of Borel and Prasad [9]. These fundamental inputs are responsible for liberties supplied by Theorem 5.2.

The main problem which remains open in this direction is to prove an analogue of Theorem 5.2 which allows $\iota(\boldsymbol{H})$ to have an infinite centralizer; such a theorem would have quite interesting number theoretic applications, see [36]. Some progress has been made in this direction recently, the reader is invited to consult [1, 34, 35], for instance.

6. EFFECTIVE UNIPOTENT DYNAMICS

In view of Theorem 3.1, let us assume that G has noncompact semisimple subgroups, e.g., G is a noncompact semisimple linear Lie group. In light of the results discussed in Section 4, the analysis of the quantitative behavior of unipotent orbits in G/Γ is reduced to orbits of groups which are not horospherical. Not surprisingly, however, this task has proven quite challenging. In this section, we will discuss some recent progress made in this direction. The general theme of results in this section revolves around exploiting and effectivizing the *polynomial like behavior* of unipotent orbits.

6.1. Effective versions of the Oppenheim conjecture

The Oppenheim conjecture, proved by Margulis [74], states that if Q is a nondegenerate, indefinite quadratic form which is not a rational multiple of a form with integer coefficients, then for every $\varepsilon > 0$, there exists some $v \in \mathbb{Z}^3 \setminus \{0\}$ so that $|Q(v)| < \varepsilon$. Generalizations were also proved by Dani and Margulis prior to Ratner's theorems.

Later, Eskin, Margulis, and Mozes [37,38] proved quantitative (equidistribution) versions of the Oppenheim conjecture which relies on Ratner's equidistribution theorem [88–90], linearization techniques of Dani and Margulis [28], and a system of inequalities for a certain Margulis function—an ingenious idea introduced in [37] which has become an indispensable tool in homogeneous dynamics and beyond, see Section 7.2. Similar results for inhomogeneous forms have also been established [75,76].

Effective results in this context have also been actively pursued. Indeed, the analytic approach (using the Hardy–Littlewood circle method) which had been employed prior to Margulis' work is by its nature effective. However, this approach is generally only applicable if either the number of variables is large or the form has special features, see, e.g., [60]. More recently, Buterus, Götze, Hille, and Margulis [18] have proved effective version of the Oppenheim conjecture (as well as the equidistribution versions [37]), with polynomial error rates, provided that the number of variable is at least 5. Their proof combines analytic techniques with ideas from geometry of numbers in the form of inequalities which are reminiscent of [37]. Analytic methods were also used in [92] and [11] to obtain polynomial estimates for almost every form in certain families of forms in dimensions 3 and 4. The case of general forms in 3 and 4 variables, however, seem to be out of the reach of analytic methods.

Lindenstrauss and Margulis [69] proved an effective version of the Oppenheim conjecture for ternary form with polylog error rates.

Theorem 6.1 ([69]). *There exist absolute constants $A \geq 1$ and $\kappa > 0$ so that the following holds:*

Let Q be an indefinite, ternary quadratic form with $\det Q = 1$ *and let $\varepsilon > 0$. There exists $T_0(\varepsilon) > 0$ so that for any $T \geq T_0(\varepsilon)\|Q\|^A$ at least one of the following holds:*

(1) *For every* $\xi \in [-(\log T)^{\kappa}, (\log T)^{\kappa}]$, *there is a primitive integer vector* $v \in \mathbb{Z}^3$ *with* $0 < \|v\| < T^A$ *satisfying*

$$\left|Q(v) - \xi\right| \ll (\log T)^{-\kappa}.$$

(2) *There is an integral quadratic form* Q' *with* $|\det Q'| < T^{\varepsilon}$ *so that*

$$\left\|Q - \lambda Q'\right\| \ll \|Q\| T^{-1}$$

where $\lambda = |\det Q'|^{-1/3}$.

The implied multiplicative constants are absolute and $\|\cdot\|$ *denotes a norm on* $\mathrm{Mat}_3(\mathbb{R})$.

Note in particular that if Q is a reduced, indefinite, ternary quadratic form which is not proportional to an integral form but has algebraic coefficients, then part (1) in Theorem 6.1 holds true for Q, see [69, COR. 1.12].

The aforementioned dichotomy is again present in Theorem 6.1: unless there is an explicit obstruction (part (2) in Theorem 6.1), one obtains an effective density result.

The proof in [69] is rather involved and is based on effectivizing Margulis' original proof of the Oppenheim conjecture, as well as the subsequent works by Dani and Margulis. This approach, is based on the study of the action of $\mathrm{SO}(Q)$, the isometry group of Q, on $X = \mathrm{SL}_3(\mathbb{R})/\mathrm{SL}_3(\mathbb{Z})$, and relies on the notion of minimal sets from topological dynamics. Minimal sets are not suitable for quantitative arguments. Indeed, the paper [69] replaces this qualitative notion with a Diophantine condition in terms of the rate of escape to infinity under a certain one-parameter $\mathbb{R}$-diagonalizable subgroup. This novel ingredient plays a crucial role in obtaining an effective account—similar elements in more general contexts will be discussed in Section 6.2. It is worth mentioning that relying only on this input, one gets a rate that is $\ll \log(\log T)$. The stronger bound obtained in [69] is made possible thanks to a combinatorial lemma [69, §9] which is of independent interest.

6.2. Linearization of unipotent orbits

As it was mentioned before, Margulis and Dani developed a topological approach to settle certain special cases of Raghunathan's conjecture which relied on the notion of minimal sets. One of the first steps in effectivizing this topological argument would therefore be to replace minimal sets with an explicit Diophantine condition. This was established by Lindenstrauss, Margulis, Mohammadi, and Shah in [72] which may be thought of as an effective version of the *linearization technique* of Dani and Margulis [28].

The linearization technique has its roots in the techniques developed by Margulis [73] in his proof of the nondivergence of unipotent orbits. These nondivergence results are effective. Indeed, they were sharpened by Dani in [22,23] and have been given a very explicit and effective form by Kleinbock and Margulis in [65]. However, the author is not aware of an effective treatment of the main results in [28] prior to [72].

Let us recall the setting in [72]. Let $\boldsymbol{G}$ be a connected $\mathbb{Q}$-group, and put $G = \boldsymbol{G}(\mathbb{R})$. We assume that $\Gamma \subset G$ is an arithmetic lattice. More specifically, fix an embedding $\iota : \boldsymbol{G} \to$

SL_N defined over $\mathbb{Q}$ so that $\iota(\Gamma) \subset \mathrm{SL}_N(\mathbb{Z})$. Using ι, we identify $\boldsymbol{G}$ with $\iota(\boldsymbol{G}) \subset \mathrm{SL}_N$, and hence will always assume that $G \subset \mathrm{SL}_N(\mathbb{R})$.

Define the following family:

$$\mathcal{H} = \{\boldsymbol{H} \subset \boldsymbol{G} : \boldsymbol{H} \text{ is a connected } \mathbb{Q}\text{-subgroup and } \mathrm{R}(\boldsymbol{H}) = \mathrm{R}_u(\boldsymbol{H})\},$$

where $\mathrm{R}(\boldsymbol{H})$ (resp. $\mathrm{R}_u(\boldsymbol{H})$) denotes the solvable (resp. unipotent) radical of $\boldsymbol{H}$. Alternatively, $\boldsymbol{H} \in \mathcal{H}$ if and only if $\boldsymbol{H}$ is a connected $\mathbb{Q}$-subgroup which is generated by unipotent subgroups over the algebraic closure of $\mathbb{Q}$. By a theorem of Borel and Harish-Chandra, $\boldsymbol{H}(\mathbb{R}) \cap \Gamma$ is a lattice in $\boldsymbol{H}(\mathbb{R})$ for every $\boldsymbol{H} \in \mathcal{H}$. We always assume that $\boldsymbol{G} \in \mathcal{H}$.

Let $U \subset G$ be a (connected) unipotent subgroup of G, and put $X = G/\Gamma$. For every $\boldsymbol{H} \in \mathcal{H}$, put $H = \boldsymbol{H}(\mathbb{R})$. Define

$$N_G(U, H) := \{g \in G : Ug \subset gH\}.$$

Note that $N_G(U, H)$ is an $\mathbb{R}$-subvariety of G. Moreover, if $H \lhd G$ and $U \subset H$, then $N_G(U, H) = G$.

Put

$$\mathcal{S}(U) = \Big(\bigcup_{\substack{\boldsymbol{H}\in\mathcal{H} \\ \boldsymbol{H}\neq\boldsymbol{G}}} N_G(U, H)\Big)/\Gamma \quad \text{and} \quad \mathcal{G}(U) = X \setminus \mathcal{S}(U).$$

Points in $\mathcal{S}(U)$ are called *singular* with respect to U, and points in $\mathcal{G}(U)$ are called *generic* with respect to U. These are, a priori, different from the measure-theoretically generic points in the sense of Furstenberg for the action of U on X equipped with m_X (see, e.g., [50, P. 98] for a definition); however, any measure theoretically generic point is generic in this explicit sense as well. The aforementioned remarkable theorem of Ratner [90] states that for every $x \in \mathcal{G}(U)$, we have $\overline{Ux} = X$.

Dani and Margulis [28] proved that U orbits of points in $\mathcal{G}(U)$ *avoid* $\mathcal{S}(U)$. The paper [72] makes this principle quantitative with polynomial rates.

We need some more notation to state this quantitative result. Let $\mathfrak{g} = \mathrm{Lie}(G)$ and put $\mathfrak{g}(\mathbb{Z}) := \mathfrak{g} \cap \mathfrak{sl}_N(\mathbb{Z})$. Let $\|\cdot\|$ denote the max norm on $\mathfrak{sl}_N(\mathbb{R})$ with respect to the standard basis. This induces a family of norms on $\wedge\mathfrak{sl}_N(\mathbb{R})$, which we continue to denote by $\|\cdot\|$.

Let $\boldsymbol{H} \in \mathcal{H}$ be a nontrivial proper subgroup of $\boldsymbol{G}$, and put

$$\rho_H := \wedge^{\dim \boldsymbol{H}} \mathrm{Ad} \quad \text{and} \quad V_H := \wedge^{\dim \boldsymbol{H}} \mathfrak{g}.$$

The representation ρ_H is defined over $\mathbb{Q}$.

Let $\mathbf{v}_{\boldsymbol{H}}$ be a primitive integral vector in $\wedge^{\dim \boldsymbol{H}} \mathrm{Lie}(\boldsymbol{G})$ corresponding to the Lie algebra of $\boldsymbol{H}$, i.e., we fix a $\mathbb{Z}$-basis for $\mathrm{Lie}(H) \cap \mathfrak{sl}_N(\mathbb{Z})$, and let $\mathbf{v}_{\boldsymbol{H}}$ be the corresponding wedge product. The vector $\mathbf{v}_{\boldsymbol{H}}$ embeds diagonally in $\wedge^{\dim \boldsymbol{H}} \mathfrak{g}$; we denote this diagonally embedded vector by v_H. Define

$$\eta_H(g) := \rho_H(g) v_H \text{ for every } g \in G.$$

In order to simplify the exposition, let us assume that U is a one-parameter unipotent subgroup of G. Fix some $z \in \mathfrak{g}$ with $\|z\| = 1$ so that $U = \{u(t) = \exp(tz) : t \in \mathbb{R}\}$. With this notation, for an element $\boldsymbol{H} \in \mathcal{H}$, we have

$$N_G(U, H) = \{g \in G : z \wedge \eta_H(g) = 0\}.$$

As it was observed before, $N_G(U, H)$ is a variety; therefore, it could change drastically under small perturbations of U. However, *effective* notions must be stable under small perturbations. One of the innovations of [72] is the introduction of the following effective notion of generic points:

Definition 6.2. Let $\varepsilon : \mathbb{R}^+ \to (0, 1)$ be a monotone decreasing function, and let $t \in \mathbb{R}^+$. A point $g\Gamma$ is called (ε, t)*-Diophantine* for the action of $U = \{\exp(tz) : t \in \mathbb{R}\}$ if for all $\boldsymbol{H} \in \mathcal{H}$ with $\{e\} \neq \boldsymbol{H} \neq \boldsymbol{G}$,

$$\|z \wedge \eta_H(g)\| \geq \varepsilon(\|\eta_H(g)\|) \quad \text{if } \|\eta_H(g)\| < e^t. \tag{6.1}$$

A point is ε*-Diophantine* if it is (ε, t)-Diophantine for all $t > 0$.

Note that this is a condition on the *pair* $(U, g\Gamma)$. Unless $U \subset \boldsymbol{H}(\mathbb{R})$ for some (proper) $\boldsymbol{H} \lhd \boldsymbol{G}$, the set $\mathcal{G}(U)$ is nonempty; moreover, any $x \in \mathcal{G}(U)$ is ε-Diophantine for some ε as above. In most interesting examples, the singular set $\mathcal{S}(U)$ is a dense subset of X. Therefore, $\mathcal{G}(U)$ is usually a G_δ-set without any interior points. For any $t \in \mathbb{R}^+$, on the other hand, the set of (ε, t)-Diophantine points in Definition 6.2 is a nice closed set with interior points (indeed, it is the closure of its interior points).

As was discussed in Section 2, and we have seen in prior sections, finitary statements require a measure of complexity for obstructions. In [72], the following measure of arithmetic complexity for subgroups in $\mathcal{H}$ is used. Define

$$\operatorname{ht}(\boldsymbol{H}) := \|\mathbf{v}_{\boldsymbol{H}}\|. \tag{6.2}$$

That is, the height of a $\mathbb{Q}$-group $\boldsymbol{H}$ is given by the height of the corresponding point in the Grassmanian of $\operatorname{Lie}(\boldsymbol{G})$, see [8, §1.5]. It is worth mentioning that for subgroups $\boldsymbol{H} \in \mathcal{H}$, $\operatorname{ht}(\boldsymbol{H})$ is closely related to the volume of the periodic orbit $H\Gamma/\Gamma$ as it was defined in Section 2, see [32, §17], [31, APP. B], and [81, §6.2].

The space X is not necessarily compact; to deal with this issue, we fix an exhaustion of X by compact subsets as follows. For every $\eta > 0$, define

$$X_\eta = \Big\{g\Gamma \in X : \min_{0 \neq v \in \mathfrak{g}(\mathbb{Z})} \|\operatorname{Ad}(g)v\| \geq \eta\Big\}.$$

By (a generalization of) Mahler's compactness criterion, X_η is compact for every $\eta > 0$, see, e.g., [72, LEMMA 2.8]. Moreover, $\bigcup_{\eta > 0} X_\eta = G/\Gamma$.

For every $g \in \mathrm{SL}_N(\mathbb{R})$, in particular for every $g \in G$, we let

$$|g| = \max\{\|g\|, \|g^{-1}\|\},$$

where $\|\cdot\|$ denotes the max norm on $\mathrm{SL}_N(\mathbb{R})$ with respect to the standard basis.

The following is the main result in [72] in the case of real groups.

Theorem 6.3 ([72]). *There are constants $A, D > 1$ depending only on N, and $E > 1$ depending on N, G, and Γ, so that the following holds. Let $g \in G$, $t > 0$, $k \geq 1$, and $0 < \eta < 1/2$. Assume $\varepsilon : \mathbb{R}^+ \to (0, 1)$ satisfies for every $s > 0$ that*

$$\varepsilon(s) \leq \eta^A s^{-A}/E.$$

Then at least one of the following three possibilities holds:

(1)

$$\bigl|\{\xi \in [-1, 1] : u(e^k\xi)g\Gamma \notin X_\eta \text{ or } \\ u(e^k\xi)g\Gamma \text{ is not } (\varepsilon, t)\text{-Diophantine}\}\bigr| < E\eta^{1/D}.$$

(2) *There exist a nontrivial proper subgroup $\boldsymbol{H} \in \mathscr{H}$ with*

$$\operatorname{ht}(\boldsymbol{H}) \leq E\bigl(|g|^A + e^{At}\bigr)\eta^{-A}$$

so that the following hold for all $\xi \in [-1, 1]$:

$$\bigl\|\eta_H\bigl(u(e^k\xi)g\bigr)\bigr\| \leq E\bigl(|g|^A + e^{At}\bigr)\eta^{-A},$$
$$\bigl\|z \wedge \eta_H\bigl(u(e^k\xi)g\bigr)\bigr\| \leq Ee^{-k/D}\bigl(|g|^A + e^{At}\bigr)\eta^{-A},$$

where $U = \{\exp(tz) : t \in \mathbb{R}\}$.

(3) *There exist a nontrivial proper normal subgroup $\boldsymbol{H} \lhd \boldsymbol{G}$ with*

$$\operatorname{ht}(\boldsymbol{H}) \leq Ee^{At}\eta^{-A}$$

so that

$$\|z \wedge v_H\| \leq \varepsilon\bigl(\operatorname{ht}(\boldsymbol{H})^{1/A}\eta/E\bigr)^{1/A}.$$

Indeed, the paper [72] proves versions of this theorem for friendly measures [72, THM. 1.7] as well as S-arithmetic versions of this theorem [72, §3]. In particular, in view of [72, THM. 3.2], and by using the restriction of scalars from number fields to $\mathbb{Q}$, the results in [72] are applicable also in the case of groups defined over a general number field.

The arguments in [72] rely on polynomial behavior of unipotent orbits as did the arguments in [28]. However, in addition to being *polynomially* effective, the results also differ from [28] in the following sense. They provide a compact subset of $\mathscr{G}(U)$ which is *independent* of the base point and to which a unipotent orbit returns unless there is an algebraic obstruction, see [72, THMS. 1.1 AND 1.5]. Regarding nondivergence properties of unipotent orbits, such uniformity is well known and is due to Dani (see [23, 27]), but in this context it was not known prior to [72].

These features have been made possible using two main new ingredients. First is the use of an effective notion of a generic point, Definition 6.2. The second ingredient is the use of a certain subgroup in $\mathscr{H}$ which controls the *speed* of unipotent orbits in the representation space V_H, see [72, §4.7]. In addition, the arguments in [72] rely on effective versions of Nullstellensatz [77, THM. IV], as well as some local nonvanishing theorems related to Lojasiewicz inequality [15, 54, 55].

6.3. Effective density of unipotent orbits

The paper [72] is the first in a series of papers which provide a general effective orbit closure theorem for unipotent orbits on arithmetic quotients. The second paper, which is in preparation, crucially relies on the results of [72].

The rate we obtain (for density of a unipotent orbit) are an iteration of logarithms in the size of the flow parameter where the number of iterations depends on $\dim G$.

7. ARITHMETIC COMBINATORICS AND POLYNOMIAL BOUNDS

The discussion in Section 6 allude to the fact that effectivizing the *existing* arguments from unipotent dynamics often does not yield a polynomial rate. Indeed, beyond the notable settings we discussed in Sections 3–5, polynomial rates of density or equidistribution in this context are rather rare. In this section we discuss some recent progress made in this direction.

7.1. Random walks by toral automorphisms

Let $\Gamma \subset \mathrm{SL}_d(\mathbb{Z})$ be a Zariski-dense subgroup which acts strongly irreducibly on $\mathbb{R}^d$ (that is, no nontrivial subspace of $\mathbb{R}^d$ is invariant under a finite index subgroup of Γ). Let ν be a finitely supported probability measure on Γ whose support generates Γ.

Furstenberg [47] showed that

$$\lambda_1(\nu) = \lim_{n\to\infty} \frac{1}{n} \log \|g_1 \cdots g_n\| \quad \nu^{\mathbb{N}}\text{-a.s.}$$

is positive.

In a landmark paper [12], Bourgain, Furman, Lindenstrauss, and Mozes proved an equidistribution theorem for random walks on $\mathbb{T}^d$ corresponding to ν, with polynomial rates.

Theorem 7.1 ([12]). *For every $0 < \lambda < \lambda_1(\nu)$, there exists a constant $C = C(\nu, \lambda)$ so that if for a point $x \in \mathbb{T}^d$ the measure $\mu_n = \nu^{(n)} * \delta_x$ satisfies that for some $a \in \mathbb{Z}^d \setminus \{0\}$,*

$$\left|\widehat{\mu}_n(a)\right| > t > 0 \quad \textit{with } n > C \log\bigl(2\|a\|/t\bigr),$$

then x admits a rational approximation p/q where $p \in \mathbb{Z}^d$ and $q \in \mathbb{N}$ satisfying

$$\left\| x - \frac{p}{q} \right\| < e^{-\lambda n} \quad \textit{and} \quad |q| < (2\|a\|/t)^C.$$

Indeed, the main results in [12] allow for a more general class of subgroups Γ and measures ν. Let us also mention that the results in [12] have been further generalized in subsequent works, see, e.g., [57, 58].

The argument in [12] is quite involved and relies on several ingredients. Here we only highlight one the main steps in the proof, which concerns bootstrapping the information about one large Fourier coefficient to a (large scale) structure for the set of large Fourier coefficients. Suppose $|\widehat{\mu}_n(a)| > t$ for some large n and some nonzero a. Then using quantitative theory of random matrix products, one can show that for a suitable choice of $n_1 \le n$

the measure μ_{n_1} has Fourier coefficients which are $> t/2$ on a subset with a (small) positive dimension [12, PROP. 6.2]. The next task is to deduce from this a possibly smaller scale $n_2 < n_1$, so that μ_{n_2} has large (polynomial in t) Fourier coefficients on a set whose large scale dimension is d. This is carried out in two steps, the first, and arguably more difficult, step is to bootstrap the dimension to $d - \varepsilon$ for a small ε (depending on ν) [12, PROP. 6.3]. The paper [12] uses ideas from additive combinatorics, namely discretized ring conjecture [10] to establish this improvement. After this is obtain, one can use more or less classical estimates from Fourier analysis to improve the dimension from $d - \varepsilon$ to d, [12, PROPS. 6.5 AND 6.11].

The three stages in the above outline, namely the initial dimension, bootstrapping the dimension, and from high dimension to positive density, are reminiscent of the three stages present in the work of Bourgain and Gamburd on random walks on compact groups [13, 14]—these three stages will be revisited in the next section.

7.2. Quotients of $\mathrm{SL}_2(\mathbb{C})$ and $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$

We now turn to the question of density (or more ambitiously equidistribution) results in quotients of *semisimple* groups, with polynomial rates. For reasons we already discussed, this has proven quite a challenging task.

Lindenstrauss and Mohammadi [71] have very recently obtained first results in the literature which provide a polynomial rate for density of general orbits in a homogeneous space of a semisimple group, beyond the settings we discussed in Sections 4 and 5.

Let us fix some notation. Let

$$G = \mathrm{SL}_2(\mathbb{C}) \quad \text{or} \quad G = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}),$$

and let $\Gamma \subset G$ be a lattice. Put $X = G/\Gamma$.

Let dist be the right-invariant metric on G which is defined using the killing form. This metric induces a metric dist_X on X. The injectivity radius of a point $x \in X$ may be defined using this metric. For every $\eta > 0$, let

$$X_\eta = \{x \in X : \text{injectivity radius of } x \text{ is} \geq \eta\};$$

this is closely related to the definition in Section 6.2, see, e.g., [71, §3] and references there.

Let $H \subset G$ be one of the following:

$$\mathrm{SL}_2(\mathbb{R}) \subset \mathrm{SL}_2(\mathbb{C}) \quad \text{or} \quad \big\{(g,g) : g \in \mathrm{SL}_2(\mathbb{R})\big\} \subset \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R}).$$

Let $P \subset H$ be the group of upper triangular matrices in H.

As before, let $\|\cdot\|$ denote the maximum norm on $\mathrm{Mat}_2(\mathbb{C})$ or $\mathrm{Mat}_2(\mathbb{R}) \times \mathrm{Mat}_2(\mathbb{R})$ with respect to the standard basis. For every $R > 0$ and every subgroup $L \subset G$, let

$$B_R^L = \big\{g \in L : \|g - I\| \leq R\big\}.$$

The following is one of the main results in [71]:

Theorem 7.2 ([71]). *Assume that Γ is an arithmetic lattice. For every $0 < \delta < 1/2$, every $x_0 \in X$, and large enough T (depending explicitly on δ and the injectivity radius of x_0), at least one of the following holds:*

(1) *For every* $x \in X_{T^{-\kappa\delta}}$, *we have*

$$\operatorname{dist}_X\big(x, B^P_{T^A}.x_0\big) \le CT^{-\kappa\delta}.$$

(2) *There exists* $x' \in X$ *such that* Hx' *is periodic with* $\operatorname{vol}(Hx') \le T^\delta$, *and*

$$\operatorname{dist}_X(x', x_0) \le CT^{-1}.$$

The above A, κ, *and* C *are positive constants depending on* X.

The proof of Theorem 7.2 has a similar flavor to [49] by Gamburd, Jakobson, and Sarnak, as well as to the work of Bourgain and Gamburd [13, 14] and the aforementioned work of Bourgain, Furman, Lindenstrauss, and Mozes [12], see Theorem 7.1.

In particular, the three stages of the proof which were discussed in Section 7.1 are present here as well: in the first step, a Diophantine condition (in the form of a closing lemma) is used to show that unless part (2) in Theorem 7.2 holds, one can produce positive dimension at a certain scale (*initial dimension*). The arithmeticity of Γ is used in this step.

The second step, is the bootstrap phase in the following form: by passing to a larger scale and translating $B^P_{T^\delta}.x_0$ with a random element of controlled size, one can obtain a set with *large dimension*. This step is carried out using a Margulis function argument. As it was mentioned before, Margulis functions were introduced in the context of homogeneous dynamics in [37] by Eskin, Margulis, and Mozes, and have become an indispensable tool in homogeneous dynamics and beyond.

The third step is to deduce effective density from large dimension. Two main ingredients are present in this step: first is a projection theorem which is based on the works of Wolff and Schlag [94, 99] and is an adaptation of [61]. This is used to move the additional dimension supplied by the bootstrap phase to the direction of a horospherical subgroup of G. The second ingredient is an argument due to Venkatesh [98] and is based on the following quantitative decay of correlations for the ambient space X: There exists $\kappa_X > 0$ so that

$$\left| \int \varphi(gx)\psi(x)\,\mathrm{d}m_X - \int \varphi\,\mathrm{d}m_X \int \psi\,\mathrm{d}m_X \right| \ll_G \mathcal{S}(\varphi)\mathcal{S}(\psi)e^{-\kappa_X \operatorname{dist}(e,g)} \tag{7.1}$$

for all $\varphi, \psi \in C^\infty(X)$, where $\mathcal{S}$ is a certain Sobolev norm and dist is our fixed right G-invariant metric on G.

See, e.g., [64, §2.4] and references there for (7.1); we note that κ_X is an absolute constant if Γ is a congruence subgroup, see [17, 19, 51].

Periodic orbits

The techniques developed in [71] can also be used to prove an effective density theorem for periodic orbits of H.

Let us first recall the following nondivergence result: there exists some $\eta_X > 0$ so that for every periodic orbit Y, we have

$$\mu_Y(X_{\eta_X}) \ge 0.9, \tag{7.2}$$

where μ_Y denotes the H-invariant probability measure on Y, see, e.g., [71, LEMMA 3.6].

Theorem 7.3 ([71]). *Let $Y \subset X$ be a periodic H-orbit in X. Then for every $x \in X_{\mathrm{vol}(Y)^{-\kappa}}$, we have*

$$\mathrm{dist}_X(x, Y) \leq C\,\mathrm{vol}(Y)^{-\kappa},$$

where $\kappa \geq \kappa_X^2/L$ (for an absolute constant L) and C depends explicitly on κ_X, $\mathrm{vol}(X)$*, and the minimum of the injectivity radius of points in X_{η_X}. If Γ is congruence, κ is absolute.*

If Γ is an arithmetic lattice, Theorem 7.3 is a rather special case of the results we discussed in Section 5. Note, however, that Theorem 7.3 does *not* require Γ to be arithmetic—recall that arithmeticity of Γ was only used in the first step of the proof of Theorem 7.2. In particular, unlike [31, 32], Theorem 7.3 does not rely on property (τ).

We also draw the reader's attention to the use of Margulis functions in establishing isolation properties for periodic (or more generally intermediate) orbits in [41] and [83].

We end this exposition with the following application of Theorem 7.3.

Totally geodesic planes in hybrid manifolds

Gromov and Piatetski-Shapiro [56] constructed examples of nonarithmetic hyperbolic manifolds by gluing together pieces of noncommensurable arithmetic manifolds. Let Γ_1 and Γ_2 be two torsion free lattices in $\mathrm{Isom}(\mathbb{H}^3)$—recall that $\mathrm{Isom}(\mathbb{H}^3)$ is an index 2 subgroup of $\mathrm{O}(3,1)$ and that $\mathrm{SL}_2(\mathbb{C})$ is locally isomorphic to $\mathrm{O}(3,1)$. Let $M_i = \mathbb{H}^3/\Gamma_i$. Assume further that for $i = 1, 2$, there exists 3-dimensional submanifolds with boundary $N_i \subset M_i$ so that

- The Zariski closure of $\pi_1(N_i) \subset \Gamma_i$ contains $\mathrm{O}(3,1)^\circ$ where $\mathrm{O}(3,1)^\circ$ is the connected component of the identity in $\mathrm{O}(3,1)$.
- Every connected component of ∂N_i is a totally geodesic embedded surface in M_i which separates M_i.
- ∂N_1 and ∂N_2 are isometric.

Let M be the manifold obtained by gluing N_1 and N_2 using the isometry between ∂N_1 and ∂N_2. Then M carries a complete hyperbolic metric; thus, we consider $\pi_1(M)$ as a lattice in $\mathrm{O}(3,1)$. Let $\Gamma' = \pi_1(M) \cap \mathrm{O}(3,1)^\circ$, and let Γ denote the inverse image of Γ' in $G = \mathrm{SL}_2(\mathbb{C})$.

If Γ_1 and Γ_2 are arithmetic and noncommensurable, then M is nonarithmetic, i.e., Γ is a nonarithmetic lattice in G. A totally geodesic plane in M lifts to a periodic orbit of $H = \mathrm{SL}_2(\mathbb{R})$ in $X = G/\Gamma$.

Theorem 7.4. *Let M be a hyperbolic* 3*-manifold obtained by gluing the pieces N_1 and N_2 from noncommensurable arithmetic manifolds along $\Sigma = \partial N_1 = \partial N_2$ as described above. The number of totally geodesic planes in M is at most*

$$L\big(\mathrm{area}(\Sigma)\,\mathrm{vol}(X)\eta_X^{-1}\kappa_X^{-1}\big)^{L/\kappa_X^2},$$

where L is absolute and $X = G/\Gamma$ is as above.

In qualitative form, this finiteness theorem was proved by Fisher, Lafont, Miller, and Stover [44, THM. 1.4], see also [4, §12].

ACKNOWLEDGMENTS

The author is scheduled is to deliver a presentation based on this work in the Dynamics section and the Lie Theory and Generalizations section of the 2022 International Congress of Mathematicians.

The results presented above are outcomes of several years of collaborations with friends, colleagues, and mentors to whom we are ever so grateful. Special thanks are due to M. Einsiedler, A. Eskin, E. Lindenstrauss, G. Margulis, H. Oh, and A. Salehi Golsefidy, N. Shah, and A. Venkatesh.

We also would like to thank H. Oh for her comments on earlier versions of this article.

FUNDING

This work was partially supported by the NSF.

REFERENCES

[1] M. Aka, M. Einsiedler, H. Li, and A. Mohammadi, On effective equidistribution for quotients of $\mathrm{SL}(d,\mathbb{R})$. *Israel J. Math.* **236** (2020), no. 1, 365–391.

[2] L. Auslander, L. Green, and F. Hahn, *Flows on homogeneous spaces*. With the assistance of L. Markus and W. Massey, and an appendix by L. Greenberg. Ann. of Math. Stud. 53, Princeton University Press, Princeton, NJ, 1963.

[3] Y. Benoist and H. Oh, Effective equidistribution of S-integral points on symmetric varieties. *Ann. Inst. Fourier (Grenoble)* **62** (2012), no. 5, 1889–1942.

[4] Y. Benoist and H. Oh, Geodesic planes in geometrically finite acylindrical 3-manifolds. 2018, arXiv:1802.04423.

[5] Y. Benoist and J.-F. Quint, Mesures stationnaires et fermés invariants des espaces homogènes. *Ann. of Math. (2)* **174** (2011), no. 2, 1111–1162.

[6] Y. Benoist and J.-F. Quint, Stationary measures and invariant subsets of homogeneous spaces (II). *J. Amer. Math. Soc.* **26** (2013), no. 3, 659–734.

[7] Y. Benoist and J.-F. Quint, Stationary measures and invariant subsets of homogeneous spaces (III). *Ann. of Math. (2)* **178** (2013), no. 3, 1017–1059.

[8] E. Bombieri and W. Gubler, *Heights in diophantine geometry*. New Math. Monogr., Cambridge University Press, 2006.

[9] A. Borel and G. Prasad, Finiteness theorems for discrete subgroups of bounded covolume in semi-simple groups. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 119–171.

[10] J. Bourgain, The discretized sum-product and projection theorems. *J. Anal. Math.* **112** (2010), 193–236.

[11] J. Bourgain, A quantitative Oppenheim theorem for generic diagonal quadratic forms. *Israel J. Math.* **215** (2016), no. 1, 503–512.

[12] J. Bourgain, A. Furman, E. Lindenstrauss, and S. Mozes, Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus. *J. Amer. Math. Soc.* **24** (2011), no. 1, 231–280.

[13] J. Bourgain and A. Gamburd, On the spectral gap for finitely-generated subgroups of SU(2). *Invent. Math.* **171** (2008), no. 1, 83–121.

[14] J. Bourgain and A. Gamburd, Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(\mathbb{F}_p)$. *Ann. of Math. (2)* **167** (2008), no. 2, 625–642.

[15] W. D. Brownawell, Local Diophantine Nullstellen inequalities. *J. Amer. Math. Soc.* **1** (1988), no. 2, 311–322.

[16] M. Burger, Horocycle flow on geometrically finite surfaces. *Duke Math. J.* **61** (1990), no. 3, 779–803.

[17] M. Burger and P. Sarnak, Ramanujan duals. II. *Invent. Math.* **106** (1991), no. 1, 1–11.

[18] P. Buterus, F. Götze, T. Hille, and G. Margulis, Distribution of values of quadratic forms at integral points. 2019, arXiv:1004.5123.

[19] L. Clozel, Démonstration de la conjecture τ. *Invent. Math.* **151** (2003), no. 2, 297–328.

[20] S. G. Dani, Invariant measures of horospherical flows on noncompact homogeneous spaces. *Invent. Math.* **47** (1978), no. 2, 101–138.

[21] S. G. Dani, Invariant measures and minimal sets of horospherical flows. *Invent. Math.* **64** (1981), no. 2, 357–385.

[22] S. G. Dani, On orbits of unipotent flows on homogeneous spaces. *Ergodic Theory Dynam. Systems* **4** (1984), no. 1, 25–34.

[23] S. G. Dani, On orbits of unipotent flows on homogeneous spaces. II. *Ergodic Theory Dynam. Systems* **6** (1986), no. 2, 167–182.

[24] S. G. Dani, Orbits of horospherical flows. *Duke Math. J.* **53** (1986), no. 1, 177–188.

[25] S. G. Dani and G. A. Margulis, Values of quadratic forms at primitive integral points. *Invent. Math.* **98** (1989), no. 2, 405–424.

[26] S. G. Dani and G. A. Margulis, Orbit closures of generic unipotent flows on homogeneous spaces of $\mathrm{SL}(3,\mathbb{R})$. *Math. Ann.* **286** (1990), no. 1–3, 101–128.

[27] S. G. Dani and G. A. Margulis, Asymptotic behaviour of trajectories of unipotent flows on homogeneous spaces. *Proc. Indian Acad. Sci. Math. Sci.* **101** (1991), no. 1, 1–17.

[28] S. G. Dani and G. A. Margulis, Limit distributions of orbits of unipotent flows and values of quadratic forms. *I. M. Gelfand Seminar, Adv. Soviet Math.* **16** (1993), part. 1, 91–137. Amer. Math. Soc., Providence, RI.

[29] W. Duke, Z. Rudnick, and P. Sarnak, Density of integer points on affine homogeneous varieties. *Duke Math. J.* **71** (1993), no. 1, 143–179.

[30] M. Einsiedler, A. Katok, and E. Lindenstrauss, Invariant measures and the set of exceptions to Littlewood's conjecture. *Ann. of Math. (2)* **164** (2006), no. 2, 513–560.

[31] M. Einsiedler, G. Margulis, A. Mohammadi, and A. Venkatesh, Effective equidistribution and property (τ). *J. Amer. Math. Soc.* **33** (2020), no. 1, 223–289.

[32] M. Einsiedler, G. Margulis, and A. Venkatesh, Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces. *Invent. Math.* **177** (2009), no. 1, 137–212.

[33] M. Einsiedler and A. Mohammadi, Effective arguments in unipotent dynamics. In *Dynamics, geometry, number theory: the impact of Margulis on modern mathematics*, The University of Chicago Press, 2022.

[34] M. Einsiedler, R. Rühr, and P. Wirth, Distribution of shapes of orthogonal lattices. *Ergodic Theory Dynam. Systems* **39** (2019), no. 6, 1531–1607.

[35] M. Einsiedler and P. Wirth, Effective equidistribution of closed hyperbolic surfaces on congruence quotients of hyperbolic spaces. In *Dynamics, geometry, number theory: the impact of Margulis on modern mathematics*, The University of Chicago Press, 2022.

[36] J. S. Ellenberg and A. Venkatesh, Local–global principles for representations of quadratic forms. *Invent. Math.* **171** (2008), no. 2, 257–279.

[37] A. Eskin, G. Margulis, and S. Mozes, Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture. *Ann. of Math. (2)* **147** (1998), no. 1, 93–141.

[38] A. Eskin, G. Margulis, and S. Mozes, Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture. *Ann. of Math. (2)* **147** (1998), no. 1, 93–141.

[39] A. Eskin and C. McMullen, Mixing, counting, and equidistribution in lie groups. *Duke Math. J.* **71** (1993), no. 1, 181–209.

[40] A. Eskin and M. Mirzakhani, Invariant and stationary measures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Publ. Math. Inst. Hautes Études Sci.* **127** (2018), 95–324.

[41] A. Eskin, M. Mirzakhani, and A. Mohammadi, Isolation, equidistribution, and orbit closures for the $\mathrm{SL}(2,\mathbb{R})$ action on moduli space. *Ann. of Math. (2)* **182** (2015), no. 2, 673–721.

[42] A. Eskin, M. Mirzakhani, and A. Mohammadi, Effective counting of simple closed geodesics on hyperbolic surfaces. 2021, arXiv:1905.04435.

[43] A. Eskin and H. Oh, Representations of integers by an invariant polynomial and unipotent flows. *Duke Math. J.* **135** (2006), no. 3, 481–506.

[44] D. Fisher, J.-F. Lafont, N. Miller, and M. Stover, Finiteness of maximal geodesic submanifolds in hyperbolic hybrids. 2018, arXiv:1802.04619.

[45] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows. *Duke Math. J.* **119** (2003), no. 3, 465–526.

[46] L. Flaminio and G. Forni, Equidistribution of nilflows and applications to theta sums. *Ergodic Theory Dynam. Systems* **26** (2006), no. 2, 409–433.

[47] H. Furstenberg, Noncommuting random products. *Trans. Amer. Math. Soc.* **108** (1963), 377–428.

[48] H. Furstenberg, The unique ergodicity of the horocycle flow. In *Recent advances in topological dynamics (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund)*, pp. 95–115, Lecture Notes in Math. 318, 1973.

[49] A. Gamburd, D. Jakobson, and P. Sarnak, Spectra of elements in the group ring of SU(2). *J. Eur. Math. Soc. (JEMS)* **1** (1999), no. 1, 51–85.

[50] E. Glasner, *Ergodic theory via joinings*. Math. Surveys Monogr. 101, American Mathematical Society, Providence, RI, 2003.

[51] A. Gorodnik, F. Maucourant, and H. Oh, Manin's and Peyre's conjectures on rational points and adelic mixing. *Ann. Sci. Éc. Norm. Supér. (4)* **41** (2008), no. 3, 383–435.

[52] A. Gorodnik and H. Oh, Rational points on homogeneous varieties and equidistribution of adelic periods. *Geom. Funct. Anal.* **21** (2011), no. 2, 319–392.

[53] B. Green and T. Tao, The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)* **175** (2012), no. 2, 465–540.

[54] M. J. Greenberg, Rational points in Henselian discrete valuation rings. *Publ. Math. Inst. Hautes Études Sci.* **31** (1966), 59–64.

[55] M. J. Greenberg, Strictly local solutions of Diophantine equations. *Pacific J. Math.* **51** (1974), 143–153.

[56] M. Gromov and I. I. Piatetski-Shapiro, Non-arithmetic groups in Lobachevsky spaces. *Publ. Math. Inst. Hautes Études Sci.* **66** (1987), 93–103.

[57] W. He and N. de Saxcé, Linear random walks on the torus. 2020, arXiv:1910.13421.

[58] W. He, T. Lakrec, and E. Lindenstrauss, Affine random walks on the torus. 2020, arXiv:2003.03743.

[59] G. A. Hedlund, Fuchsian groups and transitive horocycles. *Duke Math. J.* **2** (1936), no. 3, 530–542.

[60] H. Iwaniec, On indefinite quadratic forms in four variables. *Acta Arith.* **33** (1977), no. 3, 209–229.

[61] A. Käenmäki, T. Orponen, and L. Venieri, A Marstrand-type restricted projection theorem in $\mathbb{R}^3$. 2017, arXiv:1708.04859.

[62] A. Katz, Quantitative disjointness of nilflows from horospherical flows. 2019, arXiv:1910.04675.

[63] D. Kleinbock and G. Margulis, On effective equidistribution of expanding translates of certain orbits in the space of lattices. In *Number theory, analysis and geometry*, edited by D. Goldfeld, J. Jorgenson, P. Jones, D. Ramakrishnan, K. Ribet, and J. Tate, pp. 385–396, Springer, Boston, MA, 2012.

[64] D. Y. Kleinbock and G. A. Margulis, Bounded orbits of nonquasiunipotent flows on homogeneous spaces. In *Sinaĭ's Moscow Seminar on Dynamical Systems*, pp. 141–172, Amer. Math. Soc. Transl. Ser. 2 171, Amer. Math. Soc., Providence, RI, 1996.

[65] D. Y. Kleinbock and G. A. Margulis, Flows on homogeneous spaces and Diophantine approximation on manifolds. *Ann. of Math. (2)* **148** (1998), no. 1, 339–360.

[66] M. Lee and H. Oh, Orbit closures of unipotent flows for hyperbolic manifolds with Fuchsian ends. 2020, arXiv:1902.06621.

[67] A. Leibman, Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, 201–213.

[68] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)* **163** (2006), no. 1, 165–219.

[69] E. Lindenstrauss and G. Margulis, Effective estimates on indefinite ternary forms. *Israel J. Math.* **203** (2014), no. 1, 445–499.

[70] E. Lindenstrauss and M. Mirzakhani, Ergodic theory of the space of measured Laminations. *Int. Math. Res. Not.* **2008** (2008).

[71] E. Lindenstrauss and A. Mohammadi, Polynomial effective density in quotients of $\mathbb{H}^3$ and $\mathbb{H}^2 \times \mathbb{H}^2$. 2021, arXiv:2112.14562

[72] E. Lindenstrauss, A. Mohammadi, G. Margulis, and N. Shah, Quantitative behavior of unipotent flows and an effective avoidance principle. 2019, arXiv:1904.00290.

[73] G. A. Margulis, The action of unipotent groups in a lattice space. *Mat. Sb.* **86** (1971), no. 128, 552–556.

[74] G. A. Margulis, Indefinite quadratic forms and unipotent flows on homogeneous spaces. *Dynamical systems and ergodic theory (Warsaw, 1986). Banach Center Publ.* **23** (1989), 399–409.

[75] G. Margulis and A. Mohammadi, Quantitative version of the Oppenheim conjecture for inhomogeneous quadratic forms. *Duke Math. J.* **158** (2011), no. 1, 121–160.

[76] J. Marklof, Pair correlation densities of inhomogeneous quadratic forms. *Ann. of Math. (2)* **158** (2003), no. 2, 419–471.

[77] D. W. Masser and G. Wüstholz, Fields of large transcendence degree generated by values of elliptic functions. *Invent. Math.* **72** (1983), no. 3, 407–464.

[78] T. McAdam, Almost-prime times in horospherical flows on the space of lattices. *J. Mod. Dyn.* **15** (2019), 277–327.

[79] C. T. McMullen, A. Mohammadi, and H. Oh, Geodesic planes in hyperbolic 3-manifolds. *Invent. Math.* **209** (2017), no. 2, 425–461.

[80] C. T. McMullen, A. Mohammadi, and H. Oh, Geodesic planes in the convex core of an acylindrical 3-manifold. 2021, arXiv:1802.03853.

[81] A. Mohammadi, A. S. Golsefidy, and F. Thilmany, Diameter of homogeneous spaces: an effective account. 2018, arXiv:1811.06253.

[82] A. Mohammadi and H. Oh, Matrix coefficients, counting and primes for orbits of geometrically finite groups. *J. Eur. Math. Soc. (JEMS)* **17** (2015), no. 4, 837–897.

[83] A. Mohammadi and H. Oh, Isolations of geodesic planes in the frame bundle of a hyperbolic 3-manifold. 2020, arXiv:2002.06579.

[84] S. Mozes and N. Shah, On the space of ergodic invariant measures of unipotent flows. *Ergodic Theory Dynam. Systems* **15** (1995), no. 1, 149–159.

[85] W. Parry, Dynamical systems on nilmanifolds. *Bull. Lond. Math. Soc.* **2** (1970), 37–40.

[86] G. Prasad, Volumes of S-arithmetic quotients of semi-simple groups. *Publ. Math. Inst. Hautes Études Sci.* **69** (1989), 91–117.

[87] M. S. Raghunathan, *Discrete subgroups of Lie groups*. Springer, New York–Heidelberg, 1972.

[88] M. Ratner, On measure rigidity of unipotent subgroups of semisimple groups. *Acta Math.* **165** (1990), no. 3–4, 229–309.

[89] M. Ratner, On Raghunathan's measure conjecture. *Ann. of Math. (2)* **134** (1991), no. 3, 545–607.

[90] M. Ratner, Raghunathan's topological conjecture and distributions of unipotent flows. *Duke Math. J.* **63** (1991), no. 1, 235–280.

[91] P. Sarnak, Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series. *Comm. Pure Appl. Math.* **34** (1981), no. 6, 719–739.

[92] P. Sarnak, Values at integers of binary quadratic forms. In *Harmonic analysis and number theory (Montreal, PQ, 1996)*, pp. 181–203, Conf. Proc., Can. Math. Soc. 21, Amer. Math. Soc., Providence, RI, 1997.

[93] P. Sarnak and A. Ubis, The horocycle flow at prime times. *J. Math. Pures Appl. (9)* **103** (2015), no. 2, 575–618.

[94] W. Schlag, On continuum incidence problems related to harmonic analysis. *J. Funct. Anal.* **201** (2003), 480–521.

[95] A. Strömbergsson, On the uniform equidistribution of long closed horocycles. *Duke Math. J.* **123** (2004), no. 3, 507–547.

[96] A. Strömbergsson, An effective Ratner equidistribution result for $\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$. *Duke Math. J.* **164** (2015), no. 5, 843–902.

[97] W. A. Veech, Minimality of horospherical flows. *Israel J. Math.* **21** (1975), no. 2–3, 233–239.

[98] A. Venkatesh, Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math. (2)* **172** (2010), no. 2, 989–1094.

[99] T. Wolff, Local smoothing type estimates on L^p for large p. *Geom. Funct. Anal.* **10** (2000), no. 5, 1237–1288.

[100] D. Zagier, Eisenstein series and the Riemann zeta function. In *Automorphic forms, representation theory and arithmetic (Bombay, 1979)*, pp. 275–301, Tata Inst. Fund. Res. Stud. Math. 10, Tata Inst. Fundamental Res., Bombay, 1981.

AMIR MOHAMMADI

Department of Mathematics, The University of California, San Diego, CA 92093, USA, ammohammadi@ucsd.edu

STABILITY AND RECURSIVE SOLUTIONS IN HAMILTONIAN PDES

MICHELA PROCESI

ABSTRACT

In this survey we shall consider Hamiltonian dispersive partial differential equations on compact manifolds and discuss the existence, close to an elliptic fixed point, of special recursive solutions, which are superpositions of oscillating motions, together with their stability/instability properties. One can envision such equations as chains of harmonic oscillators coupled with a small nonlinearity, thus one expects a complicated interplay between chaotic and recursive phenomena due to resonances and small divisors, which are studied with methods from KAM theory.

We shall concentrate mainly on the stability properties of the fixed point, as well as the existence and stability of quasiperiodic and almost periodic solutions. After giving an overview on the literature, we shall present some promising recent results and discuss possible extensions and open problems.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37K55; Secondary 37K45, 35B35, 35Q41

KEYWORDS

Birkhoff normal form, almost-periodic solutions, nonlinear Schrödinger equation, almost global existence

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3552–3574
DOI 10.4171/ICM2022/55

1. INTRODUCTION

A huge variety of physical systems is modeled by Hamiltonian dispersive partial differential equations (PDEs), such as the nonlinear Schrödinger (NLS) and wave (NLW) equations, Euler and water wave equations, KdV, etc. A good point of view, which has produced many advancements in the last 30 years, is to take a dynamical systems perspective and understand qualitative behavior by studying special invariant objects, such as finite- and infinite-dimensional tori, chaotic and diffusion orbits, etc. This perspective is particularly uselful for PDEs on compact domains, where one expects a complicated interplay between chaotic and recursive phenomena. For concreteness, we shall concentrate on NLS equations on a compact Riemannian manifold $(\mathcal{M}, g)$ without boundary, namely

$$\mathrm{i}u_t - \Delta_g u + V(x)u + f\big(|u|^2\big)u = 0 \tag{NLS}$$

where $f(y)$ is analytic in a neighborhood of zero, with $f(0) = 0$, and V is an appropriately regular, real potential $V : \mathcal{M} \to \mathbb{R}$, so that $u = 0$ is an elliptic fixed point.

Of course, (NLS) is still a simplified model, since more physical examples have derivatives in the nonlinearity; this is true for most PDEs modeling hydrodynamics and, indeed, there are results in this more general setting, mostly confined to spheres or flat tori $\mathbb{T}^n := \mathbb{R}^n \setminus \mathbb{Z}^n$. In fact, in all the results we discuss, we shall impose some simplifying condition, such as choosing simple manifolds (for instance, tori or spheres, or more generally, simple compact Lie groups), and/or simplify the model, for instance, by using a convolution (instead of multiplicative) potential.

In studying the dynamics of (NLS) close to zero, one expects a complicated interplay between chaotic and recursive phenomena, with the qualitative behavior of solutions depending in a subtle way on the geometry of $\mathcal{M}$ and on V. For instance, for stability results one typically needs to use V as a "source of parameters," say by modulating V so that the eigenvalues of the elliptic operator $-\Delta_g + V$ satisfy some nonresonance conditions (such as lower bounds on the integer combinations of eigenvalues). At the same time, to deal with the nonlinearity, one needs rather precise information on products of eigenfunctions, particularly on the coefficients that give the representation in the eigenfunction basis of the product of two eigenfunctions. As a drawback, it is usually very difficult to get results for a fixed value of V, for instance, $V = 0$.

Recalling that $-\Delta_g + V(x)$ is self-adjoint and with pure point spectrum, let $(\psi_j)_{j\in I}$ be its eigenfunctions (I is some countable index set) and ω_j the corresponding eigenvalues. Then we pass to the "Fourier side," $u = \sum_{j\in I} u_j \psi_j$. Writing NLS in terms of this basis, we have an infinite chain of harmonic oscillators coupled by a nonlinear term. It is easily verified that the associated equations are Hamiltonian with respect to the standard symplectic form $\Omega := \mathrm{i}\sum_{j\in I} du_j \wedge d\bar{u}_j$, with Hamiltonian

$$H_{\mathrm{NLS}} := \sum_{j\in I} \omega_j |u_j|^2 + P(u), \quad u = (u_j)_{j\in I}, \tag{1}$$

where P is a suitable nonlinearity with a zero of order at least four.

If we ignore the nonlinearity, the dynamics is very simple. All the linear actions $|u_j|^2$ are constants of motion and the dynamics is $u_j(t) = u_j(0)e^{\mathrm{i}\omega_j t}$, hence all solutions are superpositions of oscillations. For typical V's, the linear frequencies ω_j are rationally independent and the solutions live on tori, with dimension depending on the number of nonzero actions.

Now if we take into account the nonlinearity, the u_j's interact, exchanging energy; we want to study how, close to the origin, the dynamics differs from the linear one and over which time scales. To make this quantitative, we fix a phase space $\mathtt{h} \subset L^2$ of sufficiently regular functions $\mathcal{M} \to \mathbb{C}$ (typically, a spectrally defined Sobolev space prescribing sufficiently fast decay of the linear actions $|u_j|^2$ as $j \to \infty$). If $\mathtt{h}$ is sufficiently regular then NLS is at least locally well posed and one expects the dynamics to be close to the linear one at least close to zero and for finitely long time.

If we look at a finite-dimensional truncation of (1), then the classical Kolmogorov–Arnold–Moser (KAM) theory gives a rather clear picture: under some (generic) nondegeneracy assumptions,[1] close to the origin most of the phase space is foliated by Lagrangian invariant tori (with dimension half of that of the phase space). In particular, the system is not ergodic and most initial data give rise to quasiperiodic solutions that densely fill some invariant torus and are, therefore, perpetually stable.[2] Possible chaotic behavior is restricted outside a set of asymptotically full measure at the origin. Moreover, the origin and the maximal tori are stable, with nearby trajectories staying close for exponentially long times. All the finite-dimensional results strongly depend on the dimension, and one cannot naïvely perform finite-dimensional truncations in (1) and then take limits. In fact, in the infinite-dimensional setting, the general picture is so far rather obscure and the main questions still remain unanswered. All linear solutions are perpetually stable and typical ones lie on maximal infinite-dimensional invariant tori. What is their fate under perturbation? Is it still true that typical initial data produce perpetually stable solutions? What are the stability times?

Of course, even the concept of a "typical solution" depends on our choice of the measure on the phase space. Moreover, even at the linear level, simple topological issues such as whether the maximal tori give a foliation, or whether the dynamics on the tori is dense, depend strongly on the choice of the phase space and its topology.

In this survey we discuss some partial answers to these fundamental questions. We concentrate on three issues:

1. *Stability of zero.* A good way to capture the transfer of energy between Fourier modes is to study the time evolution of the norm $|\cdot|_{\mathtt{h}}$; indeed, if $\mathtt{h}$ is sufficiently regular, a growth in the norm represents transfers between low and high modes. With this in mind, we take any initial datum $u_0 \in B_\delta(\mathtt{h})$ and give estimates on the time $T(\delta)$ such that the flow $u(t,\cdot)$ of the NLS equation is well defined and belongs to $u_0 \in B_{2\delta}(\mathtt{h})$. A rough estimate[3] gives $T(\delta) \geq \delta^{-2}$; to get better lower bounds, a good strategy is to find a change of variables on $B_\delta(\mathtt{h})$ which conjugates the Hamiltonian to $N + R$, where N is *the normal form* preserving the norm $|\cdot|_{\mathtt{h}}$ while R is *the remainder* which is small and affects the dynamics

[1] On ω and/or on the nonlinearity.

[2] Namely the linear actions have a small variation for all times.

[3] Coming from well posedness and the fact that the nonlinearity is at least cubic.

over very long times. In constructing such a change of variables, one encounters small divisors, i.e., in their analytic expression one has integer combinations of linear frequencies in the denominator, so a major point will be to impose sufficiently strong irrationality conditions and ensure some lower bounds. This is done by appropriately modulating the external parameters.

2. *Small quasiperiodic solutions.* We look for special global solutions living on finite-dimensional invariant tori. More precisely, we look for a sufficiently regular map $U : \mathbb{T}^n \to \mathtt{h}$ and a frequency $\omega \in \mathbb{R}^n$ such that $U(\mathbb{T}^n)$ is invariant under the NLS dynamics, which, when restricted to the torus, is the linear translation by ωt. We work close to zero in order to take advantage of the fact that all the solutions of the linearized equation which are supported on finitely many Fourier modes are indeed quasiperiodic. Then starting from such approximately invariant tori, one wishes to prove that nearby there exist truly invariant ones. This is done by iterative approximations using a quadratic scheme. Again one needs to control small divisors by modulating the external parameters (typically, one only needs as many parameters as the dimension of the torus but there are a number of results where one only needs one parameter, or even none). Note that these solutions are very special, even in the case of an integrable PDE they are not typical.

3. *Small almost-periodic solutions.* Starting from quasiperiodic solutions with an arbitrary number n of frequencies, it is very natural to wonder whether one can pass to the limit as $n \to \infty$ thus obtaining an almost-periodic solution. Since almost-periodic solutions are "typical" for integrable systems, a main question is how rare such solutions are in a nonintegrable setting. Unfortunately, up to now all results are for PDEs on the circle and show the existence of few and very regular solutions.

Having proved the existence of these special global solutions, an interesting point is to study their stability properties, thus giving an insight on the nearby dynamics for finite but long times. A strategy is to perform changes of variables to put the system in normal form in a neighborhood of the solution. A dual point of view is to look for unstable/chaotic orbits driven by the presence of resonant terms in the nonlinearity.

We shall concentrate on (NLS); however, all the questions described above are tied mainly to the Hamiltonian formulation (1), thus they can be reformulated for PDEs on unbounded domains, when $-\Delta_g + V(x)$ has a pure point spectrum $\omega_j \to \infty$. An interesting (and widely studied) example is the harmonic oscillator, namely $\mathcal{M} = \mathbb{R}^n$ and $V(x) = |x|^2$.

In the next sections we shall give a brief (and necessarily incomplete) survey on the three questions described above, together with some open problems.

2. LONG TIME STABILITY

The problem of *long-time* stability for infinite-dimensional dynamical systems has been studied by many authors, starting from [13] for infinite chains with a finite range coupling. In the PDE context, after the first results in [5,6,28], a breakthrough was in the papers [7,9] where the authors proved polynomial bounds on the stability times for a rather wide class of *tame-modulus* PDEs depending on parameters. Their result applies to the NLS equations on tori, where they show that for any $\mathtt{N} \gg 1$ there exist many values of the parameters for which any δ-small initial datum in H^p (with $p = p(\mathtt{N})$ tending to infinity as $\mathtt{N} \to \infty$) stays 2δ small for times $T \geq C(\mathtt{N}, p)\delta^{-\mathtt{N}}$.

An interesting question is how such results perform in applications to PDEs with derivatives in the nonlinearity; a series of results in this direction were proved for the Klein–Gordon equation on Zoll manifolds in [8,41–43]. The method developed in these papers, based on a good control of the small divisors, together with ideas from paradifferential calculus, does not apply to the case where the PDE has a superlinear dispersion relation.

Recently there has been a lot of progress regarding[4] quasilinear and fully nonlinear PDEs on $\mathcal{M} = \mathbb{S}^1$, we mention [20,21] on the water waves and [52] for quasilinear NLS. Regarding higher-dimensional quasilinear PDEs, we mention [50,53] for Klein–Gordon and Schrödinger equations on higher-dimensional tori. While most results deal with parameter families of PDEs (and hold for most values of the parameters), we mention [14] on perturbations of the integrable 1D NLS.

If one wants to go beyond polynomial bounds, up to now the literature is restricted to PDEs on tori, with initial data which are at least C^∞. In [47] the authors considered the case of analytic initial data and proved subexponential bounds of the form $T \geq e^{c\ln(\frac{1}{\delta})^{1+\beta}}$ for classes of NLS equations in $\mathbb{T}^d$. Such bounds have been discussed also in [25,36] in Gevrey class for the 1D NLS.

In order to describe the results more in detail, let us restrict to the simplest possible case of a translation invariant NLS with a convolution potential when $\mathcal{M} = \mathbb{S}^1$ so the Fourier decomposition is $u(x) = \sum_{j\in\mathbb{Z}} u_j e^{\mathrm{i}jx}$. We consider

$$\mathrm{i}u_t - u_{xx} + V \star u + f\big(|u|^2\big)u = 0, \quad V \star u := \sum_{j\in\mathbb{Z}} u_j V_j e^{\mathrm{i}jx}, \quad (V_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z}, \mathbb{R}), \tag{2}$$

so the NLS Hamiltonian (1) has frequencies $\omega_j = \omega_j(V) = j^2 + V_j$ and $P := \int_{\mathbb{T}} F(|u(x)|^2)dx$ with $F(y) := \int_0^y f(s)ds$. An important feature is that the equation now has two constants of motion

$$L = \sum_{j\in\mathbb{Z}} |u_j|^2, \quad M = \sum_{j\in\mathbb{Z}} j|u_j|^2,$$

corresponding respectively to gauge invariance $u(x) \to e^{\mathrm{i}\tau}u(x)$ and translation invariance $u(x) \to u(x+\tau)$.

As we have explained before, the stability results depend on imposing sufficiently good nonresonance conditions, otherwise one can produce counterexamples where the actions have a *fast drift*; see, for instance, [57]. For this purpose, we shall assume a very strong condition, proposed by Bourgain in [32], which is tailored to 1D PDEs and gives good estimates for many choices of the phase space. More precisely, recalling that $\omega_j(V) = j^2 + V_j$,

[4] We say that a PDE is semilinear if the highest order derivatives occur in the linear part, quasilinear if the same order derivatives appear in the linear and nonlinear parts but with degree one, otherwise fully nonlinear if the highest derivative has degree higher than one.

for $\gamma > 0$ we define the set of Diophantine frequencies

$$\mathtt{D}_\gamma := \left\{ V \in \left[-\frac{1}{2}, \frac{1}{2}\right]^{\mathbb{Z}} : \left|\omega(V) \cdot \ell\right| > \gamma \prod_{j \in \mathbb{Z}} \frac{1}{1 + |\ell_j|^2 \langle j \rangle^2}, \ \forall \ell \in \mathbb{Z}^{\mathbb{Z}} : 0 < |\ell| < \infty \right\}. \tag{3}$$

It results (see [25,31]) that $\mathtt{D}_\gamma$ is large with respect to the natural probability product measure on $[-\frac{1}{2}, \frac{1}{2}]^{\mathbb{Z}}$. From now on we shall assume that $V \in \mathtt{D}_\gamma$.

Theorem 1 (Sobolev stability, [9]). *For any p large enough and any initial datum $u(0) = u_0$ satisfying*

$$|u_0|_{H^p} := |u_0|_{L^2} + \left|\partial_x^p u_0\right|_{L^2} \leq \delta \leq \delta_0 \sim p^{-3p}, \tag{4}$$

the solution $u(t)$ of (NLS)$_V$ *with initial datum $u(0) = u_0$ exists and satisfies*

$$\left|u(t)\right|_{H_p} \leq 4\delta \quad \textit{for all times } |t| \leq T \sim p^{-5p} \delta^{-\frac{2(p-1)}{\tau_{\mathrm{S}}}}. \tag{5}$$

An interesting feature of this result is that the stability time is related to the regularity. The estimates given here are those for (2) with the Diophantine conditions (3); however, a similar phenomenon appears in all the known literature.

Let us now increase the regularity and consider Gevrey initial data, let us fix $0 < \theta < 1$, and define the function space

$$\mathtt{H}_{s,a} := \left\{ u(x) = \sum_{j \in \mathbb{Z}} u_j e^{\mathrm{i}jx} \in L^2 : |u|_{s,a}^2 := \sum_{j \in \mathbb{Z}} |u_j|^2 \langle j \rangle^2 e^{2a|j| + 2s\langle j \rangle^\theta} < \infty \right\}, \tag{6}$$

with the assumption $a \geq 0$, $s > 0$. We remark that if $a > 0$, this is a space of analytic functions, while if $a = 0$ the functions have Gevrey regularity.

Theorem 2 (Gevrey stability, [25,36]). *Fix any $a \geq 0$, $s > 0$. For any u_0 such that*

$$|u_0|_{s,a} \leq \delta \leq \delta_0 \ll 1,$$

the solution $u(t)$ of (2) with initial datum $u(0) = u_0$ exists and satisfies

$$\left|u(t)\right|_{s,a} \leq 2\delta \quad \textit{for all times } |t| \leq \frac{\mathtt{T}_0}{\delta^2} e^{(\ln \frac{\delta_0}{\delta})^{1+\theta/4}}.$$

As can be expected, as $s \to 0$ one has $\delta_0, T_0^{-1} \to 0$, on the other hand, modulating the parameter a does not give significantly improved bounds. This leads to two very natural questions: Can one get better bounds for analytic initial data? Conversely, can we lower the regularity still obtaining superpolynomial stability times?

A reasonable strategy for tackling the second question is proposed in [25] where we discussed a BNF approach for (2) on abstract weighted functions spaces. Given a positive sequence $\mathtt{w} = (\mathtt{w}_j)_{j \in \mathbb{Z}}$, with $1 \leq \mathtt{w}_j \nearrow \infty$, let us consider the Hilbert space

$$\ell_{\mathtt{w}}^2 := \left\{ u := (u_j)_{j \in \mathbb{Z}} \in \ell^2(\mathbb{C}) : |u|_{\mathtt{w}}^2 := \sum_{j \in \mathbb{Z}} \mathtt{w}_j^2 |u_j|^2 < \infty \right\}. \tag{7}$$

By the Fourier transform, such spaces identify corresponding function spaces of periodic functions. For instance, if $\mathtt{w}_j = \langle j \rangle^p$, then[5] $\mathcal{F}(\ell_{\mathtt{w}}^2)$ identifies with the Sobolev space H_p. Similarly, if $\mathtt{w}_j := \langle j \rangle e^{a|j| + s\langle j \rangle^\theta}$, we are in the Gevrey/analytic case of $H_{s,a}$.

[5] The Fourier transform $\mathcal{F}$ identifies sequences with functions $\mathcal{F}((u_j)_{j \in \mathbb{Z}}) = \sum_{j \in \mathbb{Z}} u_j e^{\mathrm{i}jx}$.

In this context we gave some computationally heavy, but very explicit conditions on $\mathtt{w}$ which ensure that a BNF theorem can be applied and allows computing the stability times. We concentrated on the two cases above, but if one runs the same computations with $\mathtt{w}_j := \langle j \rangle e^{p \log(1+\langle j \rangle)^2}$ then one gets times of order $\delta^{-\ln\ln(\delta^{-1})}$.

It would be interesting to understand whether such bounds are optimal. A natural strategy would be to construct solutions to (NLS) whose Sobolev norm increases in time. There has been a lot of interest in this question and in particular on whether one can construct solutions whose norm becomes arbitrarily large, or even diverges as $t \to \infty$. A mechanism for ensuring finite, but arbitrarily large growth was constructed in [33] for the cubic parameterless NLS on $\mathbb{T}^2$ (see also [61] for the noncubic case and [59] for a case with convolution potential). The idea is to look for solutions which are approximately supported on Fourier modes $\mathcal{S} \subset \mathbb{Z}^2$ which are resonant (i.e., some linear combinations of ω_j's with $j \in \mathcal{S}$ are zero or very small). Then the interactions between Fourier modes due to the nonlinearity become dominant and the Sobolev norm varies. The very beautiful approach of [33] seems strongly tied to the NLS equation, if one only wants to find solutions which only, say, double their Sobolev norm then there are more robust mechanisms. One idea, see [63], is to prove the existence of secondary tori which transfer energy between two sets of Fourier modes periodically in time. Another very interesting approach, see [58], is to construct chaotic orbits generalizing "Arnold diffusion" to infinite dimension.

2.1. Questions and open problems

Q1. Can one obtain stability times on a fixed Sobolev space H_p, with $T(\delta)$ growing faster than polynomially as $\delta \to 0$?

Q2. Can one prove stability for most V for the NLS with a multiplicative potential?

This was considered in [9]. However, in order to obtain a stability time of order $\delta^{-\mathtt{N}}$, the authors had to restrict the potential to a small ball (in an appropriate norm) with radius going to zero as $\mathtt{N} \to \infty$. The delicate point here is what kind of irrationality conditions can be imposed on the linear frequencies ω_j, which in this case are the periodic spectrum of the Sturm–Liouville operator $-\partial_{xx} + V(x)$ where V is an analytic function.

Q3. Can one extend the stability results to general manifolds in higher dimension?

Q4. Can one extend the subexponential Gevrey bounds to quasilinear PDEs?

Q5. What kind of bounds can be given on instability times?

2.2. An idea of the strategies

To conclude this section, let us briefly illustrate the Birkhoff normal form procedure in its simplest form applied to (2). For this purpose, we consider an analytic translation-invariant Hamiltonian written as an absolutely convergent power series

$$H(u) = \sum_{(\alpha,\beta)\in\mathcal{M}} H_{\alpha,\beta} u^\alpha \bar{u}^\beta, \quad u^\alpha := \prod_{j\in\mathbb{Z}} u_j^{\alpha_j}, \tag{8}$$

where $\mathcal{M} := \{(\alpha,\beta) \in \mathbb{N}^{\mathbb{Z}} \times \mathbb{N}^{\mathbb{Z}} \mid |\alpha| = |\beta| < +\infty,\ \sum_{j\in\mathbb{Z}} j(\alpha_j - \beta_j) = 0\}$, satisfying the reality condition $H_{\alpha,\beta} = \overline{H}_{\beta,\alpha}$, $\forall\,(\alpha,\beta) \in \mathcal{M}$.

Given a Hamiltonian as in (8), we denote by X_H its Hamiltonian vector field with respect to the symplectic form $\Omega := \mathrm{i}\sum_{j\in I} du_j \wedge d\bar{u}_j$. We say that $H \in \mathcal{H}_r(\ell^2_{\mathrm{w}})$ for $r > 0$ if the Hamiltonian vector field $X_{\underline{H}}$ of the Cauchy majorant of the Hamiltonian is a bounded analytic map $B_r(\ell^2_{\mathrm{w}}) \to \ell^2_{\mathrm{w}}$:

$$|H|_{r,\ell^2_{\mathrm{w}}} := \frac{1}{r}\Big(\sup_{|u|_{\ell^2_{\mathrm{w}}}\le r} |X_{\underline{H}}|_{\ell^2_{\mathrm{w}}}\Big) < \infty, \quad \underline{H}(u) := \sum_{(\alpha,\beta)\in\mathcal{M}} |H_{\alpha,\beta}| u^\alpha \bar{u}^\beta. \tag{9}$$

The space $\mathcal{H}_r(\ell^2_{\mathrm{w}})$ is closed with respect to Poisson brackets and, moreover, if $S \in \mathcal{H}_r(\ell^2_{\mathrm{w}})$ has a sufficiently small norm then it generates a well-defined time-one flow $B_r(\ell^2_{\mathrm{w}}) \to \ell^2_{\mathrm{w}}$. Finally, we say that a Hamiltonian H has scaling (degree) $\mathtt{d}(H) \ge \mathtt{d}$ if[6]

$$H = \sum_{(\alpha,\beta)\in\mathcal{M}:|\alpha|+|\beta|\ge \mathtt{d}+2} H_{\alpha,\beta} u^\alpha \bar{u}^\beta;$$

note that the scaling degree is additive with respect to Poisson brackets.

Now we recall that the NLS Hamiltonian (1) has the form $H = D_\omega + P$ with P of scaling ≥ 2 and $D_\omega := \sum_{j\in\mathbb{Z}} \omega_j |u_j|^2$ having scaling zero.

Let us conjugate H by the time-one flow Φ^1_S with generating Hamiltonian S. Denoting[7] $\mathrm{ad}_S : H \mapsto \{S, H\}$, the Lie exponentiation formula reads

$$H \circ \Phi^1_S = e^{\{S,\cdot\}} H = D_\omega + P + \{S, D_\omega\} + \sum_{h=2}^{\infty} \frac{\mathrm{ad}_S^{h-1}}{h!}\{S, D_\omega\} + \sum_{k=1}^{\infty} \frac{\mathrm{ad}_S^k}{k!} P.$$

Now at least at the level of formal power series, the last two summands have scaling ≥ 4 (just by the additivity of the scaling degree), so our goal is to cancel the term $P + \{S, D_\omega\}$ (which has scaling ≥ 2) up to a remainder which is either action preserving or of scaling ≥ 4. Let

$$\mathcal{R}_r(\ell^2_{\mathrm{w}}) := \Big\{ H \in \mathcal{H}_r(\ell^2_{\mathrm{w}}) \;\Big|\; H = \sum_{\alpha\ne\beta} H_{\alpha,\beta} u^\alpha \bar{u}^\beta \Big\}, \tag{10}$$

introduce the decomposition $\mathcal{H}_r(\ell^2_{\mathrm{w}}) = \mathcal{R}_r(\ell^2_{\mathrm{w}}) \oplus \mathcal{K}_r(\ell^2_{\mathrm{w}})$ and the continuous projections $\Pi_{\mathcal{K}} H := \sum_{\alpha=\beta} H_{\alpha,\beta} u^\alpha \bar{u}^\beta$, $\Pi_{\mathcal{R}} H := \sum_{\alpha\ne\beta} H_{\alpha,\beta} u^\alpha \bar{u}^\beta$. Now all Hamiltonians in $\mathcal{K}_r(\ell^2_{\mathrm{w}})$ are action preserving while for any $R \in \mathcal{R}_r(\ell^2_{\mathrm{w}})$, at least formally, one has

$$R + \{S, D_\omega\} = 0 \quad \Leftrightarrow \quad S = -\mathrm{i} \sum_{(\alpha,\beta)\in\mathcal{M}} \frac{R_{\alpha,\beta}}{\omega\cdot(\alpha-\beta)} u^\alpha \bar{u}^\beta,$$

this is called the "homological equation." Thus we choose S_0 so that $\{S_0, D_\omega\} + \Pi_{\mathcal{R}} P = 0$ and, provided that we can show that it is well defined and has a sufficiently small norm, we have found a change of variables $e^{\mathrm{ad}_{S_0}} : D_\omega + P \rightsquigarrow D_\omega + Z_1 + P_1$ where Z is action

[6] Note that saying that H has scaling $\ge \mathtt{d}$ means that its Taylor series has minimal degree of homogeneity $\ge \mathtt{d} + 2$.

[7] The Poisson brackets are defined as $\{S, H\} := dS(X_H)$.

preserving and now P_1 has scaling ≥ 4. Following the same scheme, if we choose S_1 so that $\{S_1, D_\omega\} + \Pi_{\mathcal{R}} P_1 = 0$ and again S_1 has sufficiently small norm then, composing the two changes of variables, we conjugate $D_\omega + P \rightsquigarrow D_\omega + Z_2 + P_2$ where now P_2 has scaling ≥ 6.

Assuming that $P \in \mathcal{H}_{r_0}(\ell^2_{\mathtt{w}_0})$, for some r_0, $\mathtt{w}_0 = (\mathtt{w}_{0,j})_{j\in\mathbb{Z}}$, does not imply that S_1, S_2 are such. We have reduced the problem to finding a correct weighted space such that $S_1, S_2 \in \mathcal{H}_r(\ell^2_{\mathtt{w}})$ for r small. Note that since they have scaling ≥ 2 and ≥ 4, respectively, once S_1, S_2 are well defined their norm can be made arbitrarily small by just taking r small.

Let us consider the simple example of $\mathtt{w} = \mathtt{w}_s = (\langle j\rangle^2 e^{s\langle j\rangle^\theta})_{j\in\mathbb{Z}}$. Direct computations show that $P \in \mathcal{H}_{r_0}(\ell^2_{\mathtt{w}_0})$, for some $r_0 > 0$. Now there are two key points:

Immersions. If $H \in \mathcal{H}_{r_0}(\ell^2_{\mathtt{w}_0})$ then $H \in \mathcal{H}_r(\ell^2_{\mathtt{w}_s})$ for all $r \leq r_0$ and $s \geq 0$ and the norm is decreasing in s and increasing in r.

Homological equation. If $R \in \mathcal{H}_r(\ell^2_{\mathtt{w}_s})$ then the solution S of the homological equation belongs to $\mathcal{H}_r(\ell^2_{\mathtt{w}_{s+\sigma}})$ for all $\sigma > 0$ and

$$|S|_{\ell^2_{\mathtt{w}_{s+\sigma}}} \leq e^{C\sigma^{-\frac{3}{\theta}}} |R|_{\ell^2_{\mathtt{w}_s}}. \tag{11}$$

Thus for any given $\sigma > 0$, there exists r_2 such that, for $|r| \leq r_2$, both S_1, S_2 are well defined and small and the composition of their time-one flows maps $B_r(\ell^2_{\mathtt{w}_{2\sigma}}) \to \ell^2_{\mathtt{w}_{2\sigma}}$. This gives all the necessary estimates and one can repeat this procedure $\mathtt{N}$ times. At the end, for $|r| \leq r_{\text{fin}}$, we get a change of variables $B_r(\ell^2_{\mathtt{w}_{\mathtt{N}\sigma}}) \to \ell^2_{\mathtt{w}_{\mathtt{N}\sigma}}$ which conjugates the Hamiltonian to $D_\omega + Z_{\mathtt{N}} + R_{\mathtt{N}}$ where $Z_{\mathtt{N}}$ depends only on the actions and $R_{\mathtt{N}}$ has scaling $2\mathtt{N} + 2$. Of course, we have also estimates on the norms of $Z_{\mathtt{N}}$, $R_{\mathtt{N}}$, and $R_{\mathtt{N}} \sim r^{2\mathtt{N}+2}$. Now if we want a stability estimate in $\ell^2_{\mathtt{w}_s}$, we first leave $\mathtt{N}$ as a free parameter and fix $\sigma = s\mathtt{N}^{-1}$. This gives a stability estimate $\sim r^{-(2\mathtt{N}+2)}$. Finally, by optimizing $\mathtt{N}$, one gets the subexponential bounds. Now if we take any weight $\mathtt{w}$ and follow the same strategy, we only need to verify the immersions and control the homological equation, this is what we do in [25].

The main difference in the Sobolev case is that in solving the homological equation, if $R \in \mathcal{H}(\ell^2_{\mathtt{w}})$ with $\mathtt{w}_j = \langle j\rangle^p$, then $S \in \mathcal{H}(\ell^2_{\mathtt{w}'})$ with $\mathtt{w}'_j = \langle j\rangle^{p+\tau}$ with τ fixed. This is a typical feature in the setting with finite regularity, in this context it produces the relation between stability time appearing in [9].

3. QUASIPERIODIC SOLUTIONS

There is by now a vast literature on quasiperiodic solutions for NLS (mainly confined to the case when $\mathcal{M}$ is a torus or a sphere), covering also PDEs without external parameters and quasilinear PDEs. The first results in this direction (in the early 1990s, we mention, for example, Kuksin, Wayne, Craig, Bourgain, Pöschel) were for semilinear 1D PDEs with with periodic or Dirichlet boundary conditions. There were essentially two approaches, both quadratic iteration schemes generalizing Newton's steepest descent method:

(1) Extend *KAM theory of elliptic tori* to the infinite-dimensional setting (see [67–70,73,80]) thus proving not only existence but also linear stability.

This amounts to looking for an analytic symplectic change of variables which conjugates the Hamiltonian to a normal form where the invariant torus is flat, namely there exist a set of indexes $\mathcal{S} \subset \mathbb{Z}$ of cardinality n and a set of symplectic variables such that the torus in these variables is $u_j = 0$ for all $j \notin \mathcal{S}$ and $|u_j| = \text{const.}$ for $j \in \mathcal{S}$. Finally, the dynamics on the torus is a linear translation (with Diophantine frequency) and linearized dynamics in the normal directions to the torus is diagonal and elliptic.

(2) Look for the torus embedding $U : \mathbb{T}^n \mapsto \mathtt{h}$ (the phase space) as the solution of a nonlinear functional equation $F(U) = 0$. Apply a Newton method to construct successive approximations, provided that one has some control on the left inverse for the linearized operator $\mathrm{d}F(U)$ at an approximate quasiperiodic solution. The main difficulty is that $\mathrm{d}F(U)$ is a small perturbation of a diagonal operator whose spectrum accumulates to zero, thus there is a small divisor problem which is dealt with by a multiscale analysis. This is the so-called *Craig–Wayne–Bourgain* (CWB) approach, see [28, 40] and the papers [16, 18] for a more modern point of view. Of course, these two approaches have many similarities and can be combined in an effective way; see, for instance, [17].

To make the statements more concrete, let us restrict to the NLS equation (2). We fix a set $\mathcal{S} \subset \mathbb{Z}$ of cardinality n and assume, for simplicity, that $V_j = 0$ for all $j \notin \mathcal{S}$; finally, we fix an appropriate phase space (say, $\ell^2_{\mathtt{w}}$ for some weight, e.g., $\mathtt{w}_j = \langle j \rangle$). We look for solutions close to the n-dimensional approximately invariant torus $\mathcal{T}_n$ such that $|u_j| = 0$ for all $j \notin \mathcal{S}$ and $|u_j|^2 = I_j > 0$ otherwise. In a neighborhood of such torus, we can pass to "elliptic-action angle variables" $\chi :\ (\theta, y, z) \to u$, with $u_j = \sqrt{I_j + y_j}e^{\mathrm{i}\theta_j}$, for $j \in \mathcal{S}$ while $z_j = u_j$ for $j \notin \mathcal{S}$. In these variables the NLS Hamiltonian reads

$$H_{\mathrm{NLS}} = \sum_{j \in \mathcal{S}} (j^2 + V_j) y_j + \sum_{j \notin \mathcal{S}} j^2 |u_j|^2 + \mathcal{P}. \tag{12}$$

Now the KAM scheme ensures, for many values of V, the existence of a bounded symplectic change of variables, defined in a neighborhood of $\mathcal{T}_n$, which conjugates the Hamiltonian to

$$\widetilde{\omega}(V) \cdot y + \sum_{j \in \mathbb{Z} \setminus \mathcal{S}} \Omega_j(V)|z_j|^2 + \mathcal{P}_{\mathrm{fin}}, \quad \mathcal{P}_{\mathrm{fin}} = O(y^2 + yz + z^3), \tag{13}$$

where $\widetilde{\omega}, \Omega$ are appropriate real functions of V. This means not only that $\mathcal{T}_n$ is now invariant, but also that the dynamics in the normal directions z_j is (at least at the linear level) the rotation by $e^{\mathrm{i}\Omega_j t}$.

Conversely, with the CWB method one can conjugate the NLS Hamiltonian to a normal form like (12), but where the quadratic terms in z are neither diagonal nor independent of θ.

While the first approach is technically simpler and gives a stronger result, it requires stronger hypotheses which give some control on the difference of distinct eigenvalues of the linearized equation at an approximately invariant torus, that are not verified for many physically interesting PDEs. Indeed, in the case of manifolds of dimension greater than one, the first results were by the CWB method, we mention, for instance, [29, 31] for the NLS on tori and [19] for a forced NLS on simple compact Lie groups.

Regarding linear stability issues, note that if one proves existence of solutions (via CWB) then one can prove the linear stability a posteriori, for instance, by proving that the PDE linearized at the quasiperiodic solution is "reducible," i.e., can be conjugated to constant coefficients (or even diagonalized) via a time quasiperiodic change of variables on the phase space. Then the stability can be inferred by solving the linear dynamics, which becomes trivial. In this setting if one wants to conjugate via a close to identity change of variables (since the solution is small, the linearized operator is close to diagonal, and one hopes to apply some perturbative argument), one has to deal with small divisors related to the differences of eigenvalues, just as in the KAM case. This is just like diagonalization algorithms for finite-dimensional matrices close to a diagonal one, where one needs distinct eigenvalues in order to apply perturbative arguments. Of course, in infinite dimension, differences of eigenvalues may also accumulate to zero (and typically do in our setting), so the best hope is to impose some nonuniform lower bound. Thus, proving linear stability for the solutions of PDEs in dimension higher than one is typically a rather difficult question, due to the multiplicity of the eigenvalues, the idea is to introduce a partition of the eigenvalues into clusters so that one has control on the difference of eigenvalues in different clusters while the dynamics inside a cluster is stable.

A breakthrough was in [44, 45] where the authors proved reducibility for the NLS equation with a convolution potential on $\mathbb{T}^d$. This requires a subtle analysis and the introduction of the class of Töplitz–Lipschitz functions. Their approach is based on a good control of the asymptotics of eigenvalues of operators of the form $-\Delta + V(x, \omega t)$ where V is periodic in all its variables and $x \in \mathbb{T}^d$, see also [12] for a discussion on general flat tori. As far as I am aware, the only other manifolds on which there are reducibility results are spheres (see [51] for Zoll manifolds). Instead of the reducibility, one can concentrate on the control of Sobolev norms for the corresponding linear operator. This has been discussed by many authors, see [10, 11, 24, 30, 41].

Regarding the question of parameterless PDEs, most results are in the 1D case starting from [70, 75]. Let us briefly discuss the completely resonant (NLS) with $V = 0$ and $\mathcal{M} = \mathbb{T}^d$. The idea is to first perform one step of Birkhoff normal form in order to extract parameters from the initial data. Unfortunately, if $d > 1$, the normal form is *not integrable* and actually has a rather complicated structure. This is well known and used, for instance, in [33] in order to prove explosion of Sobolev norms. Building on the paper [56] for the case $d = 2$, in [76–78] we discussed this problem and showed that in the neighborhood of appropriately chosen initial data the NLS Hamiltonian after one step of BNF is indeed integrable, satisfies the twist condition, and has appropriately controlled distinct eigenvalues.

Interestingly, the good initial data are found by first choosing the Fourier support $\mathcal{S} \subset \mathbb{Z}$ in a generic way (i.e., outside the zero set of some nontrivial polynomial) and then by choosing the actions on such support in some Cantor set. This allows proving the following theorem for any equation of the type (NLS) with $\mathcal{M} = \mathbb{T}^d$ and $V = 0$ (see also [79]):

Theorem ([78]). *Fix any n and any choice of generic frequencies $S = \{\mathtt{j}_1, \dots, \mathtt{j}_d\} \subset \mathbb{Z}^n$. For ε sufficiently small, there exists a compact set $\mathcal{C}_\varepsilon \in [\varepsilon/2, \varepsilon]^n$ of positive measure,* *parametrizing bijectively a set of analytic quasiperiodic solutions of NLS of the type*

$$\mathcal{C}_\varepsilon \ni \xi \mapsto u(\xi, x, t) = \sum_{j \in \mathcal{S}} \sqrt{\xi_j} e^{\mathrm{i}t(|j|^2 + \omega_j(\xi))} e^{\mathrm{i} j \cdot x} + O(\xi^2).$$

Moreover, the linearized NLS operator at a quasiperiodic solution is conjugated to a constant coefficient block-diagonal form with uniform bounds on the dimension of the blocks.

In (NLS), the nonlinearity is analytic and so the quasiperiodic solutions we have discussed are at least C^∞. If one considers nonlinearities with finite (but rather high) regularity then one can obtain analogous results (both KAM and Nash–Moser) for finite regularity solutions.

All the results described above are for semilinear PDEs. In order to deal with the quasilinear case, where the derivatives in the nonlinearity have the same order of the linear part, one needs to introduce new perspectives. A real breakthrough appeared in [2], where the authors introduced ideas from pseudodifferential calculus (see also [64]) to produce a general method applicable to PDEs on the circle, we mention [2,3,49,54], as well as [1,22,48] for the water wave equation. There have been some extensions of these results to higher dimensions; we mention [4,38,71].

3.1. Questions and open problems

Q6. Can one develop a "general" pseudodifferential approach to deal with quasilinear dispersive PDEs in high dimension?

Even on tori, the results up to now rely on special features of the equations. An interesting strategy was developed in [11] for a linear NLS (see also [72]).

Q7. Can one study the NLS with external parameters or even a multiplicative potential as in [16] when $\mathcal{M}$ is a compact Lie group? And having done this, can one prove a reducibility result? Can one deal with the parameterless case?

These questions are largely open and interesting even in the case when $\mathcal{M}$ is an irrational torus.

Q8. It is expected that the solutions described in this section are linearly stable or have at most a finite number of linearly unstable directions. What kind of normal form can be achieved close to the tori? This was discussed, for instance, in [15]. What can be said about nonlinear stability/instability?

In [46] the authors discuss polynomial stability times close to a periodic plane wave solution. For the NLS on $\mathbb{T}^2$, there are a number of instability results, stemming from the paper [33]; we mention [62] close to the plane wave solutions and [60] close to one-dimensional quasiperiodic solutions.

4. ALMOST PERIODIC SOLUTIONS

By definition, *almost-periodic solutions* are solutions which are limits (in the uniform topology in time) of quasiperiodic solutions. A very naïve approach would be to find them by just constructing quasiperiodic solutions supported on invariant tori of dimension n and then take the limit $n \to \infty$. Unfortunately, the KAM procedure (of, say, [70,74]) is not uniform in the dimension n, and, by taking the limit, one just falls on the elliptic fixed point.

A refined version of this very natural idea is to construct a sequence of invariant tori of growing dimension using at each step the invariant torus of the previous one as an unperturbed solution: in this way, the $(n+1)$th and nth tori are extremely close, leading to very regular solutions. This was done by Pöschel [75] by using the KAM method and by Bourgain [28] via the Nash–Moser approach, getting solutions which decay at least superexponentially (see also [55] for solutions with exponential decay).

A different approach was proposed by Bourgain in [32] to study the translation-invariant NLS (2). The idea is to construct a converging sequence of infinite-dimensional approximately-invariant manifolds and prove that the limit is the support of the desired almost-periodic solution. The fact that one does not restrict to neighborhoods of finite-dimensional tori allows for a better control of the small-divisors and hence the construction of more general, i.e., *less regular* solutions in spaces of Gevrey regularity. A drawback of Bourgain's approach is that it works only for "maximal tori," namely one has to impose some lower bounds on the actions of the approximately invariant ∞-dimensional tori. His statement concerns the quintic NLS on $\mathbb{S}^1$, i.e., (2) with $f(y) = 3y^2$. Here we give a slightly more general version taken form [35], where the authors prove also stability of the tori.

Theorem (Gevrey almost periodic solutions, [32]). *Fix $s > 0$, $0 < \theta < 1$. For all $0 < r \ll 1$ small enough, and any "approximate initial datum"*

$$u_0(x) = \sum_{j\in\mathbb{Z}} \sqrt{I_j} e^{\mathrm{i}jx} \quad \textit{such that } r/2 < \sqrt{I_j} e^{s\langle j\rangle^\theta} \langle j\rangle^2 < r, \tag{14}$$

there exists a set of positive measure in $[-1/2, 1/2]^{\mathbb{Z}}$ (depending on u_0) such that for all V in such a set there exists one almost-periodic solution with $|\sqrt{I_j} - |u_j(t)|| \ll re^{-s\langle j\rangle^\theta}\langle j\rangle^{-2}$.

Similar results were proved in [37] for the wave equation.

In [26] we gave a more precise description of these solutions, proving that for all frequencies $\omega \in \mathtt{D}_\gamma$ and *for any* approximate initial datum u_0 in a small ball in $\ell^\infty_{\mathtt{w}_s} := \{u : (e^{s\langle j\rangle^\theta}\langle j\rangle^2 u_j)_{j\in\mathbb{Z}} \in \ell^\infty\}$, there exists a potential $V = V(\omega, u_0) \in \ell^\infty$ such that the corresponding NLS equation has an almost periodic solution of frequency ω close to $\sqrt{I}$. Furthermore, in [26] we developed a strategy which allows constructing in a unified context tori of Gevrey regularity and of any dimension essentially supported on the Fourier modes belonging to any subset $\mathcal{S} \subseteq \mathbb{Z}$. Essentially, this amounts to Bourgain's result, but in (14) we only need $\sqrt{I_j}e^{s\langle j\rangle^\theta}\langle j\rangle^2 < r$. This is an interesting novelty because, with Bourgain's condition, the acceptable u_0 are of zero measure.

In [26] we discussed Gevrey solutions, but one can find even less regular solutions, see [34]. However, the question of finding maximal tori which are not C^∞ is still open. If one looks for "nonmaximal" tori, approximately supported on an infinite set $\mathcal{S}$, then one can reach very low regularity. Again, just as in the quasiperiodic case, the choice of the support can be used as a precious additional source of parameters. Given a function $u : \mathbb{R}^2 \to \mathbb{C}$

which is 2π-periodic in x and such that the map $t \mapsto u(t,\cdot) \in \mathcal{F}(\ell^1)$ is continuous,[8] we say that u is a weak solution of (2) if, for any smooth compactly supported function $\chi : \mathbb{R}^2 \to \mathbb{R}$, one has

$$\int_{\mathbb{R}^2} (-\mathrm{i}\chi_t + \chi_{xx})u - \big(V * u + f\big(|u|^2\big)u\big)\chi \, dx \, dt = 0. \tag{15}$$

Theorem ([27]). *For almost every Fourier multiplier V, there exist infinitely many small-amplitude weak almost-periodic solutions of* (2). *Infinitely many of such solutions are not classical and infinitely many are classical.*

Unfortunately, such solutions are not in any way typical and, in fact, correspond to very special infinite-dimensional elliptic tori.

The question whether full-dimensional tori exist in the Sobolev class is still open. Apart from the interest *per se*, these low regularity solutions could be used in order to find solutions for parameterless PDEs. Essentially one wants to solve the “countertem equation” $V(\omega, u_0) = 0$ by finding $u_0 = u_0(\omega)$.

4.1. Questions and open problems

Q9. Are almost-periodic solutions generic in some Banach space? For example, is it true that for many convolution potentials the tori cover a positive measure set (with respect to the probability product measure in the Gevrey space $B_r(\ell^\infty_{w_s})$)?

Q10. Can one construct maximal tori with Sobolev regularity?

Q11. Can one construct almost-periodic solutions for the NLS on higher-dimensional manifolds?

At least in the case of tori, most of the strategy proposed by Bourgain can be generalized, the main point here seems to be the choice of a smart Diophantine condition.

Q12. Can one construct almost-periodic solutions for parameterless NLS equations? Here even the case of 1D NLS with generic multiplicative potentials would be interesting.

Q13. Can one deal with unbounded nonlinearities?

This has been discussed in the case of a forced quasilinear Airy equation in [39], generalizing the approach for the quasiperiodic case.

Q14. Can one construct almost-periodic solutions for small perturbations of integrable PDEs?

In the case of quasiperiodic solutions, there are a number of results, we mention [23,65,66,68]. In order to cover the almost-periodic case, the main point is to control convexity properties for the Hamiltonian in action angle variables.

4.2. An idea of the strategies

Let us first discuss the linear case. Recall that we are restricting to 1D NLS with convolution potential so that the linear actions $|u_j|^2$ are constants of motions and the dynamics

[8] Here $\mathcal{F}$ is the usual Fourier transform.

is

$$u_j(t) = u_j(0)e^{\mathrm{i}\omega_j t}, \quad j \in \mathbb{Z}, \quad \omega = (\omega_j)_{j\in\mathbb{Z}}, \, \omega_j := j^2 + V_j.$$

Let us call $\mathcal{S}_0 := \{j \in \mathbb{Z} \mid u_j(0) \neq 0\}$.

If $\mathcal{S}_0$ is a finite set, the corresponding solution $u(t,x) := \sum_{j\in\mathbb{Z}} u_j(t)e^{\mathrm{i}jx}$ is quasiperiodic and analytic both in time and space.

If $\mathcal{S}_0$ is infinite, the regularity of $u(t,x)$ obviously depends on that of the initial datum. If $u(0) := (u_j(0))_{j\in\mathbb{Z}} \in \ell^1$ then $u(t,x)$ is a weak solution of (2). Moreover, such a solution is a time almost-periodic function, being the limit of the quasiperiodic truncations $\sum_{|j|\le n} u_j(0)e^{\mathrm{i}\omega_j t + \mathrm{i}jx}$ as $n \to \infty$. Note, finally, that the regularity (in any reasonable weighted space) is that of the initial datum. The support of each solution is an invariant torus: given an initial datum $u(0)$, set $I := (I_j)_{j\in\mathbb{Z}}$ with $I_j = |u_j(0)|^2$, then the motion is supported on $\mathcal{T}_I := \{u : |u_j|^2 = I_j \ \forall j \in \mathbb{Z}\}$.

Now the nature of these invariant tori strongly depends on the choice of the phase space. When discussing the stability of zero, a natural context was to work with the Hilbert spaces $\ell^2_{\mathtt{w}}$, which induce on the tori the product topology. In this context, however, this produces a number of problems, related to the density of finite-dimensional tori. In KAM algorithms, one typically wants to “Taylor expand” close to the approximately-invariant tori, but this requires a Banach manifold structure, so even though the product topology is the natural one with respect to the group structure, the KAM algorithm seems to require a finer choice, e.g., weighted spaces based on ℓ^∞,

$$\ell^\infty_{\mathtt{w}} := \Big\{u := (u_j)_{j\in\mathbb{Z}} \in \ell^\infty(\mathbb{C}) : |u|_{\mathtt{w}} := \sup_{j\in\mathbb{Z}} \mathtt{w}_j |u_j| < \infty\Big\}. \tag{16}$$

Given a sequence $I = (I_j)_{j\in\mathbb{Z}}$ with $I_j \ge 0$ and $\sqrt{I} := (\sqrt{I_j})_{j\in\mathbb{Z}} \in \ell^\infty_{\mathtt{w}}$, we consider the torus

$$\mathcal{T}_I := \big\{u \in \ell^\infty_{\mathtt{w}} : |u_j|^2 = I_j \ \forall j \in \mathbb{Z}\big\}. \tag{17}$$

Now the map

$$\mathrm{i} : \mathbb{T}^{\mathcal{S}_0} \to \mathcal{T}_I \subset \mathtt{w}_p, \quad \varphi = (\varphi_j)_{j\in\mathcal{S}_0} \mapsto \mathrm{i}(\varphi), \quad \mathrm{i}_j(\varphi) := \begin{cases} \sqrt{I_j}e^{\mathrm{i}\varphi_j} & \text{for } j \in \mathcal{S}_0, \\ \mathrm{i}_j(\varphi) := 0 & \text{otherwise,} \end{cases} \tag{18}$$

is an analytic immersion provided that we endow $\mathbb{T}^{\mathcal{S}_0}$ with the ℓ^∞-topology. Note that, assuming also that $\inf_j \sqrt{I_j}\mathtt{w}_j > 0$, the map i is an embedded torus, in a neighborhood of which one can construct local action angle variables. By construction, the linear dynamics on the torus $\mathcal{T}_I$ is $\varphi \to \varphi + \omega t$.

Since the map $t \mapsto \omega t \in \mathbb{T}^{\mathcal{S}_0}$ is not even continuous (endowing $\mathbb{T}^{\mathcal{S}_0}$ with the ℓ^∞-topology and recalling that $\omega_j \sim j^2$), the regularity of $t \mapsto \mathrm{i}(\omega t)$ depends on the choice of the actions I_j. If we assume $\inf_j \sqrt{I_j}\mathtt{w}_j > 0$, then it is not continuous with respect to the strong[9] topology.

[9] Note that the map is continuous with respect to the product topology, which coincides with the weak-$*$ topology on bounded sets.

In contrast with the finite-dimensional case, even if ω has rationally independent entries, it is not straightforward to understand whether this invariant torus is densely filled[10] by the solution's orbit or not. In fact, this issue is related to the asymptotic behavior of ω. For example, if we require that $\mathcal{S}_0$ is a very sparse set then the density follows, see [27].

We say that $\mathcal{T}_I$ is a KAM torus of frequency $\omega \in \mathbb{R}^{\mathbb{Z}}$ for the Hamiltonian N if it has the form $\sum_{j\in\mathbb{Z}} \omega_j |u_j|^2 + P$, with $P = O(|u|^2 - I)^2$, so that the Hamiltonian vector field X_P vanishes on the torus $\mathcal{T}_I$. Indeed, under the hypotheses above, $\mathcal{T}_I$ is invariant and the restricted dynamics is linear with frequency ω, namely

$$u_j(t) = u_j(0)e^{\mathrm{i}\omega_j t}, \quad \bigl|u_j(0)\bigr|^2 = I_j, \quad j \in \mathbb{Z}. \tag{19}$$

Note that in this definition the only relevant frequencies are those corresponding to nonzero actions.

Let us now fix the support of the solution by taking a subset $\mathcal{S} \subset \mathbb{Z}$ and consider $\sqrt{I} \in \bar{B}_r(\ell^\infty_{\mathrm{w}})$ with $I_j = 0$ for $j \in \mathcal{S}^c$. We say that the torus $\mathcal{T}_I$ is an elliptic KAM torus of frequency $\nu \in \mathbb{R}^{\mathcal{S}}$ for the Hamiltonian N with normal frequency $(\Omega_j)_{j\in\mathbb{Z}\setminus\mathcal{S}}$ if, setting for notational convenience $u_j = v_j$ for $j \in \mathcal{S}$ and $u_j = z_j$ otherwise, one has (compare with (13) with $y = |v|^2 - I$)

$$N = \sum_{j\in\mathcal{S}} \nu_j |v_j|^2 + \sum_{j\in\mathbb{Z}\setminus\mathcal{S}} \Omega_j |z_j|^2 + R, \quad R = O\bigl(\bigl(|v|^2 - I\bigr)^2 + \bigl(|v|^2 - I\bigr)z + z^3\bigr).$$

We can now state our version of KAM theorem for infinite tori. We shall concentrate on the elliptic tori and in particular on the low regularity case. To this purpose, for $p > 1$, we consider the Sobolev space $\ell^\infty_{\mathrm{w}_p}$ where now $\mathrm{w}_p = \langle j\rangle^p$. In order to work in this low regularity setting, we need to impose some conditions on $\mathcal{S}$ requiring that it is sufficiently sparse (for instance, $\mathcal{S} = 2^{\mathbb{N}}$). For any such $\mathcal{S}$, for all $r > 0$ sufficiently small, for every $\sqrt{I} \in B_r(\ell^\infty_{\mathrm{w}_p})$ with $I_j = 0$ for $j \in \mathcal{S}^c$, we have

Theorem (Sobolev case [27]). *There exists a positive measure Cantor-like set in* $[-1/2, 1/2]^{\mathbb{Z}}$ *and for all V in this set there exists a close to identity change of variables* $\Phi : B_r(\ell^\infty_{\mathrm{w}_p}) \to \ell^\infty_{\mathrm{w}_p}$ *such that* $\mathcal{T}_I$ *is an elliptic KAM torus* $H_{\mathrm{NLS}} \circ \Phi$.

To give an idea of the proof, let us restrict to the maximal case. By the very definition of a KAM torus, we wish to decompose a regular Hamiltonian as a sum of regular terms with an increasing "order of zero" at $\mathcal{T}_I$. Namely, given a Hamiltonian $H \in \mathcal{H}_r(\ell^\infty_{\mathrm{w}})$, we wish to write it as sum of three terms, all in $\mathcal{H}_r(\ell^\infty_{\mathrm{w}})$,

$$H = H^{(-2)} + H^{(0)} + H^{(\geq 2)}$$

so that $X_{H^{(-2)}}$ is not tangent to $\mathcal{T}_I$, $H^{(0)}$ vanishes at $\mathcal{T}_I$ and its vector field is tangent but not necessarily null, while $H^{(\geq 2)} = O(|u|^2 - I)^2$ (this means that the corresponding vector field vanishes at $\mathcal{T}_I$). The main point is to make a power series expansion centered at I without introducing a singularity at zero. Start from a regular Hamiltonian $H(u)$ expanded in Taylor

[10] In the product topology such solutions are always dense.

series at $u = 0$ and rewrite every monomial as $|u|^{2m}u^\alpha \bar{u}^\beta$ with α, β with distinct support. Then define an *auxiliary Hamiltonian* $\mathtt{H}(u, w)$ (here $w = (w_j)_{j\in\mathbb{Z}}$ are auxiliary "action" variables) by the substitution $|u|^{2m}u^\alpha\bar{u}^\beta \rightsquigarrow w^m u^\alpha \bar{u}^\beta$.

Since we are considering functions on an ℓ^∞ space, it turns out that $\mathtt{H}(u, w)$ is analytic in both u and w. In particular, we can Taylor expand with respect to w at the point $w = I$, with I being in the domain of analyticity.

Then we set $H^{(-2)}(u) := \mathtt{H}(u, I)$, $H^{(0)}(u) := D_w \mathtt{H}(u, I)[|u|^2 - I]$, and $H^{(\geq 2)}(u)$ is what is left. As an example, the Hamiltonian $H = |u_1|^2|u_2|^4 \operatorname{Re}(u_1\bar{u}_3)$ has auxiliary Hamiltonian $\mathtt{H}(u, w) = w_1 w_2^2 \operatorname{Re}(u_1\bar{u}_3)$ and decomposes into

$$H^{(-2)} := I_1 I_2^2 \operatorname{Re}(u_1\bar{u}_3), \quad H^{(0)} := \big[I_2^2\big(|u_1|^2 - I_1\big) + 2I_1 I_2\big(|u_2|^2 - I_2\big)\big] \operatorname{Re}(u_1\bar{u}_3),$$
$$H^{(\geq 2)} := \big(|u_1|^2\big(|u_2|^2 - I_2\big) + 2I_2\big(|u_1|^2 - I_1\big)\big)\big(|u_2|^2 - I_2\big) \operatorname{Re}(u_1\bar{u}_3).$$

The above decomposition is, *at a formal level*, the same introduced by Bourgain in [32], but in [26] we show that it is, in fact, a direct sum decomposition $\mathcal{H}_r(\ell^\infty_{\mathtt{w}}) = \mathcal{H}_r^{(-2)}(\ell^\infty_{\mathtt{w}}) \oplus \mathcal{H}_r^{(0)}(\ell^\infty_{\mathtt{w}}) \oplus \mathcal{H}_r^{(\geq 2)}(\ell^\infty_{\mathtt{w}})$, with explicit control on the projections. An important point is that all our construction works independently of the "dimension" of $\mathcal{T}_I$, namely it never requires conditions of the form $I_j \neq 0$.

Let us compare this decomposition with that used for finite-dimensional tori. Note that in this case $\mathcal{S} = \mathbb{Z}$ so there are no "normal variables" z.

We consider the example above and pass to action–angle variables $\chi : (\theta, y) \to u$, with $u_j = \sqrt{I_j + y_j}e^{\mathrm{i}\theta_j}$. Then the terms canceled in a classical KAM scheme would be the first two terms in the Taylor expansion at $y = 0$ of $\mathcal{P}(\theta, y) := H \circ \chi = (I_1 + y_1)(I_2 + y_2)^2\sqrt{(I_1 + y_1)(I_3 + y_3)} \cos(\theta_1 - \theta_3)$, namely

$$\mathcal{P}(\theta, 0) = I_1^{\frac{3}{2}} I_2^2 \sqrt{I_3} \cos(\theta_1 - \theta_3),$$
$$\mathcal{P}_y(\theta, 0)[y] = \left(\frac{3}{2} I_2\sqrt{I_3}y_1 + 2I_1\sqrt{I_3}y_2 + \frac{I_1 I_2}{2\sqrt{I_3}}y_3\right) I_2\sqrt{I_1} \cos(\theta_1 - \theta_3).$$

Direct computations show that

$$\begin{aligned} H^{(-2)} \circ \chi + H^{(0)} \circ \chi &= I_1 I_2^2 \sqrt{(I_1 + y_1)(I_3 + y_3)} \cos(\theta_1 - \theta_3) \\ &\quad + \big(I_2^2 y_1 + 2I_1 I_2 y_2\big)\sqrt{(I_1 + y_1)(I_3 + y_3)} \cos(\theta_1 - \theta_3) \\ &= \mathcal{P}(\theta, 0) + \mathcal{P}_y(\theta, 0)[y] + O(y^2), \end{aligned}$$

and, obviously, $H^{(\geq 2)} \circ \chi$ is at least quadratic in y. In conclusion, we are canceling more terms than is strictly necessary, but in doing so we avoid introducing the singularity $I = 0$.

Now our result is proved by an iterative procedure. To get a feeling of the proof, let us consider the 1D NLS case (2) (recalling that $H = D_\omega + P$, with P small and $D_\omega = \sum \omega_j |u_j|^2$) and perform the first step. Just like in the case of the stability of zero and of quasiperiodic solutions, once we have identified the terms $P^{(-2)}, P^{(0)}$ (which are the obstacles to $\mathcal{T}_I$ being a KAM torus), we perform a change of variables e^{ad_S} to cancel them. It is convenient to look for $S = S^{(-2)} + S^{(0)}$ (namely such that the component of degree ≥ 2 is zero).

Our aim is to make the projections $\Pi^{(-2)}, \Pi^{(0)}$ of the new Hamiltonian, $e^{\mathrm{ad}_S}(D_\omega + P)$, “quadratically smaller” with respect to $P^{(-2)}, P^{(0)}$.

Now one can directly verify that this is achieved by choosing S as the solution of a homological equation, which now is (recall the projections on $\mathcal{R}$ defined in (10))

$$\Pi_{\mathcal{R}} P^{(-2)} + \{S^{(-2)}, D_\omega\} = 0, \tag{20}$$

$$\Pi_{\mathcal{R}} P^{(0)} + \{S^{(0)}, D_\omega\} + \Pi^{0,\mathcal{R}}\{S^{(-2)}, P^{\geq 2}\} = 0. \tag{21}$$

Now $P^{(-2)}, P^{(0)} \in \mathcal{H}_r(\ell^\infty_{\mathtt{w}})$ (they are analytic in a neighborhood of zero). If we chose Gevrey regularity $\mathtt{w} = \mathtt{w}_s = \langle j\rangle^2 e^{s\langle j\rangle^\theta}$, this allows us to solve the homological equation, using the estimates (11), which hold in ℓ^∞ as well. We reach a new Hamiltonian of the form $\sum_{j\in\mathbb{Z}}(\omega_j + \lambda_j)|u_j|^2 + P_1$, with $\lambda \in \ell^\infty$ and for $r_1 < r$ and $\sigma_1 > 0$, one has that $P_1^{(-2)}, P_1^{(0)}$ are quadratically smaller in the (nested) space $\mathcal{H}_{r_1}(\ell^\infty_{\mathtt{w}_{\sigma_1}})$. At the next step one repeats this procedure just with a slightly different frequency, decreasing at each step n the radius r_n and increasing σ_n in a summable way. Actually, in [26] we write all the equations in terms of the final frequency, and use a counterterm theorem á la Herman.

ACKNOWLEDGMENTS

I wish to thank L. Biasco, L. Corsi, R. Feola, E. Haus, A. Maspero, J. E. Massetti and R. Montalto for many helpful discussions and suggestions.

REFERENCES

[1] P. Baldi, M. Berti, E. Haus, and R. Montalto, Time quasi-periodic gravity water waves in finite depth. *Invent. Math.* **214** (2018), no. 2, 739–911.

[2] P. Baldi, M. Berti, and R. Montalto, KAM for quasi-linear and fully nonlinear forced KdV. *Math. Ann.* **359** (2014), 471–536.

[3] P. Baldi, M. Berti, and R. Montalto, KAM for autonomous quasilinear perturbations of KdV. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33** (2016), 1589–1638.

[4] P. Baldi and R. Montalto, Quasi-periodic incompressible Euler flows in 3D. 2020, arXiv:2003.14313.

[5] D. Bambusi, Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations. *Math. Z.* **230** (1999), no. 2, 345–387.

[6] D. Bambusi, On long time stability in Hamiltonian perturbations of nonresonant linear PDEs. *Nonlinearity* **12** (1999), 823–850.

[7] D. Bambusi, Birkhoff normal form for some nonlinear PDEs. *Comm. Math. Phys.* **234** (2003), no. 2, 253–285.

[8] D. Bambusi, J. M. Delort, B. Grébert, and J. Szeftel, Almost global existence for Hamiltonian semi-linear Klein–Gordon equations with small Cauchy data on Zoll manifolds. *Comm. Pure Appl. Math.* **60** (2007), 1665–1690.

[9] D. Bambusi and B. Grébert, Birkhoff normal form for partial differential equations with tame modulus. *Duke Math. J.* **135** (2006), no. 3, 507–567.

[10] D. Bambusi, B. Grebert, A. Maspero, and D. Robert, Growth of Sobolev norms for abstract linear Schrödinger equations. *J. Eur. Math. Soc. (JEMS)* **23** (2017), no. 2, 557–583.

[11] D. Bambusi, B. Langella, and R. Montalto, Growth of Sobolev norms for unbounded perturbations of the Laplacian on flat tori. 2020, arXiv:2012.02654.

[12] D. Bambusi, B. Langella, and R. Montalto, Spectral asymptotics of all the eigenvalues of Schrödinger operators on flat tori. 2020, arXiv:2007.07865.

[13] G. Benettin, J. Fröhlich, and A. Giorgilli, A Nekhoroshev-type theorem for Hamiltonian systems with infinitely many degrees of freedom. *Comm. Math. Phys.* **119** (1988), no. 1, 95–108.

[14] J. Bernier, E. Faou, and B. Grébert, Rational normal forms and stability of small solutions to nonlinear Schrödinger equations. *Ann. PDE* **6** (2020), 14.

[15] M. Berti and L. Biasco, Branching of Cantor manifolds of elliptic tori and applications to PDEs. *Comm. Math. Phys.* **305** (2011), no. 3, 741–796.

[16] M. Berti and Ph. Bolle, Quasi-periodic solutions for Schrödinger equations with Sobolev regularity of NLS on $\mathbb{T}^d$ with a multiplicative potential. *J. Eur. Math. Soc. (JEMS)* **15** (2013), 229–286.

[17] M. Berti and Ph. Bolle, A Nash–Moser approach to KAM theory. In *Hamiltonian partial differential equations and applications*, pp. 255–284, Fields Inst. Commun. 75, Fields Inst. Res. Math. Sci., Toronto, ON, 2015.

[18] M. Berti and Ph. Bolle, *Quasi-periodic solutions of nonlinear wave equations on the d-dimensional torus*. EMS Monogr. Math., EMS Press, 2020.

[19] M. Berti, L. Corsi, and M. Procesi, An abstract Nash–Moser theorem and quasi-periodic solutions for NLW and NLS on compact Lie groups and homogeneous manifolds. *Comm. Math. Phys.* **334** (2015), no. 3, 1413–1454.

[20] M. Berti and J. M. Delort, *Almost global existence of solutions for capillarity-gravity water waves equations with periodic spatial boundary conditions*. Springer, 2018.

[21] M. Berti, R. Feola, and F. Pusateri, Birkhoff normal form and long time existence for periodic gravity WaterWaves. *Comm. Pure Appl. Math.* (to appear), 2018, arXiv:1810.11549.

[22] M. Berti, L. Franzoi, and A. Maspero, Traveling quasi-periodic water waves with constant vorticity. *Arch. Ration. Mech. Anal.* **240** (2021), 99–202.

[23] M. Berti, T. Kappeler, and R. Montalto, Large KAM tori for quasi-linear perturbations of KdV. *Arch. Ration. Mech. Anal.* **239** (2021), 1395–1500.

[24] M. Berti and A. Maspero, Long time dynamics of Schrödinger and wave equations on flat tori. *J. Differential Equations* **267** (2019), 1167–1200.

[25] L. Biasco, J. E. Massetti, and M. Procesi, An Abstract Birkhoff Normal Form Theorem and Exponential Type Stability of the 1D NLS. *Comm. Math. Phys.* **375** (2020), no. 3, 2089–2153.

[26] L. Biasco, J. E. Massetti, and M. Procesi, Almost-periodic invariant tori for the NLS on the circle. *Ann. Inst. Henri Poincaré C* **38** (2021), no. 3, 711–758.

[27] L. Biasco, J. E. Massetti, and M. Procesi, Small amplitude weak almost periodic solutions for the 1D NLS. 2021, arXiv:2106.00499.

[28] J. Bourgain, Construction of approximative and almost periodic solutions of perturbed linear Schrödinger and wave equations. *Geom. Funct. Anal.* **6** (1996), no. 2, 201–230.

[29] J. Bourgain, Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. *Ann. of Math. (2)* **148** (1998), no. 2, 363–439.

[30] J. Bourgain, Growth of Sobolev Norms in Linear Schrödinger Equations with Quasi-Periodic Potential. *Comm. Math. Phys.* **204** (1999), 207–247.

[31] J. Bourgain, *Green's function estimates for lattice Schrödinger operators and applications*. Ann. of Math. Stud. 158, Princeton University Press, Princeton, 2005.

[32] J. Bourgain, On invariant tori of full dimension for 1D periodic NLS. *J. Funct. Anal.* **229** (2005), no. 1, 62–94.

[33] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, Transfer of energy to high frequencies in the cubic defocusing nonlinear Schrödinger equation. *Invent. Math.* **181** (2010), no. 1, 39–113.

[34] H. Cong, The existence of full dimensional KAM tori for Nonlinear Schrödinger equation. 2021, arXiv:2103.14777.

[35] H. Cong, J. Liu, Y. Shi, and X. Yuan, The stability of full dimensional KAM tori for nonlinear Schrödinger equation. *J. Differential Equations* **264** (2018), no. 7, 4504–4563.

[36] H. Cong, L. Mi, and P. Wang, A Nekhoroshev type theorem for the derivative nonlinear Schrödinger equation. *J. Differential Equations* **268** (2020), no. 9, 5207–5256.

[37] H. Cong and X. Yuan, The existence of full dimensional invariant tori for 1-dimensional nonlinear wave equation. *Ann. Inst. Henri Poincaré C* **38** (2020), no. 3, 759–786.

[38] L. Corsi and R. Montalto, Quasi-periodic solutions for the forced Kirchhoff equation on $\mathbb{T}^d$. *Nonlinearity* **31** (2018), 5075–5109.

[39] L. Corsi, R. Montalto, and M. Procesi, Almost-periodic response solutions for a forced quasi-linear Airy equation, *J. Dyn. Diff. Equat.* **33** (2021), 1231–1267.

[40] W. Craig and C. E. Wayne, Newton's method and periodic solutions of nonlinear wave equations. *Comm. Pure Appl. Math.* **46** (1993), no. 11, 1409–1498.

[41] J.-M. Delort, Growth of Sobolev norms of solutions of linear Schrödinger equations on some compact manifolds. *Int. Math. Res. Not. IMRN* **2010** (2010), no. 12, 2305–2328.

[42] J.-M. Delort, A quasi-linear Birkhoff Normal Forms method. application to the quasi-linear Klein–Gordon equation on $\mathbb{S}^1$. *Astérisque* **341** (2012), 119 p.

[43] J.-M. Delort and J. Szeftel, Long-time existence for small data nonlinear Klein–Gordon equations on tori and spheres. *Int. Math. Res. Not. IMRN* **37** (2004), 1897–1966.

[44] L. H. Eliasson and S. B. Kuksin, On reducibility of Schrödinger equations with quasiperiodic in time potentials. *Comm. Math. Phys.* **286** (2009), 125–135.

[45] L. H. Eliasson and S. B. Kuksin, KAM for the nonlinear Schrödinger equation. *Ann. of Math. (2)* **172** (2010), 371–435.

[46] E. Faou, L. Gauckler, and C. Lubich, Sobolev stability of plane wave solutions to the cubic nonlinear Schrödinger equation on a torus. *Comm. Partial Differential Equations* **38** (2013), no. 7, 1123–1140.

[47] E. Faou and B. Grébert, A Nekhoroshev-type theorem for the nonlinear Schrödinger equation on the torus. *Anal. PDE* **6** (2013), no. 6, 1243–1262.

[48] R. Feola and F. Giuliani, Time quasi-periodic traveling gravity water waves in infinite depth. *Mem. Amer. Math. Soc.* (to appear), 2020, arXiv:2005.08280.

[49] R. Feola, F. Giuliani, and M. Procesi, Reducible KAM Tori for the Degasperis–Procesi Equation. *Comm. Math. Phys.* **377** (2020), no. 3, 1681–1759.

[50] R. Feola, B. Grébert, and F. Iandoli, Long time solutions for quasi-linear Hamiltonian perturbations of Schrödinger and Klein–Gordon equations on tori. 2020, arXiv:2009.07553.

[51] R. Feola, B. Grébert, and T. Nguyen, Reducibility of Schrödinger equation on a Zoll manifold with unbounded potential. *J. Math. Phys.* **61** (2020), 071501.

[52] R. Feola and F. Iandoli, Long time existence for fully nonlinear NLS with small Cauchy data on the circle. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **22** (2021), no. 1, 109–182.

[53] R. Feola and R. Montalto, Quadratic lifespan and growth of Sobolev norms for derivative Schrödinger equations on generic tori. 2021, arXiv:2103.10162.

[54] R. Feola and M. Procesi. *J. Differential Equations* **259** (2015), no. 7, 3389–3447.

[55] J. Geng and X. Xu, Almost periodic solutions of one dimensional Schrödinger equation with the external parameters. *J. Dynam. Differential Equations* **25** (2013), no. 2, 435–450.

[56] J. Geng, X. Xu, and J. You, An infinite dimensional KAM theorem and its application to the two dimensional cubic Schrödinger equation. *Adv. Math.* **226** (2011), no. 6, 5361–5402.

[57] F. Giuliani, Transfers of energy through fast diffusion channels in some resonant PDEs on the circle. *Discrete Contin. Dyn. Syst. Ser. A* **41** (2021), no. 11, 5057–5085.

[58] F. Giuliani, M. Guardia, P. Martin, and S. Pasquali, Chaotic-like transfers of energy in Hamiltonian PDEs. *Comm. Math. Phys.* **384** (2021), no. 2, 1227–1290.

[59] M. Guardia, Growth of Sobolev norms in the cubic nonlinear Schrödinger equation with a convolution potential. *Comm. Math. Phys.* **329** (2014), no. 1, 405–434.

[60] M. Guardia, Z. Hani, E. Haus, A. Maspero, and M. Procesi, Strong nonlinear instability and growth of Sobolev norms near quasiperiodic finite-gap tori for the 2D cubic NLS equation. *J. Eur. Math. Soc. (JEMS)* (to appear), 2018, arXiv:1810.03694.

[61] M. Guardia, E. Haus, and M. Procesi, Growth of Sobolev norms for the analytic NLS on $\mathbb{T}^2$. *Adv. Math.* **301** (2016), 615–692.

[62] Z. Hani, Long-time instability and unbounded Sobolev orbits for some periodic nonlinear Schrödinger equations. *Arch. Ration. Mech. Anal.* **211** (2014), no. 3, 929–964.

[63] E. Haus and M. Procesi, KAM for beating solutions of the quintic NLS. *Comm. Math. Phys.* **354** (2017), no. 3, 1101–1132.

[64] G. Iooss, P. I. Plotnikov, and J. F. Toland, Standing waves on an infinitely deep perfect fluid under gravity. *Arch. Ration. Mech. Anal.* **177** (2005), 367–478.

[65] T. Kappeler and R. Montalto, On the stability of periodic multi-solitons of the KdV equation. *Comm. Math. Phys.* 2020, arXiv:2009.02721.

[66] T. Kappeler and J. Pöschel, *KdV & KAM*. Ergeb. Math. Grenzgeb. (3) 45, Springer, Berlin, 2003.

[67] S. B. Kuksin, Perturbation of conditionally periodic solutions of infinite-dimensional Hamiltonian systems. *Izv. Ross. Akad. Nauk Ser. Mat.* **52** (1988), no. 1, 41–63.

[68] S. B. Kuksin, A KAM theorem for equations of the Korteweg–de Vries type. *Rev. Math. Phys.* **10** (1998), 1–64.

[69] S. B. Kuksin, Fifteen years of KAM for PDE. In *Geometry, topology, and mathematical physics*, pp. 237–258, Amer. Math. Soc. Transl. Ser. 2 212, Amer. Math. Soc., Providence, RI, 2004.

[70] S. Kuksin and J. Pöschel, Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. *Ann. of Math. (2)* **143** (1996), 149–179.

[71] R. Montalto, The Navier–Stokes equation with time quasi-periodic external force: existence and stability of quasi-periodic solutions. 2020, arXiv:2005.13354.

[72] L. Parnovski and R. Shterenberg, Complete asymptotic expansion of the integrated density of states of multidimensional almost-periodic Schrödinger operator. *Ann. of Math.* **176** (2012), 1039–1096.

[73] J. Pöschel, On elliptic lower-dimensional tori in Hamiltonian systems. *Math. Z.* **202** (1989), 559–608.

[74] J. Pöschel, A KAM-theorem for some nonlinear partial differential equations. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (4)* **23** (1996), 119–148.

[75] J. Pöschel, On the construction of almost periodic solutions for a nonlinear Schrödinger equation. *Ergodic Theory Dynam. Systems* **22** (2002), 1537–1549.

[76] C. Procesi and M. Procesi, A normal form for the Schrödinger equation with analytic non-linearities. *Comm. Math. Phys.* **312** (2012), no. 2, 501–557.

[77] C. Procesi and M. Procesi, A KAM algorithm for the non-linear Schrödinger equation. *Adv. Math.* **272** (2015), 399–470.

[78] C. Procesi and M. Procesi, Reducible quasi-periodic solutions for the Non Linear Schrödinger equation. *Boll. Unione Mat. Ital.* **9** (2016), no. 2, 189–236.

[79] W. M. Wang, Energy supercritical nonlinear Schrödinger equations: Quasiperiodic solutions. *Duke Math. J.* **165** (2016), no. 6, 1129–1192.

[80] E. Wayne, Periodic and quasi-periodic solutions of nonlinear wave equations via KAM theory. *Comm. Math. Phys.* **127** (1990), 479–528.

MICHELA PROCESI

Dipartimento di Matematica e Fisica Università di Roma Tre 00156, Roma, Italy, procesi@mat.uniroma3.it

DYNAMICS AND “ARITHMETICS” OF HIGHER GENUS SURFACE FLOWS

CORINNA ULCIGRAI

ABSTRACT

We survey some recent advances in the study of (area-preserving) flows on surfaces, in particular on the typical dynamical, ergodic, and spectral properties of smooth area-preserving (or *locally Hamiltonian*) flows, as well as recent breakthroughs on *linearization* and *rigidity* questions in higher genus. We focus in particular on the *Diophantine-like conditions* which are required to prove such results, which can be thought of as a generalization of *arithmetic conditions* for flows on tori and circle diffeomorphisms. We will explain how these conditions on higher genus flows and their Poincaré sections (namely generalized interval exchange maps) can be imposed by controlling a renormalization dynamics, but are of more subtle nature than in genus one since they often exploit features which originate from the nonuniform hyperbolicity of the renormalization.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37-02; Secondary 37E35, 37E05, 11J70, 37A25, 37A30

KEYWORDS

Flows on surfaces, ergodic theory, area-preserving, locally Hamiltonian, multi-valued Hamiltonian, parabolic dynamics, interval exchange transformations, generalized interval exchange transformations, linearization, Diophantine conditions, renormalization, Rauzy-Veech induction, Rauzy-Veech cocycle

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3576–3609
DOI 10.4171/ICM2022/143

1. INTRODUCTION

Flows on surfaces are among the most basic and fundamental examples of dynamical systems. First of all, they are among the lowest possible dimensional smooth systems; furthermore, many models of systems of physical origin are described by flows on surfaces, starting from celestial mechanics, up to solid state physics or statistical mechanics models. The beginning of the study of surface flows can be dated back to Poincaré [63] at the end of the 19th century, and coincides with the birth of dynamical systems as a research field. Poincaré was in particular interested in the study of flows on *tori*, or surfaces of *genus one*. Several famous systems in physics lead naturally to the study of flows on surfaces of *higher genus*, which, in this survey, will mean genus $g \geq 2$. Examples include the Ehrenfest model in statistical mechanics (related to a linear flow on a translation surface of genus five), or the Novikov model in solid state physics, which is described by locally Hamiltonian flows, a class which will be one of the central themes of this survey (see Section 3.1).

There is a rich history of results on the topological and qualitative behavior of trajectories (see, for example, [60] and the references therein), as well as on the ergodic theory of certain well-studied classes of flows (for example, in genus one, in relation with KAM theory, see Section 2, and linear flows on translation surfaces, whose study is intertwined with Teichmüller dynamics, see Section 3). Many fundamental problems, though, in particular on the mathematical characterization of chaos (such as dynamical, spectral, and rigidity questions) in various natural classes of surface flows, in particular smooth flows preserving a smooth measure, were only recently understood and many others are still open (see Section 3.1).

One of the reasons for this late development is perhaps that, in order to investigate fine chaotic or rigidity properties of flows in higher genus, one needs to impose quite delicate assumptions on the behavior of orbits on different scales. To capture these multiscale features, the concept of *renormalization* plays a crucial role (see Section 4). In the case of genus one, the assumptions on the flow often take the form of *Diophantine conditions* or, more generally, of *arithmetic conditions* on the rotation number (see Section 5) and control how well the flow orbits are approximated by *periodic orbits*. The renormalization point of view on these conditions is that they can be described in terms of continued fraction theory and therefore studying the dynamics of the Gauss map, or, equivalently, geometrically, studying the geodesic flow on the modular surface, both of which are classically well understood.

In higher genus, on the other hand, one had to wait for the development of the rich and fruitful theory of renormalization in Teichmüller dynamics (see Section 4). This theory provides a renormalization framework (initially developed to study ergodic properties of rational billiards, interval exchange transformations, and translation flows), which can be exploited to understand when a surface flow is *renormalizable* (see Sections 3.2 and 4) and when it preserves a smooth invariant measure; in the latter case, then, it allows imposing conditions on a (smooth) surface flows to guarantee the presence of particular chaotic properties (see Section 3.1). The type and nature of what we refer to as *Diophantine-like conditions* in higher genus, which is much more delicate than in genus one and often involves assumptions on *hyperbolicity* of the renormalization, will be the leading theme of this survey. These conditions are sometimes also called *arithmetic conditions*, by analogy with the genus one case, even though the relation with classical arithmetic and Diophantine equations is lost when the genus is greater than one.

In what follows, we first start in the next Section 2 with the classical case of flows on *genus one surfaces*, recalling some of the classical results on the *linearization problem* and ergodic properties and discussing the related arithmetic conditions. Then, in Section 3, we will briefly overview some of the rapid developments in our understanding of ergodic, spectral, and disjointness properties of (smooth) area-preserving flows on higher genus surfaces (see Section 3.1), as well as linearization and rigidity problems in higher genus (in Section 3.2). After having introduced the notion of *renormalization* in this setting (see Section 4), we then focus in Section 5 on the Diophantine-like conditions behind these results.

2. FLOWS ON SURFACES OF GENUS ONE AND CLASSICAL ARITHMETIC CONDITIONS

A central idea introduced by Poincaré was that the study of a surface flow can be often *reduced* to the study of a one-dimensional discrete dynamical system, by taking what we nowadays call a *Poincaré* section and considering the *Poincaré* first return map of the flow to the section (when and where it is defined). If we start from a flow $\varphi_{\mathbb{R}} := (\varphi_t)_{t\in\mathbb{R}}$ on a torus, i.e., on a compact, orientable surface S of genus one, and assume that it does not have fixed points, or closed orbits (or, more generally, Reeb components, see [60]), there is a (global) section given by a closed transverse curve and the Poincaré first return map to it is a diffeomorphism $f : S^1 \to S^1$ of the circle $S^1 \cong \mathbb{R}/\mathbb{Z}$. The simplest example of *circle diffeomorphism* (or *circle diffeo* for short) is a (rigid) *rotation*, i.e., the map $R_\alpha(x) = x + \alpha \mod 1$ on $\mathbb{R}/\mathbb{Z} = [0,1]/\sim$. A key concept associated to circle diffeomorphisms is that of *rotation number*: if μ is an *invariant* probability *measure* for the circle diffeo f (which always exists by Krylov–Bogolyubov theorem), the rotation number $\rho(f)$ of f can be seen as an *average displacement* of points, namely $\rho(f) = \int_0^1 (F(x) - x)\,\mathrm{d}\mu(x) \mod 1$ where $F : \mathbb{R} \to \mathbb{R}$ is a lift of $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$. The rotation R_α can be seen as the *linear model* of a circle diffeo with rotation number α.

The *topological behavior* of trajectories of $(\varphi_t)_{t\in\mathbb{R}}$ can be completely understood and classified exploiting the *rotation number* (this is essentially the content of *Poincaré classification theorem*, see [36] for an expository account): when $\rho(f) \in \mathbb{Q}$, there exist periodic orbits (which either *foliate* the surface S, or are attracting or repelling). On the other hand, when $\rho(f) \notin \mathbb{Q}$, the dynamics of $(\varphi_t)_{t\in\mathbb{R}}$ is either *minimal* on the whole surface (i.e., all orbits are *dense*), or minimal when restricted to a *Cantor-like* invariant limit set (locally a product of a Cantor set with $\mathbb{R}$). In the latter case, we speak of *Denjoy-counterexamples*; their existence is ruled out when the diffeo (and the flow) is sufficiently smooth, for example, $\mathcal{C}^2$ in view of Denjoy's work [15] (less regularity, in particular $\mathcal{C}^1$ with bounded variation derivative, suffices, see, e.g., [36] for more details).

Arithmetic conditions for linearization of circle diffeomorphisms. To gain a finer understanding of the dynamics and describe the ergodic behavior of almost-every trajectory with respect to a smooth measure, one has to address the *linearization problem*, a classical question which is at the heart of the theory of circle diffeomorphisms. Namely, one wants to understand when a circle diffeomorphism T is *linearizable*, i.e., conjugate to a rigid rotation R_α (i.e., when there exists a homeomorphism $h : S^1 \to S^1$, called the *conjugacy*, such that $R_\alpha \circ h = h \circ T$) and what is the *regularity* of the conjugacy h. To address this question, one needs to put further assumptions both on the *regularity* of the diffeo and, in relation to it, the irrationality of the *rotation number*.

We recall that *arithmetic conditions* are conditions that prescribe how well (or how *badly*) the irrational rotation number $\alpha \in \mathbb{R}$ is approximated by *rational* numbers and morally control how well the flow orbits are approximated by *periodic orbits*. The best known such condition is perhaps the (classical) *Diophantine condition* (or DC, for short): $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ is said to be *Diophantine* (of exponent $\tau \geq 0$) iff there exists $C > 0$ such that

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{C}{q^{2+\tau}}, \quad \text{for all } p, q \in \mathbb{Z},\ q \neq 0.$$

If the above condition holds for $\tau = 0$, we say that α is *badly approximable* or *bounded-type*. Equivalently, the DC can be rephrased in terms of the continued fraction expansion $[a_0, a_1, \dots, a_n, \dots]$ of α: if q_n denotes the *convergents* of α, namely the denominators of the partial approximations $p_n/q_n := [a_0, a_1, \dots, a_n]$, the DC is equivalent to the growth control $a_{n+1} = O(q_n^\tau)$. In particular, α is of bounded type iff a_n are uniformly bounded.

The *local theory* of linearization of circle diffeos, which treats the case of diffeos $f : S^1 \to S^1$ which are $\mathscr{C}^\infty$*-close* (or analytically, or $\mathscr{C}^r$-close) to a circle rotation R_α, where $\alpha = \rho(f)$, is a rather classical application of KAM theory. The prototype result is the *local rigidity* theorem of Arnold [1], who showed that if α is Diophantine, circle diffeos which are a sufficiently small analytic deformations of R_α and have rotation number equal to α, must be *analytically* conjugate to R_α. Among the few *global results* (which do not assume that f is close to a rotation), we recall the celebrated theorem by Michael Herman [30] and Jean-Christophe Yoccoz [77], answering a question by Arnold, showing that if f is $\mathscr{C}^\infty$ (or analytic) and its rotation number $\rho(f)$ satisfies the DC, the conjugacy is $\mathscr{C}^\infty$ (resp. analytic). Furthermore, the DC turns out to be the optimal arithmetic condition for global smooth linearization. Another, more subtle arithmetic condition, called "*condition* H" in honor of Herman, was introduced by Yoccoz as the optimal condition for global *analytic* linearization of analytic diffeos, see [79].

Another famous arithmetic condition is the *Roth-type* condition, which is satisfied by irrationals $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ such that $a_n = O_\varepsilon(q_n^\varepsilon)$ for all $\varepsilon > 0$. A crucial step in the KAM approach developed by Arnold for circle diffeomorphisms is to solve a *linearized* version of the conjugacy equation $R_\alpha \circ h = h \circ T$, namely the *cohomological equation*: given a smooth $\phi : I \to \mathbb{R}$, one looks for a smooth solution $\varphi : I \to \mathbb{R}$ to the equation $\varphi \circ R_\alpha - \varphi = \phi$. The Roth-type condition turns out to be the optimal one needed to solve this cohomological equation with optimal loss of differentiability: for any $r > s + 1 \geq 1$, one can find a solution $\varphi \in \mathscr{C}^s$ for any $\phi \in \mathscr{C}^r$ as long as $\int \phi = 0$ (which is a trivial necessary condition) *if and only* *if* α is Roth-type: this equivalent characterization provides a remarkable connection between dynamical and arithmetical properties.

We remark that the Diophantine condition, the H condition, and the Roth-type condition can all be proved to have *full measure*, namely they hold for a set of $\alpha \in [0,1]$ of Lebesgue measure one (the set of badly approximable $\alpha \in [0,1]$, on the other hand, has Lebesgue measure *zero*, although full Hausdorff dimension). While full measure of the Diophantine and Roth conditions can be proved in an elementary way, it is an instructive exercise to derive it from the properties of the Gauss map $G : [0,1] \to [0,1]$ and of the Gauss invariant measure $\mathrm{d}x/\log 2(1+x)$, since this point of view can be applied to show full measure of other arithmetic conditions as well and it can be furthermore generalized to higher genus (see Sections 4 and 5).

In view of this remark, we conclude this section with a reinterpretation of Herman's linearization theorem in the language of *foliations* into flow trajectories. In this setting, the linear model of a flow on a torus is a *linear flow* on $\mathbb{R}^2/\mathbb{Z}^2$ (i.e., the flow which arises as solution of $(\dot{x_1}, \dot{x_2}) = (\theta_1, \theta_2)$, which moves points with unit speed along lines of slope θ_2/θ_1).

Theorem 2.1 (Reformulation of Herman's global theorem [30]). *For a full measure set of real numbers α, a foliation on a genus one surface which is topologically conjugate to the foliation given by a linear flow with rotation number α is also $\mathscr{C}^\infty$-conjugate to it.*

Since the regularity of a conjugacy between foliations, which sends leaves into leaves, is defined in terms of the transverse structure, this result is just a restatement of the result for the Poincaré maps of the two flows (which are circle diffeomorphisms and rotations respectively).

Ergodic properties in genus one and exceptional behavior. From the existence (and abundance) of smooth (or at least continuously *differentiable*, i.e., $\mathscr{C}^1$) linearizations, one can infer many of the smooth measure-theoretical ergodic properties of flows of genus one. In particular, one sees that, for a full measure set of rotation numbers, flows in genus one are *ergodic* (since irrational rotations are) with respect to a *smooth* invariant measure of full support (the $\mathscr{C}^1$-regularity of the conjugacy allows us indeed to *transport* the Lebesgue invariant measure to obtain the invariant measure for the diffeo, which in turns gives a *transverse measure* for the flow). Furthermore, they are *uniquely ergodic* (in view of Kronecker–Weyl theorem for rotations, e.g., [14]), i.e., this natural invariant measure is the *unique* invariant measure (up to scaling).

We remark that *exceptional* ergodic behaviors in genus one (smooth) surface flows, can be constructed for flows whose rotation numbers are irrational but not Diophantine, i.e., the so-called *Liouvillean* (rotation) numbers. When α is Liouville, exploiting the abundance of good rational approximations $(p_n/q_n)_n$ to α, for example, using the method of *periodic approximations* pioneered by Anosov and Katok and later revived by Fayad, Katok et al. (see [36] or the survey [18]), one can construct many examples with *pathological* behavior, for example, flows with a *singular* invariant measures and *time-reparametrizations* (also

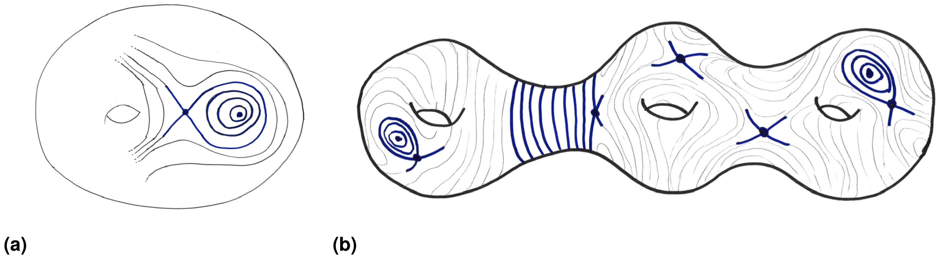


FIGURE 1
Pictorial representation of locally Hamiltonian flows on a surfaces: in (a) an Arnold flow ($g = 1$) and in (b) a flow in $g = 3$ with two minimal components and 3 periodic components.

called *time-changes*) which are weakly mixing or which have mixed spectrum (see [18] and the references therein).

Finally, before moving to higher genus, we remark that another possible way to introduce interesting dynamical features for *typical* rotation numbers is to consider flows on tori *with singularities*. The simplest type of singularity is a *stopping point*. Already such a simple perturbation, which is only a time-reparametrization of the flow, can lead to flows which are typically *mixing* (see [43]) and even to flows with *Lebesgue spectrum* (see [17]). Smooth measure preserving flows on a torus with one center and one simple saddle (see Figure 1a) were first studied by Arnold in [2] and constitute one of the most studied examples in the class of flows known as *locally Hamiltonian*: we return to them and to their typical ergodic properties in Section 3.1.

3. DYNAMICS OF FLOWS ON SURFACES OF HIGHER GENUS

Let us now consider the *higher genus* case, namely consider now a (smooth) flow $\varphi_{\mathbb{R}} := (\varphi_t)_{t\in\mathbb{R}}$ on a compact, connected orientable (closed) surface S of genus $g \geq 2$. Notice that in this case, by Euler characteristic restrictions, the flow *always* has *fixed points* (see Figure 2 for some examples). We require that singularities be *isolated* (so that in particular, by compactness, the set $\mathrm{Fix}(\varphi_{\mathbb{R}})$ of fixed points is *finite*).

Topological dynamics and quasiminimal sets. The *topological classification* of the possible behavior of trajectories of a flow on a surface (and, more generally, of surface *foliations* which are not necessarily orientable) has been a topic of research in the 20th century (starting from the 1930–1940s, up to the 1970s). In particular, through the works of Maier, Levitt, Gutierrez, Gardiner et al. (see [60] for references), one could obtain results on what possible *orbit closures* are, as well as a classification of *quasiminimal sets*, which can be defined as possible *ω-limit sets* of *nontrivial* recurrent trajectories, i.e., set of accumulation points of trajectories different from a fixed point or a closed, periodic orbit. Quasiminimal sets can be the whole surface, subsurfaces with boundary, or a Cantor-like invariant sets. Moreover, one

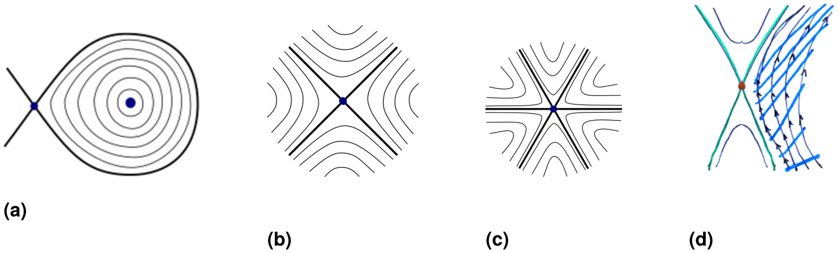


FIGURE 2
Type of singularities of a locally Hamiltonian flow: a center in (a), a simple saddle in (b) and a multisaddle in (c). Decelerations and shearing near a Hamiltonian saddle in (d).

can prove *decomposition theorems* showing that one can *cut* the surface S into subsurfaces each of which contains at most one quasiminimal set (see in particular the work by Levitt [49]). We do not enter here into the details of these topological results, but refer the interested reader, for example, to the monograph [60] and the references therein.

Interval exchanges and generalized IETs as Poincaré sections. As in the case of genus one, an essential tool to study a higher genus flow is to consider a (local) *transversal* $I \subset S$ to the flow and the *Poincaré first return map* T of the flow on I (when it is defined, for example, almost everywhere when the flow preserves a finite measure with full support; for more general situations, see [60]). Such first return maps $T : I \to I$ are one-to-one *piecewise diffeomorphisms* known as *generalized interval exchange transformations*: a map $T : I \to I$ is a generalized interval exchange transformations or, for short, a GIET, if one can partition I into intervals $I_1, \ldots, I_d$ (finitely many since we are assuming that $\varphi_{\mathbb{R}}$ has finitely many fixed points) so that the restriction T_i of T to I_i, for each $1 \le i \le d$, is a diffeomorphism onto its image which extends to a diffeo of the closure $\overline{I}_i$ (see, e.g., [55]). We say in this case that T is a d-GIET. We say, furthermore, that T is *of class* $\mathscr{C}^r$ if the restriction of T to each I_i extends to a $\mathscr{C}^r$-diffeomorphism onto the closed interval $\overline{I}_i$. The adjective *generalized* is used to distinguish them from the more commonly studied (standard) interval exchange transformations (or simply IETs), which are one-to-one piecewise *isometries*, namely GIETs such that the derivative T_i' of each branch is constant and equal to one.

Standard IETs are a generalization of circle rotations (since an IET is a rotation when $d = 2$) and play an analogous role in higher genus, providing the natural *linear model* of a GIET (see Section 3.2). Furthermore, as rotations are Poincaré maps of *linear flows* on the torus $\mathbb{R}^2/\mathbb{Z}^2$, IETs arise naturally as Poincaré maps of *linear flows* on *translation surfaces* (see the ICM proceeding [12] for an introduction to the latter).

3.1. Locally Hamiltonian flows

We will be mostly concerned with flows which preserve a (probability) measure μ of *full support*, for example, an area-form, since this is a natural setup for *ergodic theory*. Given a surface S with a fixed smooth area form ω, a *smooth area-preserving flow* $\varphi_{\mathbb{R}} = (\varphi_t)_{t\in\mathbb{R}}$ on S is a smooth flow on S which preserves the measure μ given integrating a smooth density with respect to ω. The interest in the study of these flows and, in particular, in their ergodic and mixing properties, was revived by Novikov [61] in the 1990s, in connection with problems arising in solid-state physics, as well as in pseudoperiodic topology (see, e.g., the survey [84] by A. Zorich). Smooth area-preserving flows are also called *locally Hamiltonian flows* or *multivalued Hamiltonian flows* in the literature, in view of their interpretation as flows locally given by Hamiltonian equations: one can find local coordinates (x_1, x_2) on each open set $U \subsetneq S$ in which $\varphi_{\mathbb{R}}$ is given by the solution to the equations:

$$\begin{cases} \dot{x_1} = \partial H/\partial x_2, \\ \dot{x_2} = -\partial H/\partial x_1, \end{cases}$$

where $H : U \to \mathbb{R}$ is a real-valued (*local*) Hamiltonian. For simplicity, we will assume here that H is infinitely differentiable, even though for several results $\mathscr{C}^3$ (or also $\mathscr{C}^{2+\varepsilon}$ for every $\varepsilon > 0$) suffices. It turns out that such smooth area-preserving flows on S are in one-to-one correspondence with smooth *closed* real-valued differential 1-forms: given such a 1-form η, we can associate to it the integral flow $\varphi^{\eta}_{\mathbb{R}}$ of the vector field X such that $\eta = i_X\omega$, where i_X denotes the contraction operator. Since η is closed, $\varphi^{\eta}_{\mathbb{R}}$ is area-preserving; conversely, every smooth area-preserving flow can be obtained in this way.

Topology and measure class. Let $\mathcal{F}$ denote the set of smooth closed 1-forms on S with isolated zeros. On $\mathcal{F}$ (which we can think of as the space of locally Hamiltonian flows) one can define a *topology* as well as a measure class. The *topology* is obtained by considering perturbations of closed smooth 1-forms by (small) closed smooth 1-forms. We will often restrict our attention to the subset $\mathcal{M} \subset \mathcal{F}$ of *Morse* closed 1-forms (i.e., forms which are locally the differential of a *Morse function*), which is *open and dense* in $\mathcal{F}$ with respect to this topology (see, e.g., [64]). Locally Hamiltonian flows corresponding to forms in $\mathcal{M}$ have only *nondegenerate fixed points*, i.e., *centers* and *simple saddles* (as in Figures 2a and 2b), as opposed to degenerate *multisaddles* (as in Figure 2c). Furthermore, if $\mathcal{F}_{s,l}$ denote the flows which correspond to flows in $\mathcal{M}$ with s saddle points and l centers, each $\mathcal{F}_{s,l}$ is open and their union is dense in $\mathcal{F}$ (see [64]).

A *measure-theoretical notion of typical* can be defined on each $\mathcal{F}_{s,l}$ using the *Katok fundamental class* (introduced by Katok in [35], see also [60]), i.e., the cohomology class of the 1-form η which defines the flow. Let $\mathrm{Fix}(\varphi_{\mathbb{R}})$ denote the set of *fixed points* (also called *singularities*) of the flow $\varphi_{\mathbb{R}}$ and let $k = s + l$ be its cardinality (recall that it is finite since the flow is in $\mathcal{F}$ and $k \geq 1$ when $g \geq 2$). If we fix a base $\gamma_1, \dots, \gamma_n$ of the relative homology $H_1(S, \mathrm{Fix}(\varphi_{\mathbb{R}}), \mathbb{R})$ (where $n = 2g + k - 1 = 2g + s + l - 1$) and consider the period map Per given by $\mathrm{Per}(\eta) = (\int_{\gamma_1} \eta, \dots, \int_{\gamma_n} \eta) \in \mathbb{R}^n$, we say that a property holds for a *typical* locally Hamiltonian flow in $\mathcal{F}_{s,l}$ if it holds for all η such that $\mathrm{Per}(\eta)$ belongs to a full measure set with respect to the Lebesgue measure on $\mathbb{R}^n$.

Minimal components and ergodicity. To describe (typical) chaotic behavior in locally Hamiltonian flows, it is crucial to distinguish between two open sets (complementary, up to measure zero, see [75] or [64] for more details): in the first open set, which we will denote by $\mathcal{U}_{\min}$, the typical flow is *minimal* (the term *quasiminimal* is also used in the literature), in the sense that the orbits of all points which are not fixed points are *dense* in S; flows in $\mathcal{U}_{\min}$ have only saddles, since the presence of centers prevents minimality. On the other open set, that we call $\mathcal{U}_{\neg\min}$, the flow is not minimal (there are saddle loops homologous to zero which disconnect the surface), but one can decompose the surface into a finite number of subsurfaces with boundary S_i, $i = 1, \dots, N$ such that for each i either S_i is a *periodic component*, i.e., the interior of S_i if foliated into closed orbits of $\varphi_{\mathbb{R}}$ (in Figure 1b one can see three periodic components, namely two disks and one cylinder), or S_i is such that the restriction of $\varphi_{\mathbb{R}}$ to S_i is minimal in the sense above, as pictured in the remaining two subsurfaces in Figure 1b. These are called *minimal components* and there are at most g of them (where g is the genus of S), see Section 3.1.

Notice that minimality and ergodicity of a (minimal component of a) locally Hamiltonian flow are equivalent to minimality or respectively ergodicity of an (and hence any) interval exchange transformation which appears as the Poincaré map. Classical results proved in the 1980s guarantee that almost every IET (with respect to the Lebesgue measure on the interval lengths, assuming that the permutation is irreducible) is minimal (as showed by Keane [37], see also [35]) and (uniquely) ergodic (as proved in the works by Masur [50] and Veech [76], considered early milestones of the successful application of Teichmüller dynamics to the study of IETs and translation surfaces, see the ICM proceeding [12] or the survey [85]). It then follows from definition of Katok measure class that a typical local Hamiltonian flow in $\mathcal{U}_{\min}$ is minimal and ergodic and, given a typical local Hamiltonian flow in $\mathcal{U}_{\neg\min}$, its restriction on each minimal component is ergodic.

Classification of mixing properties. Finer chaotic features of locally Hamiltonian flows, in particular mixing and spectral properties, change according to the type of singularities and depend crucially on the locally Hamiltonian parametrization of saddle points. For a (nongeneric) locally Hamiltonian flow with at least one *degenerate saddle* (an example of such a saddle is shown in Figure 2c), *mixing* (for the definition, see (3.1) with $n = 2$) was proved in the 1970s by Kochergin [43]. When, on the other hand, $\eta \in \mathcal{M}$ is a Morse 1-form, so that all saddles are *simple*, one has a dichotomy: inside the open set $\mathcal{U}_{\min}$ in which the typical flow is minimal, almost every locally Hamiltonian flow is *weakly mixing*, but it is *not mixing*; both results follow from work by the author [72,73]. On the other hand, for a full measure set of flows in $\mathcal{U}_{\neg\min}$, the restriction to each of minimal components is mixing (as proved by Ravotti [64] extending the previous work [71] by the author).

The question of mixing in higher genus was raised by V. Arnold in the 1990s, when he conjectured (see [2]) that the restriction of a typical smooth flow on a torus with one center and one simple saddle to its minimal component (namely for what we nowadays call an *Arnold flow*) was indeed mixing. His conjecture was proved shortly after by Khanin and Sinai in [67], who showed mixing under the assumption that the rotation number α is such that the entries a_n of the continued fraction expansion of α do not grow too fast, namely there exist a power $1 < \tau < 2$ and $C > 0$ such that $|a_n| \leq C n^{\tau}$. One can show (for example, exploiting the

Gauss map G and the finiteness of $\int_0^1 a_0(x)\,\mathrm{d}\,\mu_G(x)$, where μ_G is the Gauss measure, via a standard Borel–Cantelli argument) that this arithmetic condition holds for a full measure set of α. The condition was later improved by Kocerghin, see [44]. Also in the case of absence of mixing, a prototype result for flows over a full measure set of rotation numbers was proved by Kochergin [42] already in the 1970s (and much more recently extended in [45] to all irrational rotation numbers), much earlier than results in higher genus [65,70,73].

In higher genus, the above mentioned results on mixing/absence of mixing require the introduction of Diophantine-like conditions, which describe the full measure set of locally Hamiltonian flows for which the results hold. In [71], for example, we introduced a condition on a IET (see Section 5 for more details) called *Mixing Diophantine Condition* (or MDC, for short). Let us say that the restriction of a locally Hamiltonian flow $\varphi_{\mathbb{R}}$ to one of its minimal components S_i satisfies the MDC if one can find a section $I \subset S_i$ (in *good position* in the sense of [55]) such that the IET which arises as Poincaré map of $\varphi_{\mathbb{R}}$ to I satisfies the MDC. One can then prove:

Theorem 3.1 (Ulcigrai [71], Ravotti [64]). *Let $\varphi_{\mathbb{R}}$ be a flow in $\mathcal{U}_{\neg\min}$ and let S_i be a minimal component. If the restriction of $\varphi_{\mathbb{R}}$ to S_i satisfies the Mixing Diophantine Condition, then $\varphi_{\mathbb{R}}$ restricted to S_i is mixing.*

We then show in [71] (exploiting results from [3], see Section 5) that the MDC is satisfied by a full measure set of IETs. Similarly, to prove that a typical flow in $\mathcal{U}_{\min}$ is *not* mixing, a Diophantine-like condition is introduced and proved to be of full measure in [73]. Special cases of the absence of mixing result for surfaces with $g = 2$ and two isometric saddles were proved in [70] and by Scheglov in [65]. We remark that in $\mathcal{U}_{\min}$ there exist, nevertheless, exceptional mixing flows, as shown by Chaika and Wright in [13], who produced sporadic examples in $g = 5$.

Parabolic dynamics and slow chaos. Smooth area-preserving flows on surfaces also provide one of the fundamental classes of *parabolic*, or *slowly chaotic*, dynamical systems (see, e.g., the survey [75]). In systems which display *sensitive dependence* on initial conditions (the so-called *butterfly effect*), one can find many nearby initial conditions whose trajectories *diverge* with time. Contrary to hyperbolic systems, where this divergence happens (infinitesimally) at *exponential speed*, in parabolic systems the divergence speed is *slow*, namely subexponential, and in all known examples *polynomial* or subpolynomial. Slow divergence in locally Hamiltonian flows is created by Hamiltonian saddles, which create different deceleration rates of nearby trajectories and produce a form of (local) *shearing*, by *tilting* in the flow direction the image under the flow of arcs initially transverse to the dynamics, as illustrated in Figure 2d. Shearing happens not only locally, near a saddle, but globally for typical flows in $\mathcal{U}_{\neg\min}$, which (in view of the presence of saddle loops) display a global *asymmetry* in the prevalent direction of shearing. It is this geometric mechanism which is behind the proof of mixing (in this setting, but also for many other classes of parabolic flows, see the survey [74] and the references therein). Under the assumption that the restriction of $\varphi_{\mathbb{R}}$ to a minimal component S_i satisfies the Mixing Diophantine Condition, one can produce quantitative estimates on shearing of transverse arcs and, as shown by Ravotti in [64], prove quantitative mixing estimates, which show that mixing happens (at least) at *subpolynomial speed*, i.e., for any two smooth observables $f, g : S_i \to \mathbb{R}$ supported outside the saddles in $\mathrm{Fix}(\varphi_{\mathbb{R}}) \cap S_i$,

$$\left| \int_{S_i} f\big(\varphi_t(x)\big) g(x) \mathrm{d}\mu - \int_{S_i} f \mathrm{d}\mu \int_{S_i} f \mathrm{d}\mu \right| \leq \frac{C_{f,g}}{(\log t)^{\gamma}}, \quad t \geq 0.$$

This is expected to be also the optimal nature of the estimates, namely the decay is *not* expected to be polynomial or faster in this setting, but no lower bounds on the decay of correlations are currently available.

Ratner's forms of shearing. Striking consequences of shearing (such as measure and joining rigidity) were proved for another famous class of parabolic flows, namely horocycle flows on hyperbolic surfaces and their time-changes, by exploiting a *quantitative shearing* property introduced by Marina Ratner and nowadays known as *Ratner property* (or RP). In view of its importance, in the study of horocycle flows and, more generally, unipotent flows in homogeneous dynamics, it is natural to ask whether this property can be proved and exploited in other parabolic (non homogeneous) settings. For locally Hamiltonian flows, which are natural candidates, the original Ratner property is believed to fail due to the presence of singularities (see [16]). Nevertheless, a variant of the RP which has the same dynamical consequences, called *Switchable Ratner Property* (or SRP, for short), was introduced by B. Fayad and A. Kanigowski [16] and showed to hold for typical Arnold flows (as well as some flows in genus one with one degenerate singularity). As an abstract consequence of the SRP property, one can conclude that typical Arnold flows are not only mixing, but *mixing of all orders*, namely for any $n \geq 2$ and any n-tuple $A_0, \ldots, A_{n-1}$ of measurable sets,

$$\mu\big(A_0 \cap \varphi_{t_1}(A_1) \cap \cdots \cap \varphi_{t_1+\cdots+t_{n-1}}(A_{n-1})\big) \xrightarrow{t_1,t_2,\ldots,t_{n-1}\to\infty} \mu(A_0)\cdots\mu(A_{n-1}). \tag{3.1}$$

Notice that this definition reduces to the classical definition of mixing in the special case $n = 2$; whether mixing implies mixing of all orders in general is still an open problem, known as *Rohlin conjecture*.

To prove the SRP property, one needs to assume that the rotation number $\alpha = [a_0, a_1, \ldots, a_n, \ldots]$ satisfies an ad hoc arithmetic condition, namely, if q_n are the denominators of α, one requires that, for some $0 < \xi, \eta < 1$ (taken to be $\xi = \eta = 7/8$ in [16]) the following series is finite:

$$\sum_{k \notin K(\alpha)} \frac{1}{(\log q_n)^{\eta}} < +\infty, \quad \text{where } K(\alpha) := \big\{k \in \mathbb{N},\ a_{k+1} \leq C(\log q_k)^{\xi}\big\}. \tag{3.2}$$

In a joint work with A. Kanigowski and J. Kułaga-Przymus [33], we were able to generalize this result to higher genus. To do so, it is once again crucial to introduce a suitable Diophantine-like condition, which we called in [33] the *Ratner Diophantine Condition* (or RDC) and we describe in Section 5. The main result we prove is the following.

Theorem 3.2 (Kanigowski, Kułaga-Przymus, Ulcigrai [33]). *If the restriction of $\varphi_{\mathbb{R}} \in \mathcal{U}_{\neg\min}$ to a minimal component S_i satisfies the Ratner Diophantine Condition, $\varphi_{\mathbb{R}} : S_i \to S_i$ satisfies the Switchable Ratner Property and is mixing of all orders.*

We then show that the RDC is satisfied by almost every IET and therefore can conclude that, for a full measure set of locally Hamiltonian flows in $\mathcal{U}_{\neg\min}$, each restriction to a minimal component is mixing of all orders.

Quantitative estimates on slow, Ratner-type shearing were recently used (in the joint work [34] with A. Kanigowski and M. Lemańczyk) to study *disjointness of rescalings*, a property that has recently received a revival of attention in view of its role as possible tool to prove Sarnak Möbius orthogonality conjecture (see the ICM proceedings survey [48] and the references therein). In [34] we introduce a disjointness criterium based on Ratner shearing and use it (as one of the applications) to show that, in genus one, typical Arnold flows have *disjoint rescalings* and satisfy Moebius orthogonality. Disjointness of rescalings seems to be an important feature of parabolic dynamics: while specific parabolic flows may fail to be disjoint from their rescalings (primarily the horocyle flow on a hyperbolic surface), several recent results seem to indicate that this property is indeed widespread among parabolic flows (see, e.g., the results in [34] on time-changes of horocycle flows). In the context of surface flows, disjointness of rescalings has been verified in [4] for *von Neumann flows* (which can be realized as translation flows on surfaces with boundary). Whether one can extend the disjointess result proved in [34] for Arnold flows to higher genus smooth flows, remains an open problem and is likely to require a delicate control of Diophantine-like properties.

Polynomial deviations of ergodic averages. Slow chaotic behavior manifests itself not only through *slow* mixing, but also through *slow* convergence of ergodic integrals: given an ergodic area-preserving flow $\varphi_\mathbb{R}$ (or its restriction to an ergodic minimal component $S' \subset S$) and a real valued observable f with zero-mean, the ergodic integrals $I_T(f,x) := \int_0^T f(\varphi_t(x))\,\mathrm{d}t$ decay to zero *polynomially* with some exponent $0 < \nu < 1$ for almost every initial point, i.e., $|I_T(f,x)| \sim O(T^\nu)$ in the sense that

$$\limsup_{T\to\infty} \frac{\log|I_T(f,x)|}{\log T} = \nu.$$

This phenomenon, known as *polynomial deviations of ergodic averages*, was discovered experimentally in the 1990s by A. Zorich and explained (for linear flows on translation surfaces and observables corresponding to cohomology classes) in seminal work by Kontsevitch and Zorich [46,83] relating power deviations to Lyapunov exponents of renormalization (see Section 4). Forni in [23] could extend this result to integrals of sufficiently regular functions over translation flows and show that ergodic integrals can display a *power spectrum* of behaviors, i.e., there are exactly g positive exponents $0 < \nu_g \leq \cdots \leq \nu_2 < \nu_1 := 1$ (which correspond to the positive Lyapunov exponents of renormalization) and for each a subspace of finite codimension of smooth observables that present polynomial deviations as above with exponent $\nu = \nu_i$. A finer analysis of the behavior of Birkhoff sums or integrals, beyond the *size* of oscillations, appears in the works [7,54]: Bufetov in [7] shows in particular that (for *typical* translation flows and sufficiently regular observables) the *asymptotic behavior* of ergodic integrals can be described in terms of g (where g is the genus of the surface) *cocycles* $\Phi_i(t,x)$, $1 \leq i \leq g$ (also called *Bufetov functionals*): each $\Phi_i : \mathbb{R} \times S' \to \mathbb{R}$ is a cocycle over the flow $\varphi_\mathbb{R}$ (in the sense that $\Phi_i(t+s,x) = \Phi_i(t,x) + \Phi_i(s,\varphi_t(x))$ for any $x \in S'$

and $t \in \mathbb{R}$), $\Phi_1(T,x) \equiv T$ and each Φ_i has power deviations $|\Phi_i(T,x)| \sim O(T^{\nu_i})$ with exponent ν_i. Together, the cocycles encode the *asymptotic behavior* of the ergodic integrals up to subpolynomial behavior, in the sense that, for some constants $c_i = c_i(f)$,

$$\int_0^T f\big(\varphi_t(x)\big)\mathrm{d}t = c_1 T + c_2\Phi_2(T,x) + \cdots + c_g\Phi_g(T,x) + \mathrm{Err}(f,T,x), \tag{3.3}$$

where for almost every $x \in S'$ the *error term* $\mathrm{Err}(f,T,p)$ is subpolynomial, i.e., for any $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that $|\mathrm{Err}(f,T,p)| \le C_\varepsilon T^\varepsilon$. In a joint work with Frączek, we recently gave a new proof of this result in [26], which extends the result to the setting of smooth observables over locally Hamiltonian flows with Morse singularities (in $\mathcal{U}_{\min}$ as well as in $\mathcal{U}_{\neg\min}$) and also shows that the set of locally Hamiltonian flows for which the result holds can be described in terms of a Diophantine-like condition. More precisely, we define in [26] the *Uniform Diophantine Condition* (or UDC, for short; see Section 5) and show that it has full measure. We then prove the following.

Theorem 3.3 (Frączek–Ulcigrai [26]). *If the restriction of the locally Hamiltonian flow $\varphi_\mathbb{R} \in \mathcal{M}$ on a minimal component S' satisfies the* Uniform Diophantine Condition*, for each $\mathscr{C}^3$ observable $f : S' \to \mathbb{R}$, there exist g exponents ν_i and corresponding cocycles Φ_i such that the expansion* (3.3) *holds.*

Spectral theory. The study of the spectrum of the unitary operators acting on $L^2(S,\mu)$ given by $f \mapsto f \circ \varphi_t$ can shed further light on the chaotic features of the dynamics of the flow $\varphi_\mathbb{R} := (\varphi_t)_{t\in\mathbb{R}}$ and is at the heart of the study of spectral theory of dynamical systems (see [48] or [75] and the references therein). While the classification of mixing properties of locally Hamiltonian flows is essentially complete, very little is known about their spectral properties beyond the case of genus one (and some sporadic examples, such as *Blokhin examples*, essentially built gluing genus one flows, see the work [25]). The recent result [17] by Fayad, Forni, and Kanigowski for genus one suggests that it may be possible to prove that the spectrum is countable Lebesgue also in higher genus when in presence of degenerate, sufficiently strong (multisaddle) singularities. In the nondegenerate case, though, we recently proved in joint work with Chaika, Frączek, and Kanigowski [10] that a *typical* locally Hamiltonian flow on a *genus two* surface with two isomorphic *simple saddles* has *purely singular* spectrum. This result does not use explicit Diophantine-like conditions, but rather geometry and, in particular, a special symmetry (the hyperelliptic involution) that surfaces in genus two are endowed with; Liouville-type Diophantine conditions are here imposed by requesting the presence on the surface of large flat cylinders close to the direction of the flow, whose existence for typical flows is then proved by a Borel–Cantelli-type of argument (see [10] for details). Extending this result beyond genus two, though, will probably require the use of Rauzy–Veech induction (see Section 4) and the introduction of new Diophantine-like conditions, which impose some controlled form of degeneration. The nature of the spectrum of minimal components of locally Hamiltonian flows in $\mathcal{U}_{\neg\min}$ (even in genus one, i.e., for Arnold flows) is a completely open problem.

3.2. Linearization and rigidity in higher genus

A different line of problems in which Diophantine-like conditions in higher genus play a crucial role are conjectures concerning *linearization* and *rigidity* properties of higher genus flows and their Poincaré sections, GIETs (defined in Section 3). In analogy with the case of circle diffeos, we say that a GIET T is *linearizable* if it is topologically conjugate to a linear model, namely to a (standard) IET T_0.

Topological conjugacy and wandering intervals. To generalize Poincaré and Denjoy work, one needs first of all a combinatorial invariant which extends the notion of rotation number. Such an invariant can be constructed by recording the combinatorial data of a renormalization process, as we explain in Section 4. One of the crucial differences between GIETs and circle diffeomorphisms, though, is the failure of a generalization of Denjoy theorem: there are smooth GIETs that are semiconjugate to a minimal IET for which the semiconjugacy is *not* a conjugacy; in other words, they have wandering intervals (see the examples found in [6,8] in the class of periodic-type (affine) IETs and, more generally, [54]). It is important to stress that this is *not* a low-regularity phenomenon, nor is it related to special arithmetic assumptions: as shown by the key work [54] by Marmi, Moussa, and Yoccoz, wandering intervals exist even for piecewise *affine* (hence analytic) GIETs (called AIETs), for almost every topological conjugacy class. The presence of wandering intervals is on the contrary expected to be *typical* (see, e.g., the conjectures in [27,55]) and it is closely interknit with the absence of a *Denjoy Koksma inequality* and, more generally, *a priori bounds* for renormalization, see [29].

Local obstructions to linearization. As an important first step towards local linearization, we already mentioned the *cohomological equation* $\varphi \circ T - \varphi = \phi$ in Section 2, where $T = R_\alpha$ was a rotation. Whether the cohomological equation could be solved when T is an IET, under suitable assumptions, was unknown until the pioneering work of Forni [21], who brought to light the existence of a *finite* number of obstructions to the existence of a (piecewise finite differentiable) solution. We remark that obstructions to solve the cohomological equation have been since then discovered to be a characteristic phenomenon in *parabolic dynamics* (e.g., their existence have been proved by Flaminio and Forni for horocycle flows [19] and nilflows on nilmanifolds [20], see also the ICM talk [22]). Forni's work is a breakthrough that paved the way for the development of a linearization theory in higher genus.

Another breakthrough, which put the stress on the *arithmetic* aspect of linearization in higher genus, was achieved by Marmi–Moussa–Yoccoz in their work [55] (and related works [53, 57]). In [53], in particular, they reproved and extended Forni's result using the IETs renormalization described in Section 5 and introduced the *Roth-type* condition (see also Section 5), as an explicit Diophantine-like condition on the IET needed to solve the cohomological equation $\varphi \circ T - \varphi = \phi - \xi$, where ξ is a piecewise constant function which embodies the finite-dimensional *obstructions*. This result, combined with a generalization of Herman's *Schwarzian derivative trick*, then led to the proof in [55] by the same authors that, for any $r \geq 2$, the $\mathscr{C}^r$ *local conjugacy class* of almost every IET T (more precisely, of any T of *restricted* Roth-type, see Section 5) is a *submanifold of finite codimension*. Marmi, Moussa, and Yoccoz also conjectured that for $r = 1$ it is a submanifold of codimension $(d-1)+(g-1)$, where d is the number of exchanged intervals and g the genus of the surface of which T is a Poincaré section. For the measure zero class of IETs of *hyperbolic periodic type* (see Section 5), this conjecture has recently been proved by Ghazouani in [28]. The proof of this result for almost every IET will require the introduction of a new suitable Diophantine-like condition on IETs.

Rigidity of GIETs. We say that a class of (dynamical) systems is *geometrically rigid* (or also $\mathscr{C}^1$-rigid), if the existence of a topological conjugacy between two objects in the class automatically imply that the conjugacy is actually $\mathscr{C}^1$. The global linearization results by Herman and Yoccoz recalled in Section 2 shows that the class of (smooth, or at least $\mathscr{C}^3$) circle diffeomorphisms with Diophantine rotation number is geometrically rigid (and actually $\mathscr{C}^\infty$-rigid, i.e., if a smooth circle diffeo is conjugated via a homeomorphism h to R_α with α satisfying the DC, then h is $\mathscr{C}^\infty$). We already saw that this can be reinterpreted as a rigidity result for flows on surfaces of genus one (see Theorem 2.1). In joint work with S. Ghazouani, we recently proved a generalization of this result to genus two.

Theorem 3.4 (Ghazouani, Ulcigrai [29]). *Under a full measure Diophantine-like condition, a foliation on a* genus two *surface which is topologically conjugate to the foliation given by a linear flow with Morse saddles is also $\mathscr{C}^1$ conjugate to it.*

Here full measure refers to the Katok measure class on the linear flow models (see the definition given earlier in this section). For simplicity, we stated the result for flows with simple, Morse-type saddles; degenerate saddles can also be considered, but then one has to further assume that the foliations are *locally* $\mathscr{C}^1$ conjugated in a neighborhood of the multisaddle. Both these results can be reformulated at level of Poincaré sections: we introduce more precisely a rather subtle Diophantine-like conditions on (irreducible) IETs of any number of intervals $d \geq 2$, that we call the *Regular Diophantine Condition*, or RDC (we comment on it in Section 3) and show that it is satisfied by almost every (irreducible) IET on d. We then prove:

Theorem 3.5 (Ghazouani, Ulcigrai [29]). *If an irreducible d-IET T_0 with $d = 4$ or $d = 5$ satisfies the RDC, then any $\mathscr{C}^3$-generalized interval exchange map T which is topologically conjugate to T_0, and whose* boundary $B(T)$ vanishes*, is actually conjugated to T_0 via a $\mathscr{C}^1$ diffeomorphism.*

The *boundary operator* $B(T)$ which appears in this statement is a $\mathscr{C}^1$-conjugacy invariant introduced in [55]; it encodes the holonomy at singular points of the surface of which T is a Poincaré section. Requesting that $B(T)$ vanishes is therefore a necessary condition for the existence of a conjugacy of class $\mathscr{C}^1$. Theorem 3.5 solves for $d = 4, 5$ one of the open problems suggested by Marmi, Moussa, and Yoccoz in [55], where they conjecture the result to hold also for any other larger d. The result which is missing to prove the conjecture in

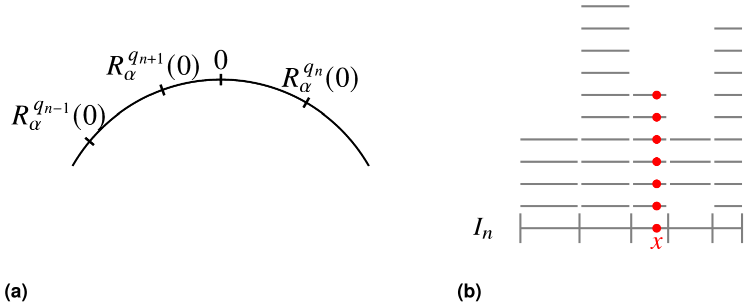


FIGURE 3
Renormalization algorithms for rotations and IETs.

its generality is a generalization of an estimate used in [54] to show existence of wandering intervals in affine IETs. The main result in [29], on the other hand (namely a dynamical dichotomy for the orbit of T under renormalization) is already proved for IETs which satisfy the RDC for any $d \geq 2$.

4. RENORMALIZATION AND COCYCLES

In this section we introduce the renormalization dynamics which is used as main tool to impose Diophantine-like conditions in higher genus. Renormalization in dynamics is a powerful tool to study dynamical systems which present forms of self-similarity (exact or approximate) at different scales. A map $T : I \to I$ of the unit interval which is (infinitely) *renormalizable* is such that one can find a (infinite) sequence of nested subintervals $I_{n+1} \subset I_n \subset \cdots \subset I$ such that the *induced dynamics* $T_n : I_n \to I_n$ (obtained by considering the first return map of T on I_n) is well defined and, up to *rescaling*, belongs to the same class of dynamical systems of the original T. Here, the rescaling, which is done so that the *rescaled* (or *renormalized*) map acts again on an interval of unit length, is given by the map $x \mapsto T_n(|I_n|x)/|I_n|$. We will now describe renormalization in the context of rotations first and then IETs. In both cases, at the level of (minimal) flows (or equivalently orientable foliations) on surfaces, the inducing process corresponds to taking shorter and shorter Poincaré sections of a given surface flow (on the torus or on a higher genus surface).

Renormalization algorithms. If $T = R_\alpha$ is a rotation by an irrational α and $q_n, n \in \mathbb{N}$, are the denominators of the convergents p_n/q_n of α, then one can consider as sequence $(I_n)_{n\in\mathbb{N}}$ the shrinking arcs on S^1 which have as endpoints $R_\alpha^{q_n}(0)$ and $R_\alpha^{q_{n+1}}(0)$. These endpoints correspond dynamically to consecutive closest returns of the orbit of 0 (see Figure 3a). The induced map T_n is then again a rotation R_{α_n}, with rotation number $\alpha_n = \mathcal{G}^n(\alpha)$, where $\mathcal{G}$ is the Gauss map $\mathcal{G}(x) = \{1/x\}$ and $\{\cdot\}$ denotes the fractional part.

Similarly, for a d-IET T, one wants to choose the nested sequence $(I_n)_{n\in\mathbb{N}}$ of inducing intervals so that the induced maps T_n are all IETs of the same number d of subintervals. Given any minimal T (or more generally any IET satisfying the *Keane condition* [37], i.e., such that the orbits of its discontinuity points are infinite and distinct), classical algorithms which produce such an infinite sequence $(I_n)_{n\in\mathbb{N}}$ are the *Rauzy–Veech induction algorithm* (see Veech [76] or [81] and the references therein) and *Zorich induction*, an acceleration of the same algorithm introduced by Zorich in [82]. For the definitions of these algorithms, which we will not use in the following, we refer the interested reader to the lecture notes [81]. One can show that, for $d = 2$, Zorich induction corresponds to the renormalization of rotations given by the Gauss map.

On the parameter space $\mathcal{I}_d$ of all d-IETs, these algorithms induce renormalization operators $\mathcal{R} : \mathcal{I}_d \to \mathcal{I}_d$, which associate to T the d-IET $\mathcal{R}(T)$ obtained by applying one step of the corresponding induction and then renormalizing the induced map to act on $[0, 1]$. Veech showed that Rauzy–Veech renormalization admits a conservative absolutely continuous invariant measure, that induces a *finite* invariant measure for the *Zorich acceleration*, as proved in [82]. The ergodic properties of the renormalization dynamics in parameter space have been intensively studied and are by now well understood, see, e.g., [80] and the references therein for a brief survey.

Rohlin towers and matrices. After n steps of induction, one can recover the original dynamics through the notion of *Rohlin towers* as follows: if I_n^i is one of the subintervals of T_n and $r := r_n^i$ is its first return time to I_n under the action of T, the intervals $I_n^i, T(I_n^i), \dots, T^{r-1}(I_n^i)$ are disjoint. Their union is called a *Rohlin tower* of step n and each of them is called a *floor* (see Figure 3b for a graphical depiction of floors and towers). Given an infinitely renormalizable T, for any n one can see $[0, 1]$ as a union of d Rohlin towers of step n, as shown in Figure 3b. Rohlin towers thus produce a sequence of *partitions* of $[0, 1]$ (into floors of towers of step n).

Renormalization produces also a sequence of $d \times d$ matrices A_n, $n \in \mathbb{N}$, with integer entries, which should be thought of as *multidimensional continued fraction* digits and describe *intersection numbers* of Rohlin towers. The matrices $(A_n)_{n\in\mathbb{N}}$ are defined so that the entries of the product $A^n := A_n \cdots A_1$ have the following dynamical meaning: the (i, j) entry $(A^n)_{ij}$ is the number of visits of the orbit of any point $x \in I_n^j$ to the initial subinterval I_0^i until its first return time r_n^j; in other words, $(A^n)_{ij}$ is the number of floors of the jth tower of level n which are contained in I_0^i. These entries generalize the classical continued fraction digits: for $d = 2$, indeed, the matrices $(A_n)_{n\in\mathbb{N}}$ associated to R_α, for n of alternate parity, have respectively the form

$$\begin{pmatrix} 1 & a_n \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 0 \\ a_n & 1 \end{pmatrix},$$

where a_n are the entries of the continued fraction expansion $\alpha = [a_0, a_1, \dots, a_n, \dots]$. Diophantine-like conditions for IETs are defined by imposing conditions on these matrices, on their growth as well as on their hyperbolicity, see in Section 5. The matrices $(A_n)_{n\in\mathbb{N}}$ are produced by the renormalization dynamics: for rotations, the entries $(a_n)_{n\in\mathbb{N}}$ of the continued fraction expansion of α satisfy $a_n = a(\mathcal{G}^n(\alpha))$, where $a(\cdot)$ is an integer-valued function on $[0,1]$. Similarly, one has now that $A_n = A(\mathcal{R}^n(T))$, where $A : \mathcal{I}_d \to \mathrm{SL}(d,\mathbb{Z})$ is a matrix-valued function on the space $\mathcal{I}_d$ of d-IETs, i.e., a *cocycle* (known as the *Rauzy–Veech cocycle*, or *Zorich cocycle* if considering the Zorich acceleration).

Positive and balanced accelerations. It turns out though that Zorich acceleration is often not sufficient (see, for example, [41] and [40] where it is shown that the classical Diophantine notions of bounded- [41] and Diophantine-type [40] do not generalize naturally when using Zorich acceleration). Two accelerations which play a key role in Diophantine-like conditions are the *positive* and the *balanced acceleration*. By *accelerations* we mean here an induction which is obtained by considering only a subsequence $(n_k)_{k\in\mathbb{N}}$ of Rauzy–Veech times. The associated (accelerated) cocycle is then obtained considering products

$$A(n_k, n_{k+1}) := A_{n_{k+1}-1} \cdots A_{n_k+1} A_{n_k}.$$

The *positive acceleration* appears in the works by Marmi, Moussa, and Yoccoz [53,55,57]. They showed that if T satisfies the Keane condition, for any n there exists $m > n$ such that $A(n,m)$ is a strictly positive matrix. The accelerated algorithm then corresponds to choosing the sequence $(n_k)_{k\in\mathbb{N}}$ setting $n_0 := 0$ and then, for $k \geq 1$, choosing n_k to be the smallest integer $n > n_{k-1}$ such that $A(n_{k-1}, n)$ is strictly positive. On the other hand, to define the *balanced acceleration*, one considers a subsequence $(n_k)_{k\in\mathbb{N}}$ of Rauzy–Veech times n for which the corresponding Rohlin towers are *balanced*, in the sense that ratios of widths $|I_n^i|/|I_n^j|$ and heights r_n^i/r_n^j are uniformly bounded above and below. We will return to these accelerations and some instances in which they are helpful in Section 5.

Combinatorial rotation numbers. We remark that the definition of Rauzy–Veech induction can be extended also to a GIET T (under the Keane condition, which guarantees that $\mathcal{R}^n(T)$ can be defined for every $n \in \mathbb{N}$) and then exploited to give a combinatorial notion of *rotation number* as well as a definition of *irrationality* in higher genus (following [55,57], see also [81]). As one computes the induced maps $(T_n)_{n\in\mathbb{N}}$, one can indeed record the sequence $(\pi_n)_{n\in\mathbb{N}}$ of permutations of the GIETs $(T_n)_{n\in\mathbb{N}}$: this sequence provides the desired *combinatorial rotation number* for $d > 2$. We say that a GIET is *irrational* if the sequence of matrices $(A_n)_{n\in\mathbb{N}}$ have a positive acceleration (or equivalently, in the terminology introduced by Marmi, Moussa, and Yoccoz, the path described by $(\pi_n)_{n\in\mathbb{N}}$ is *infinitely complete*). One can then show that two *irrational* GIETs with the same rotation number are *semiconjugated* (see, e.g., [81]), a result that generalizes a property of rotations numbers and circle diffeos and hence explains the choice of calling this higher genus combinatorial object the "*rotation number*" of a GIET.

Renormalization of Birkhoff sums. Given $T : I \to I$ and a function $f : I \to \mathbb{R}$, we denote by $S_n f := \sum_{k=0}^{n-1} f \circ T^k$ the nth Birkhoff sum (of the function f under the action of T). When $T = R_\alpha$ is a rotation (or a circle diffeo), it is standard to study first Birkhoff sums of the form $S_{q_n} f$ for q_n convergent of α, corresponding to closest returns, and then use them to *decompose* more general Birkhoff sums. Similarly, renormalization for (G)IETs can be exploited to produce *special Birkhoff sums*, namely Birkhoff sums of a special form that can be understood first, exploiting renormalization, and then used to decompose and study general Birkhoff sums. For each $n \in \mathbb{N}$, if $T_n : I_n \to I_n$ is the induced map after n steps of renormalization, the nth *special Birkhoff sum* is the induced function $S(n)f : I_n \to I_n$, defined by $S(n)f(x) := S_{r_n^i} f(x)$ if $x \in I_n^i$. Thus, since r_n^i is the height of the Rohlin tower over I_n^i, the value $S(n)f(x)$ is obtained summing the orbit *along the tower* which has x in the base, see Figure 3b. Notice that for $d = 2$, when considering Zorich acceleration, these reduce to sums of the form $S_{q_n} f(x)$. The associated *special Birkhoff sums operators* $S(n)$, $n \in \mathbb{N}$, map $\mathrm{f} : I \to \mathbb{R}$ to $S(n)f : I_n \to \mathbb{R}$. When f is piecewise constant and takes a constant value f^i on each I^i, $S(n)$ can be identified with a linear operator given by the (studied acceleration of the) Rauzy–Veech cocycle $A^n = A_n \cdots A_1$ as follows: one can show that $S(n)f$ takes constant values f_n^i on each I_n^i and the column vectors $\mathbb{f} := (f^i)_{i=1}^d$ and $\mathbb{f}_n := (f_n^i)_{i=1}^d$ are related by $\mathbb{f}_n = A^n\, \mathbb{f}$. Thus, special Birkhoff sums operators can be seen as infinite-dimensional extensions of the Rauzy–Veech cocycle (and its accelerations).

When considering a rotation R_α, to decompose $S_n f(x)$ into Birkhoff sums of the form $S_{q_k} f(y)$ where $y \in I_k$, one can write $n = \sum_{k=0}^{k_n} b_k q_k$, where k_n is the smallest integer k such that $n < q_k$ and b_k are integers such that $0 \le b_k \le a_k$ (a presentation sometimes known as *Ostrowsky decomposition*). Correspondingly, recalling that $S_{q_k} f(y) = S(k)f(y)$ when $y \in I_k$, we can write

$$S_n f(x) = \sum_{k=0}^{k_n} \sum_{j=0}^{b_k - 1} S(k) f\left(x_j^k\right), \quad \text{where } x_j^k \in I_k, \text{ for all } 0 \le j < b_k. \tag{4.1}$$

For IETs one can also get an analogous decomposition of any Birkhoff sums $S_n f(x)$ into special Birkhoff sums, which has the same form (4.1), but where $0 \le b_k \le \|A^n\| := \sum_{i,j} (A^n)_{ij}$ and the decomposition is obtained *dynamically*, by decomposing the orbit of x until time n into blocks, each of which is contained in a tower and hence corresponds to a special Birkhoff sums.

Renormalization in moduli spaces. We conclude this section mentioning that these renormalization algorithms (for rotations and IETs) describe a discretization of a renormalization dynamics on the moduli space of surfaces. In genus one, the Gauss map is well known to be related to the geodesic flow on the modular surface (which can be seen as the moduli space of flat tori), see, e.g., [66]. Similarly, (an extension of) Rauzy–Veech induction can be obtained as Poincaré map of the *Teichmüller geodesic flow* on the moduli space of translation surfaces (see, e.g., [85]).

The full measure Diophantine-like conditions that we discuss in this survey are satisfied by (Poincaré maps of) linear flows in almost every direction on almost every translation surface in these moduli space (with respect to the Lebesgue, or Masur–Veech measure, see [12]). A different question is whether these properties hold for a *given* surface in almost every direction, in particular if the surface has special properties, for example, is a torus cover (i.e., it is a square-tiled surface), or has special symmetries (e.g., it is a *Veech surface* or it belongs to an $\mathrm{SL}(2,\mathbb{R})$-invariant locus, see [12]). In these settings, while some results can be obtained by general measure-rigidity techniques (in particular, from the work [9] by

Chaika and Eskin, see also the ICM proceedings [12] and the references therein), to describe explicit Diophantine-like conditions, it is often helpful to exploit or develop *ad hoc* renormalization algorithms (for example, one can use finite extensions of the Gauss map to study square-tiled surfaces, see, e.g., [59], or construct Gauss-like maps for some Veech surfaces, see, e.g., [69]).

5. DIOPHANTINE-LIKE CONDITIONS IN HIGHER GENUS

We finally describe in this section some of the Diophantine-like conditions which were introduced to prove some of the results on typical ergodic and spectral properties of smooth area-preserving flows on surfaces (see Section 3.1) and on *linearization* (such as solvability of the cohomological equation and rigidity questions in higher genus, see Section 3.2).

5.1. Bounded-type IETs and Lagrange spectra

We start with two important classes of IETs, namely *periodic-type* and *bounded-type* IETs, both of which have measure zero in the space $\mathcal{I}_d$ of IETs (although full Hausdorff dimension in the case of bounded-type IETs), but often constitute an important class of IETs in which dynamical and ergodic properties can be tested.

One of the simplest requests on a (G)IET is that its orbit under renormalization is *periodic*, so that the sequence of Rauzy–Veech cocycle matrices $(A_n)_{n\in\mathbb{N}}$ introduced in the previous Section 4 is *periodic*, i.e., there exists $p > 0$ such that $A_{n+p} = A_n$ for every $n \in \mathbb{N}$. We will furthermore request that the *period matrix* $A := A_p \cdots A_2 A_1$ is strictly positive. These IETs are called in the literature *periodic-type* IETs (see, e.g., [68]), in analogy with *periodic-type* rotation numbers (quadratic irrationals like the golden mean $(\sqrt{5}-1)/2 = [1, 1, \dots, 1, \dots]$ which have a periodic continued fraction expansion). By construction they are *self-similar*, and one can also show that they arise as Poincaré section of foliations which are fixed by a *pseudo-Anosov* surface diffeomorphism. Notice that d-IETs of periodic type form a *measure zero set* in $\mathcal{I}_d$ (they are actually countable). One can show (in view of a Perron–Frobenius argument, e.g., following [76]) that periodic-type IETs are always *uniquely ergodic* with respect to the Lebesgue measure.

Periodic-type IETs are often the very first type of IETs used to construct *explicit examples*; see, e.g., the explicit examples of weakly mixing periodic-type IETs in [68] or the explicit examples of Roth-type IETs build in the Appendix of [53]. On the other hand, among periodic-type IETs one can also find examples with exceptional behavior. A further request, that is used to guarantee that a periodic-type T displays features similar to those of typical (in the measure theoretical sense) IETs is that T is of *hyperbolic periodic-type*: this means that the periodic matrix A has g eigenvalues of modulus greater than 1, where g is the genus of the surface of which T is a Poincaré section. Notice that g is the largest possible number of such eigenvalues, as it can be shown by either geometric or combinatorial arguments (in particular, exploiting the symplectic features of the cocycle matrices, which come from their interpretation as action of renormalization on the *relative* homology $H_1(S, \mathrm{Fix}(\varphi_{\mathbb{R}}), \mathbb{R})$, one can show that A has also g eigenvalues of modulus less than 1, while the transpose A^T acts as a permutation on a subspace of dimension $k := d - 2g$ which gives rise to a k-dimensional central space).

Bounded-type IETs equivalent characterizations. Periodic-type IETs are a special case of so called bounded-type IETs: we say that a (Keane) IET T is of *bounded-type* if the matrices of the *positive acceleration* $P_k := A(n_k, n_{k+1})$ are uniformly bounded, i.e., there exists a constant $M > 0$ such that $\|P_k\| \leq M$ for every $k \in \mathbb{N}$. From this point of view, bounded-type IETs can be seen as a generalization of *bounded-type* rotation numbers (which, recalling Section 2, are $\alpha = [a_0, a_1, \dots, a_n, \dots]$ such that for some $M > 0$ we have $|a_n| \leq M$). It turns out that this renormalization-based definition characterizes a natural class of IETs (and corresponding surfaces) from the combinatorial and geometric point of view: bounded-type IETs are *linearly recurrent* (i.e., satisfy an important notion of low complexity in word-combinatorics) and surfaces which have a bounded-type IET as a section give rise to *bounded* Teichmüller geodesics in the moduli space of translation surfaces (see, e.g., [31] for the proof of the equivalences). These natural characterizations show once more how the *positive* acceleration (and not simply Zorich acceleration) is the good one to use in this setting (see also [41] where it is shown that asking that Zorich matrices are bounded leads to a different, strictly larger class).

Furthermore, from the point of view of renormalization, the uniform bounds on the norm of the matrices P_k imply that the partitions into Rohlin towers produced by Rauzy–Veech renormalization are all *balanced* (see Section 4). From a purely dynamics perspective, the orbits of a bounded-type IET are *well-spaced*: there are uniform constants $c, C > 0$ such that, for any point x and any n, the *gaps* (i.e., the distances between closest point) of the orbit $\{T^i x, 0 \leq i < n\}$ are all comparable to n, i.e., are bounded below by c/n and above by C/n. Yet another characterization is in terms of orbits of discontinuities: if $\delta_n(T)$ denotes the smallest length of a continuity interval for T_n, $\liminf_{n\in\mathbb{N}} n\delta_n(T) > 0$, see [31] and the reference therein.

Several results in the literature were proved first assuming bounded-type (for example, the absence of mixing for flows in $\mathcal{U}_{\min}$, see [70], preceding [73]) and some properties are currently known only under the assumption of being bounded-type, for example, *absence of partial rigidity* and *mild-mixing* (see [47] and [32], respectively) for flows in $\mathcal{U}_{\min}$ (it is possible, but an open question, that these two properties fail without assuming that a Poincaré section is of bounded-type), or ergodicity of typical skew-product extensions of IETs by piecewise constant cocycles (see [11]).

Bounded-type uniform contraction and deviations estimates. One of the way in which the bounded-type assumption can be exploited is the following. It is well known that iterates of a *positive* $d \times d$ matrix $A > 0$ act on the positive cone $\mathbb{R}^d_+$ as a *strict* contraction (e.g., with respect to the Hilbert projective metric): this is the phenomenon behind the proof of Perron–Frobenius theorem, showing that A has a unique (positive) eigenvector with maximal eigenvalue. More generally, the projective action of any matrix A_i with $\|A_i\| \leq M$ has a contraction rate which depends on M only; this, in view of the connection between the entries of the cocycle products $A^n := A_n \cdots A_1$ and (special) Birkhoff sums (see Section 4), can be used, given a bounded-type IET, to prove unique ergodicity and to give uniform estimates on the rate of convergence of ergodic averages: one can, for example, show that there is a uniform constant (which can be taken to be 1) and a uniform exponent γ_M such that, for any bounded-type IET with $\|P_k\| \le M$ and any mean zero (piecewise) smooth $f : I \to \mathbb{R}$, $|S_n f(x)| \le n^{\gamma_M}$ for all $x \in I$ (see the Appendix of [11]).

The role of bounded-type conditions in the study of Lagrange spectra. Periodic-type and bounded-type rotation numbers play a central role in the study of the *Lagrange spectrum* $\mathcal{L} \subset \mathbb{R} \cup \{+\infty\}$, a classical object in both number theory and dynamics (see, for example, [31] or [58] and the reference therein). It is defined as the set $\mathcal{L} := \{L(\alpha),\ \alpha \in \mathbb{R}\}$ where $L(\alpha) := \limsup_{q,p\to\infty} 1/q|q\alpha - p|$; one can show that $L(\alpha) < \infty$ exactly when α is of bounded-type, in which case $L(\alpha)^{-1}$ provides the smallest constant such that $|\alpha - p/q| < L(\alpha)^{-1}/q^2$ has infinitely many integer solutions $p, q \in \mathbb{Z}, q \neq 0$ (and it has also an interpretation in terms of depths of excursions into the cusp of hyperbolic geodesics on the modular surface). Among the many geometric and dynamical extensions of the notion of Lagrange spectrum (see some of the references in [31]), a natural generalization to higher genus leads to Lagrange spectra of IETs and translation surfaces, which we introduced in joint work with Hubert and Marchese in [31]. The finite values of these spectra are achieved exactly by bounded-type IETs and can be computed using renormalization. We show furthermore in [31] that these spectra can be obtained as the *closure* of the values achieved by periodic-type IETs.

5.2. Roth-like conditions and type

The *Roth-type* condition, to the best of our knowledge, was historically the first full measure "arithmetic" condition to be defined and exploited in higher genus.

Roth-type condition. In the seminal paper [53], Marmi, Moussa, and Yoccoz show first of all that (a predecessor of) the positive acceleration of Rauzy–Veech induction (refer to Section 4) is well defined for all Keane IETs and use this acceleration to define the Roth-type condition and prove that it has full measure; they then show that this condition is sufficient to solve the cohomological equation after removing obstructions (see Section 3.2). Since bounded-type IETs have measure zero, to describe a full measure set of IETs, one needs to allow the norms $\|P_k\|$ of the matrices $(P_k)_k$ of the positive acceleration to grow. Marmi, Moussa, and Yoccoz show in [53] that, for almost every d-IETs in $\mathcal{I}_d$, the matrices $(P_k)_k$ grow subpolynomially, i.e., for any $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that

$$\|P_{k+1}\| \le C_\varepsilon \|Q_k\|^\varepsilon, \quad \text{where } Q_k := P_k \cdots P_1. \tag{5.1}$$

This condition should be seen as a higher genus generalization of the classical Roth-type condition, see Section 2. A d-IETs is called *Roth-type* if it satisfies (5.1) (which is equivalent to condition (a) in [53], see [57]), and two additional conditions, which concern the contraction properties of the cocycle (condition (b) in [53] imposes that the operators $S(k)$ act as contractions on mean-zero functions and guarantees unique ergodicity and the existence of a *spectral gap*, while the last one, condition (c) or *coherence*, concerns the contraction rate of the stable space and its quotient space). The presence of additional requests that concern not only the growth of the matrices but also their hyperbolicity properties seems to be an important and new feature of several Diophantine-like conditions in higher genus, see Section 5.4. While the proof that the latter two conditions are satisfied by almost every IET is a simple consequence of Forni's work **[23]** and Oseledets theorem (which can be applied in view of the work by Zorich **[82]**), the proof that the growth condition (5.1) is typical takes a large part of **[53]**; a simpler proof can be now deduced (as explained in **[52]**) from a later result by Avila–Gouezel–Yoccoz **[3]**.

Variations of the Roth-type condition. As we saw, the periodic-type condition can be refined to the more restrictive condition of *hyperbolic* periodic-type. In a similar way, one may further request, given a Roth-type IET T, that the stable space, i.e., the space $\Gamma_s(T)$ of vectors $v \in \mathbb{R}^d$ such that $A^n v \to 0$ exponentially as n grows (which, in the case of a periodic-type IET with period matrix A, is generated by the eigenvectors corresponding to the eigenvalues of A which have modulus greater than 1) has maximal dimension, namely g. The condition that one gets was called *restricted Roth-type* in **[55]**; it has full measure in view of **[23]** and was used to study the structure and codimension of local $\mathscr{C}^r$ conjugacy class of a (G)IET for $r > 2$. In the joint work **[56]** with Marmi and Yoccoz, we introduced a further weakening of the (restricted) Roth-type condition, the *absolute* (restricted) Roth type condition, expressed only in terms of the cocycle action on a $2g$-dimensional subspace which can be identified with the *absolute* homology $H_1(S, \mathbb{R})$ of the surface S of which T is section (in contrast, the original condition involves the whole cocycle, which describes the action on the *relative* homology $H_1(S, \mathrm{Fix}(\varphi_{\mathbb{R}}), \mathbb{R})$). Exploiting **[9]**, one can also show that this absolute (restricted) Roth-type condition holds on every translation surface for almost every direction (see **[9]** and **[56]**). A generalization of the *restricted* Roth-type condition, the *quasi-Roth-type* condition, was introduced in **[24]** to extend the results of **[53]** and **[55]** to Poincaré maps of surfaces for which the stable space has dimension less than g (see **[24]** for details). Let us also mention that a Roth-type condition can also be imposed on the *backward rotation number* (of a translation flow), requesting a growth rate similar to (5.1) for the *dual* cocycle. The corresponding *dual Roth-type condition* was used in **[56]** to study the asymptotic oscillations of the error term in (3.3) (which we describe in terms of a *distributional* cocycle or *distributional limit shape*, see **[56]** for details).

Type and recurrence for IETs. It is not surprising that Diophantine-like conditions can also be used to study *recurrence* questions. While for rotations these reduce to Diophantine properties in the classical arithmetic sense (namely how well a number can be approximated by rationals), given an IET T, one can study either how frequently the successive iterates $(T_n(x))_{n\in\mathbb{N}}$ return close to x (see, e.g., **[5]**), or how close the iterates of a discontinuity come to other discontinuities, see, e.g., **[51]**. The (Diophantine) *type* η of a rotation R_α is defined to be $\eta := \sup\{\beta : \liminf_{n\to\infty} n^\beta \{n\alpha\} = 0\}$. Bounded- and Roth-type numbers have type $\eta = 1$ (while Liouville ones have type $\eta = \infty$). One can show (see **[40]** and **[55]**) that requesting an IET T be of Roth-type is equivalent to asking that $\sup\{\beta : \liminf n^\beta \delta_n(T) = 0\} = 1$, where here $\delta_n(T)$ is the minimum spacing between discontinuities of T_n. It also implies (but without equivalence) that the first return time $\tau_r(x)$ of x to a ball of radius $r > 0$ satisfies the logarithmic law $\lim_{r\to 0} \log \tau_r(x)/\log(1/r) = 1$ for almost every $x \in [0,1]$ (see [40]).

5.3. Controlled growth Diophantine-like conditions

Any *balanced acceleration* of Rauzy–Veech induction (as defined in Section 4), produces, given a typical IET T, a sequence of times $(n_k)_k$ which correspond to occurrences of positive matrices A_{n_k} whose norm $\|A_{n_k}\| \le M$ is uniformly bounded (these are, furthermore, return times to a compact subset K of the parameter space for the natural extension). As for bounded-type IETs, occurrences of these positive bounded matrices give very good control of the convergence of (special) Birkhoff sums of characteristic functions $\chi_{I_0^j}$ (see the end of Section 5.1). More generally, if $x_0 \in I_n^j$ belongs to the inducing interval I_n of a balanced return time $n := n_k$ and $q := r_n^j$ is the height of the corresponding tower, the orbit $\{x_0, T(x_0), \dots, T^{q-1}(x_0)\}$ *along a tower* is so regularly spaced that one can get good estimates of the Birkhoff sums $S_q f(x_0)$ also for other classes of observables f. In order to estimate Birkhoff sums $S_n f(x)$ for other times $n \in \mathbb{N}$ and points $x \in [0,1]$, one can then *interpolate* these estimates by using the decomposition (4.1) into special Birkhoff sums. It is clear now that for this interpolation to provide good estimates for any time $n \in \mathbb{N}$, one needs to impose that the balanced times $(n_k)_k$ are sufficiently frequent so that $\|A(n_k, n_{k+1})\|$ grows in a controlled way. Notice that by balance the tower heights $r_{n_k}^j$ for $1 \le j \le d$ are all comparable and if we set $q_n := \max_j r_n^j$, the norm $\|A(n_k, n_{k+1})\|$ is proportional to $q_{n_{k+1}}/q_{n_k}$.

Mixing Diophantine condition. The main requirement of the mixing Diophantine condition introduced in [71] is that there exist a (good) positive acceleration and $C > 0$ such that

$$\left\|A(n_k, n_{k+1})\right\| \le C k^{\tau}, \quad \forall k \in \mathbb{N}, \text{ for some } 1 < \tau < 2. \tag{5.2}$$

This condition should be seen as a higher-genus generalization of the Khanin–Sinai condition $|a_k| \le Ck^{\tau}$ for mixing of Arnold flows, see Section 2. The proof that it is satisfied by a full measure set of IETs follows from a Borel–Cantelli argument analogous to that which can be used in genus one, but the input in higher genus are the highly nontrivial integrability estimates for balanced accelerations proved by Avila, Gouezel, and Yoccoz (which the authors proved to show in [3] that the Teichmüller geodesic flow is exponentially mixing): it is proved in [3] that for any $0 < \nu < 1$, there exists a suitable compact set K such that $\int_K \|A_K\|^{\nu} \mathrm{d}\mu$ is finite (where A_K is the accelerated cocycle and μ the Zorich measure).

In order to prove mixing of (minimal components of) locally Hamiltonian flows in $\mathcal{U}_{\neg\min}$ (i.e., Theorem 3.1), one needs good quantitative estimates on *shearing*: these are given by estimates of Birkhoff sums $S_n f$ over an IET which arise as Poincaré map, for a particular observable f (namely, f is taken to be the derivative of the roof function in the special flow representation of $\varphi_{\mathbb{R}}$), which turns out not to be in L^1 (indeed, the function f has singularities of type $1/x$, which are not integrable). When $n = n_k$ is a balanced time, one can control the corresponding special Birkhoff sums $S(n_k)f$ and show that each Birkhoff sum along a tower $S_q f(x)$, where $q = q^j_{n_k}$ and $x \in I^j_{n_k}$, can be controlled after removing the *closest point* contribution that, in this case, is simply $1/x$. One can indeed show that the *trimmed* Birkhoff sum $S_q f(x) - 1/x$ is asymptotic to $Cq \log q$. The mixing Diophantine condition allows to *interpolate* these estimates and to show that, also for any other $n \in \mathbb{N}$, $S_n f(x)$ grows asymptotically as $Cn \log n$ for all points x *with the exception* of points which belong to a set $\Sigma_n \subset [0,1]$ of measure going to zero. The set Σ_n of points which needs to be removed to get the desired control contains points whose orbits may be *resonant*, in the sense that it may contain a close-to-arithmetic progression near one of the singularities of f, with step $q_{n_k}/q_{n_{k+1}}$ (which can be a very small step if $q_{n_{k+1}}$ is much larger than q_{n_k}).

Ratner Diophantine condition. In order to prove that (minimal components of) locally Hamiltonian flows in $\mathcal{U}_{\neg \min}$ have the *switchable Ratner property* (e.g., Theorem 3.2, see Section 3), one needs more delicate quantitative shearing estimates. Such estimates are proven assuming first of all the mixing Diophantine condition, but the MDC is not sufficient. While mixing is an asymptotic condition and therefore it is sufficient, for all large n, to prove estimates for the Birkhoff sums $S_n f(x)$ (introduced in the previous subsection) on sets of measure tending to 1 (and hence one can remove a set Σ_n whose measure goes to zero), the (switchable) Ratner property requires estimates on arbitrarily large sets of initial points, for *all* large times $n \geq n_0$. If the series $\sum_{n\in\mathbb{N}} \mathrm{Leb}(\Sigma_n)$ were finite, the tail sets of the form $\bigcup_{n\geq n_0} \Sigma_n$ would have arbitrarily small measures, and thus one could throw away these unions for n_0 large. Unfortunately, one can check that the measures $(\mathrm{Leb}(\Sigma_n))_{n\in\mathbb{N}}$ are *not* summable. Instead, we consider a subset $K \subset \mathbb{N}$ such that $\sum_{n\notin K} \mathrm{Leb}(\Sigma_n) < +\infty$ and exploit the additional freedom given by the *switchable* Ratner condition to deal with points $x \in \Sigma_n$ when $n \in K$. This requires the introduction of a suitable Diophantine-like condition.

We say that an IET T satisfies the *Ratner Diophantine condition* (RDC) if T satisfies the mixing DC along the sequence $(n_k)_{k\in\mathbb{N}}$ of balanced induction times and there exist $0 < \xi, \eta < 1$ such that, if $B_k := A(n_k, n_{k+1})$ are the matrices of the accelerated cocycle and $q_k := \max_j r^j_{n_k}$ the maximum height of the corresponding towers, then we have

$$\sum_{k\notin K} 1/(\log q_k)^\eta < +\infty, \quad \text{where } K := \left\{k \in \mathbb{N} : \|B_k\| \leq k^\xi\right\}. \tag{5.3}$$

The assumption (5.3) guarantees in particular the summability of $\sum_{n\notin K} \mathrm{Leb}(\Sigma_n)$, so that tail sets of this series *can* be removed. When $k \in K$, using that n_k is a balanced time and $q_k/q_{k-1} \leq \|B_k\|$ is not too large, one can show that an arbitrarily large set of points x do not get close of order c/q_{k-1} to a singularity *twice* in time of order q_k, so by either going *forward* or *backward* in time one can avoid getting $O(q_k^{-1})$ close to singularities. This suffices to provide the control of $S_n f(x)$ (and therefore of *shearing*) required by the switchable Ratner property for all times.

Notice that if an IET T is of bounded type (so $\|B_k\|$ are bounded) then the RDC is automatically satisfied (since the complement of K in $\mathbb{N}$ is finite and therefore the series is a sum of finitely many terms). The Ratner DC imposes that the times k for which $\|B_k\|$ is large are not too frequent: in a sense if an IET satisfies the RDC, it behaves like an IET of bounded type modulo some error with small density (as a subset of $\mathbb{N}$), but this relaxation allows the property to hold for almost every IETs: we prove in [33] that, indeed, for suitable choices of ξ and η, the RDC is satisfied by a full measure set of IETs. Formally (when using the suitable acceleration), the assumption (5.3) looks like the Diophantine condition for rotations introduced by Kanigowski and Fayad in [16], see (3.2). The proof of full measure of the RDC is modeled on the proof of full measure of the arithmetic condition (3.2), with the role of the Gauss map played by the renormalization operator in the parameter space corresponding to the balanced acceleration. Key ingredients to make this proof work are once more the integrability estimates by Avila–Gouezel–Yoccoz [3], as well as a *quasi-Bernoulli* property of the balanced acceleration, see [33] for details.

Backward growth condition for absence of mixing. The Diophantine-like condition to prove absence of mixing of typical locally Hamiltonian flows in $\mathcal{U}_{\min}$ (see Section 3.1) is not explicitly stated in [73], but, from the proof, one can see that one needs the existence of a suitable acceleration of the balanced acceleration, whose matrices will be denoted by $(B_k)_{k\in\mathbb{N}}$, of a subsequence $(k_l)_l$ and of a constant $M > 0$ such that

$$\sum_{k=0}^{k_l} \frac{\|B_k\|}{\nu^{k_l-k}} = \sum_{j=0}^{k_l} \frac{\|B_{k_l-j}\|}{\nu^j} \leq M < +\infty, \quad \text{for all } k \in \mathbb{N}, \tag{5.4}$$

where ν is some constant with $\nu > 1$. Such a condition has two interesting features: it requires a *backward* control of the growth of the matrices of an accelerated cocycle, which has to happen *infinitely often*. Indeed, for the series (5.4) to converge and be uniformly bounded by M, one needs to ask that the norms $\|B_k\|$ when k belongs to the sequence $(k_l)_{l\in\mathbb{N}}$ are uniformly bounded; furthermore, it is sufficient to then impose that, going backward in time, they grow slower than the denominator, namely that $\|B_{k_l-j}\| \leq Ce^{\delta j}$ for $0 \leq j \leq k_l$ where δ is chosen so that $e^{\delta} < \nu$. These conditions can be shown to be of full measure by exploiting Oseledets integrability (for the *dual* cocycle).

Such backward conditions seem to appear naturally when one wants to provide good control of the deviations of the points in a finite segment $\{x, T(x), \dots, T_N(x)\}$ of an IET orbit from an arithmetic progression: one would like to show, for example, that, if we relabel the points in the orbit segment so that $0 < x_1 < x_2 < \cdots < x_N < 1$, the points x_i display polynomial deviations from an arithmetic progression, i.e., there exist $C > 0$ and $0 < \gamma < 1$ such that $|x_i - i/N| \leq C(i/N)^\gamma$. These estimates (which are used in [70,73] to show, through a cancelations mechanism, that there is a subsequence of times with no shearing and, as a consequence, that mixing fails) can be proved for all times for bounded type IETs (see [70]), but, for typical IETs, even for orbits along a balanced tower of some renormalization level n_{k_0}, it may not be possible to choose a constant C uniformly in i. Heuristically, the reason for this is that, to estimate the location of x_i, one can use a *spatial decomposition* of the interval $[0, x_i]$ into floors of renormalization towers which involves the entries of *backward* cocycle matrices (a decomposition similar to that in (4.1), but with the role of time now played by space; geometrically this can also be interpreted as swapping the role of the horizontal and vertical flows on a translation surface). The presence of an exceptionally large $\|A_k\|$, even if k is much smaller than k_0, can still spoil the deviations control, since it may correspond in the spatial decomposition to a *clustering* of points, close to an arithmetic progression of a very small step.

We point out that phenomena of similar nature, where the whole backward history of the continued fraction entries matters to control orbits, appears also in genus one, in the theory of circle diffeomorphisms. For example, in the paper [39] (in which the Herman's theory of linearization briefly recalled in Section 2) is revisited, following [38], through the renormalization perspective and optimal results are achieved for low regularity), the finiteness of a series of the form $\sum_{n=n_0}^{\infty} a_{n+1}\big(\sum_{i=0}^{n}\frac{l_n}{l_{n-i}}(l_{n-i-1})^{\eta}\big)$, where $(a_n)_{n\in\mathbb{N}}$ are the CF entries of $\alpha=[a_0,a_1,\ldots]$ $l_n := |q_n\alpha - p_n|$ and $0<\eta<1$, is used to control the *spatial decomposition* of orbit segments. It would be interesting to know if the analogy, which at this level is only formal and on the *nature* of the conditions, hides a more profound similarity.

5.4. Effective Oseledets Diophantine-like conditions

To conclude, we briefly describe the uniform and regular Diophantine-like conditions (UDC and RDC, for short), introduced and used to prove Theorems 3.5 and 3.3, respectively (see Sections 3.2 and 3.1). Both these conditions present a novel aspect: not only they impose conditions which control the *growth* of cocycle matrices of a suitable acceleration (as all the conditions we have seen in Section 5.3), as well *hyperbolicity* assumptions (as, for example, the *hyperbolic* periodic-type or the *restricted* Roth-type condition, in Sections 5.1 and 5.2), but they also impose *quantitative* forms of *hyperbolicity*, by asking for *effective* bounds on the convergence rates in the conclusion of Oseledets theorem, as we now detail.

Effective Oseledets control and the UDC. Let us say that a sequence of balanced return times $(n_k)_{k\in\mathbb{N}}$ satisfies an *effective* Oseledets control if one can find a sequence of *invariant splittings* $\mathbb{R}^d = E_s^n \oplus E_c^n \oplus E_u^n$, with $\dim E_s^n = g$, such that, for some $\theta>0$ and any $k\in\mathbb{N}$,

$$\big\|A(n_k,n)|_{E_s^{n_k}}\big\|_\infty \le Ce^{-\theta(n-n_k)} \quad \text{for every } n\ge n_k; \tag{5.5}$$

$$\big\|A(n,n_k)^{-1}|_{E_u^{n_k}}\big\|_\infty \le Ce^{-\theta(n_k-n)} \quad \text{for every } 0\le n\le n_k. \tag{5.6}$$

Thus, the cocycle contracts the stable space $E_s^{n_k}$ in the future and the unstable space $E_u^{n_k}$ in the past with a *uniform* rate θ and a uniform constant C. These times can be produced, for example, by considering returns to a set (for the natural extension) where the conclusion of Oseledets theorem (for the cocycle and its inverse) can be made uniform. An IET satisfies the *uniform Diophantine condition* (UDC) if there exists balanced times $(n_k)_k$ with effective Oseledets control and, furthermore, for every $\varepsilon>0$ there exist $C,c>0$, $\lambda>0$ and a subsequence $(k_l)_{l\in\mathbb{N}}$ which is *linearly growing* (i.e., such that $\liminf_{l\to\infty} k_l/l>0$), for which

$$\big\|A(n_k,n_{k_l})\big\| \le C_\varepsilon e^{\varepsilon|k-k_l|} \quad \text{for all } k\ge 0 \text{ and } l\ge 0; \tag{5.7}$$

$$ce^{\lambda k} \le \big\|A(0,n_k)\big\| \le Ce^{(\lambda+\varepsilon)k} \quad \text{for all } \ k\ge 0. \tag{5.8}$$

One can show that assuming that T satisfies the RDC implies, in particular that T is of (restricted) Roth-type (see [26]); on the other hand, (5.5) and (5.6) are assumptions of a new nature, and furthermore (5.8) clearly excludes IETs of bounded type; thus this is a more restrictive Diophantine-like condition, although still of full measure (see [26]).

The RDC and conditions on Diophantine series. In the regular Diophantine condition (used to study rigidity of GIETs in [29] and, in particular, to prove Theorem 3.5), we assume that T is Oseledets generic and require the existence of a special sequence of balanced times $(n_k)_k$ such that the two following *forward* and *backward* series (involving the accelerated matrices $B_k := A(n_k, n_{k+1})$, their products $B(k,l) := B_l B_{l-1} \cdots B_{k+1}$, as well as the projections Π_s^k and Π_u^k to $E_s^{n_k}$ and $E_u^{n_k}$, respectively) are uniformly bounded by some constant $M > 0$ along a linearly growing subsequence $(k_l)_{l\in\mathbb{N}}$, namely, for every $l \in \mathbb{N}$,

$$\sum_{k=1}^{k_l} \left\| B(k,k_l)_{|E_s^{n_k}} \right\| \left\| \Pi_s^k \right\| \|B_{k-1}\| \leq M, \qquad \sum_{k=k_l+1}^{\infty} \left\| B(k_l,k)^{-1}_{|E_u^k} \right\| \left\| \Pi_u^k \right\| \|B_{k-1}\| \leq M. \tag{5.9}$$

We also require a uniform lower bound on the *angles* between the subspaces E_s^n, E_u^n, and E_c^n of the splitting along the subsequence $(n_{k_l})_l$ and subexponential growth of $B(k_l, k_{l+1})$. The convergence of these series can be proved assuming that the sequence $(n_k)_k$ provides effective Oseledets control; the subsequence $(k_l)_l$ is then selected so that the uniform upper bound holds. We remark that also the UDC can be used to prove the convergence and uniform boundedness (along a linearly growing subsequence) of some series of similar (although simpler) nature (that we call *Diophantine series*, see [26] for details). Notice also the similarity between the backward series in (5.9) and the series (5.4) used to prove absence of mixing, even though the latter involves only the norm of the matrices and not their hyperbolic properties.

Examples of arithmetic conditions on classical rotation numbers which do not depend only on the asymptotic behavior of the continued fraction entries (as *Diophantine* or *Roth-type* conditions) but instead depend on values or finiteness of series involving continued fraction entries include the *Brjuno*-condition (see, e.g., [78]) and the *Perez–Marco* condition [62]. Conditions which require recurrence to a set of rotation numbers with this type of control in the theory of circle diffeos seem to appear in global rigidity results, see, for example, Condition (H) defined by Yoccoz [79].

Final remarks and questions. We saw that advancements in our understanding of both chaotic properties and linearization and rigidity questions in the context of surface flows in higher genus depend crucially on sometimes delicate Diophantine-like conditions, imposed to control the renormalization dynamics. While some of these resemble the classical counterparts, others are of new nature and involve in particular hyperbolicity features which become visible only in higher genus. A downside of this new aspect is that conditions that require Oseledets genericity assumptions are not easily checkable. If there is a way of producing explicit examples with such properties which are not of periodic type, even within a locus, remains a challenge. Since many developments are still quite recent, it is possible that some conditions can be simplified or weakened and still yield the same results; furthermore, the interdependence or inclusions between the various conditions have not been fully investigated. Finally, even though, all the conditions we described, with the only exception of bounded-type conditions, are of full measure, they are likely not to be the optimal ones required for the results for which they were introduced (we know this, for example, for the absence of mixing condition, in view of [13]). Finding optimal conditions for each of these problems is certainly interesting, but probably very difficult.

ACKNOWLEDGMENTS

I would like to acknowledge Selim Ghazouani, Giovanni Forni, and Stefano Marmi for many interesting discussions, as well as for their useful feedback on an earlier version of this text. A thank goes also to Stefan Kurz for allowing me to use Figure 3 and to P. Berk, K. Frączek, D. H. Kim, F. Trujillo, and T. Kappeler for reading and commenting this text.

FUNDING

This work was partially supported by the Swiss National Science Foundation.

REFERENCES

[1] V. I. Arnold, Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspekhi Mat. Nauk* **18** (1963), no. 6, 91–192 (Russian).

[2] V. I. Arnold, Topological and ergodic properties of closed 1-forms with incommensurable periods. *Funktsional. Anal. i Prilozhen.* **25** (1991), no. 2, 1–12.

[3] A. Avila, S. Gouëzel, and J.-C. Yoccoz, Exponential mixing for the Teichmüller flow. *Publ. Math. Inst. Hautes Études Sci.* **104** (2006), 143–211.

[4] P. Berk and A. Kanigowski, Spectral disjointness of rescalings of some surface flows. *J. Lond. Math. Soc. (2)* **103** (2021), no. 3, 901–942.

[5] M. Boshernitzan and J. Chaika, Diophantine properties of IETs and general systems: quantitative proximality and connectivity. *Invent. Math.* **192** (2013), no. 2, 375–412.

[6] X. Bressaud, P. Hubert, and A. Maass, Persistence of wandering intervals in self-similar affine interval exchange transformations. *Ergodic Theory Dynam. Systems* **30** (2010), no. 3, 665–686.

[7] A. I. Bufetov, Limit theorems for translation flows. *Ann. of Math. (2)* **179** (2014), no. 2, 431–499.

[8] R. Camelier and C. Gutierrez, Affine interval exchange transformations with wandering intervals. *Ergodic Theory Dynam. Systems* **17** (1997), no. 6, 1315–1338.

[9] J. Chaika and A. Eskin, Every flat surface is Birkhoff and Oseledets generic in almost every direction. *J. Mod. Dyn.* **9** (2015), 1–23.

[10] J. Chaika, K. Frączek, A. Kanigowski, and C. Ulcigrai, Singularity of the spectrum for smooth area-preserving flows in genus two and translation surfaces well approximated by cylinders. *Comm. Math. Phys.* **381** (2021), no. 3, 1369–1407.

[11] J. Chaika and D. Robertson, Ergodicity of skew products over linearly recurrent IETs. *J. Lond. Math. Soc. (2)* **100** (2019), no. 1, 223–248.

[12] J. Chaika and B. Weiss, The horocycle flow on the moduli space of translation surfaces. In *ICM 2022 Proceedings, Vol. 5*, pp. 3412–3430, EMS Press, 2022.

[13] J. Chaika and A. Wright, A smooth mixing flow on a surface with nondegenerate fixed points. *J. Amer. Math. Soc.* **32** (2019), no. 1, 81–117.

[14] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic theory*. Springer, 1980.

[15] A. Denjoy, Sur les courbes definies par les équations différentielles à la surface du tore. *J. Math. Pures Appl. (9)* **11** (1932), 333–375.

[16] B. Fayad and A. Kanigowski, Multiple mixing for a class of conservative surface flows. *Invent. Math.* **203** (2016), no. 2, 555–614.

[17] B. Fayad, A. Kanigowski, and G. Forni, Lebesgue spectrum of countable multiplicity for conservative flows on the torus. *J. Amer. Math. Soc.* **34** (2021), 747–813.

[18] B. Fayad and A. Katok, Constructions in elliptic dynamics. *Ergodic Theory Dynam. Systems* **24** (2004), no. 5, 1477–1520.

[19] L. Flaminio and G. Forni, Invariant distributions and time averages for horocycle flows. *Duke Math. J.* **119** (2003), no. 3, 465–526.

[20] L. Flaminio and G. Forni, On the cohomological equation for nilflows. *J. Mod. Dyn.* **1** (2007), no. 1, 37–60.

[21] G. Forni, Solutions of the cohomological equation for area-preserving flows on compact surfaces of higher genus. *Ann. of Math. (2)* **146** (1997), no. 2, 295–344.

[22] G. Forni, Asymptotic behaviour of ergodic integrals of 'renormalizable' parabolic flows. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pp. 317–326, Higher Ed. Press, Beijing, 2002.

[23] G. Forni, Deviation of ergodic averages for area-preserving flows on surfaces of higher genus. *Ann. of Math. (2)* **155** (2002), no. 1, 1–103.

[24] G. Forni, S. Marmi, and C. Matheus, Cohomological equation and local conjugacy class of diophantine interval exchange maps. 2018, arXiv:1712.02509. To appear in *Proc. Amer. Math. Soc.*, DOI 10.1090/proc/14538.

[25] K. Frączek and M. Lemańczyk, A class of special flows over irrational rotations which is disjoint from mixing flows. *Ergodic Theory Dynam. Systems* **24** (2004), no. 4, 1083–1095.

[26] K. Frączek and C. Ulcigrai, On the asymptotic growth of Birkhoff integrals for locally Hamiltonian flows and ergodicity in their extensions. 2021, arXiv:2112.05939.

[27] S. Ghazouani, Une invitation aux surfaces de dilatation. 2019, arXiv:1901.08856.

[28] S. Ghazouani, Local rigidity for periodic generalised interval exchange transformations. *Invent. Math.* **226** (2021), no. 2, 467–520.

[29] S. Ghazouani and C. Ulcigrai, A priori bounds for giets, affine shadows and rigidity of foliations in genus 2. 2021, arXiv:2106.03529.

[30] M.-R. Herman, Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations. *Publ. Math. Inst. Hautes Études Sci.* **49** (1979), 5–233.

[31] P. Hubert, L. Marchese, and C. Ulcigrai, Lagrange spectra in Teichmüller dynamics via renormalization. *Geom. Funct. Anal.* **25** (2015), no. 1, 180–255.

[32] A. Kanigowski and J. Kułaga-Przymus, Ratner's property and mild mixing for smooth flows on surfaces. *Ergodic Theory Dynam. Systems* **36** (2016), no. 8, 2512–2537.

[33] A. Kanigowski, J. Kułaga-Przymus, and C. Ulcigrai, Multiple mixing and parabolic divergence in smooth area-preserving flows on higher genus surfaces. *J. Eur. Math. Soc. (JEMS)* **21** (2019), no. 12, 3797–3855.

[34] A. Kanigowski, M. Lemańczyk, and C. Ulcigrai, On disjointness properties of some parabolic flows. *Invent. Math.* **221** (2020), no. 1, 1–111.

[35] A. B. Katok, Invariant measures of flows on oriented surfaces. *Sov. Math., Dokl.* **14** (1973), 1104–1108.

[36] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*. Encyclopedia Math. Appl. 54, Cambridge University Press, Cambridge, 1995.

[37] M. Keane, Interval exchange transformations. *Math. Z.* **141** (1975), 25–31.

[38] K. M. Khanin and Y. G. Sinaĭ, A new proof of M. Herman's theorem. *Comm. Math. Phys.* **112** (1987), no. 1, 89–101.

[39] K. Khanin and A. Teplinsky, Herman's theory revisited. *Invent. Math.* **178** (2009), no. 2, 333–344.

[40] D. H. Kim, Diophantine type of interval exchange maps. *Ergodic Theory Dynam. Systems* **34** (2014), no. 6, 1990–2017.

[41] D. H. Kim and S. Marmi, Bounded type interval exchange maps. *Nonlinearity* **27** (2014), no. 4, 637–645.

[42] A. V. Kočergin, The absence of mixing in special flows over a rotation of the circle and in flows on a two-dimensional torus. *Dokl. Akad. Nauk SSSR* **205** (1972), 512–518.

[43] A. V. Kočergin, Mixing in special flows over a shifting of segments and in smooth flows on surfaces. *Mat. Sb.* **96** (1975), no. 138, 471–502.

[44] A. V. Kochergin, Some generalizations of theorems on mixing flows with nondegenerate saddles on a two-dimensional torus. *Mat. Sb.* **195** (2004), no. 9, 19–36.

[45] A. V. Kochergin, Nondegenerate saddles, and absence of mixing. II. *Mat. Zametki* **81** (2007), no. 1, 145–148.

[46] M. Kontsevich, Lyapunov exponents and Hodge theory. In *The mathematical beauty of physics (Saclay, 1996)*, pp. 318–332, Adv. Ser. Math. Phys. 24, World Sci. Publ., River Edge, NJ, 1997.

[47] J. Kułaga, On the self-similarity problem for smooth flows on orientable surfaces. *Ergodic Theory Dynam. Systems* **32** (2012), no. 5, 1615–1660.

[48] M. Lemańczyk, Furstenberg disjointness, Ratner properties and Sarnak's conjecture. In *ICM 2022 Proceedings, Vol. 5*, pp. 3508–3528, EMS Press, 2022.

[49] G. Levitt, La décomposition dynamique et la différentiabilité des feuilletages des surfaces. *Ann. Inst. Fourier (Grenoble)* **37** (1987), no. 3, 85–116.

[50] R. Mañé, Chap. I.8. In *Ergodic theory and differentiable dynamics*, pp. 52–57, Springer, 1987.

[51] L. Marchese, The Khinchin theorem for interval-exchange transformations. *J. Mod. Dyn.* **5** (2011), no. 1, 123–183.

[52] S. Marmi, C. Matheus, and P. Moussa, Fonctions de Brjuno, échanges d'intervalles, origamis et autres objets insolites. In *'Jean-Christophe Yoccoz', Mathématiciens*, SMF, 2018.

[53] S. Marmi, P. Moussa, and J.-C. Yoccoz, The cohomological equation for Roth-type interval exchange maps. *J. Amer. Math. Soc.* **18** (2005), no. 4, 823–872 (electronic).

[54] S. Marmi, P. Moussa, and J.-C. Yoccoz, Affine interval exchange maps with a wandering interval. *Proc. Lond. Math. Soc. (3)* **100** (2010), no. 3, 639–669.

[55] S. Marmi, P. Moussa, and J.-C. Yoccoz, Linearization of generalized interval exchange maps. *Ann. of Math. (2)* **176** (2012), no. 3, 1583–1646.

[56] S. Marmi, C. Ulcigrai, and J.-C. Yoccoz, On Roth type conditions, duality and central Birkhoff sums for I.E.M. In *Quelques aspects de la théorie des systèmes dynamiques: un hommage à Jean-Christophe Yoccoz. II*, pp. 65–132, Astérisque 416, SMF, 2020.

[57] S. Marmi and J.-C. Yoccoz, Hölder regularity of the solutions of the cohomological equation for Roth type interval exchange maps. *Comm. Math. Phys.* **344** (2016), no. 1, 117–139.

[58] C. Matheus, The Lagrange and Markov spectra from the dynamical point of view. In *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, pp. 259–291, Lecture Notes in Math. 2213, Springer, Cham, 2018.

[59] C. Matheus, M. Möller, and J.-C. Yoccoz, A criterion for the simplicity of the Lyapunov spectrum of square-tiled surfaces. *Invent. Math.* **202** (2015), no. 1, 333–425.

[60] I. Nikolaev and E. Zhuzhoma, *Flows on 2-dimesional manifolds*. Lecture Notes in Math. 1705, Springer, 1999.

[61] S. P. Novikov, The Hamiltonian formalism and a multivalued analogue of Morse theory. *Uspekhi Mat. Nauk* **37** (1982), no. 5, 3–49 (Russian).

[62] R. Pérez Marco, Sur les dynamiques holomorphes non linéarisables et une conjecture de V. I. Arnol'd. *Ann. Sci. Éc. Norm. Supér. (4)* **26** (1993), no. 5, 565–644.

[63] H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*. Les Grands Classiques Gauthier-Villars, Librairie Scientifique et Technique Albert Blanchard, Paris, 1987.

[64] D. Ravotti, Quantitative mixing for locally Hamiltonian flows with saddle loops on compact surfaces. *Ann. Henri Poincaré* **18** (2017), no. 12, 3815–3861.

[65] D. Scheglov, Absence of mixing for smooth flows on genus two surfaces. *J. Mod. Dyn.* **3** (2009), no. 1, 13–34.

[66] C. Series, The modular surface and continued fractions. *J. Lond. Math. Soc. (2)* **31** (1985), no. 1, 69–80.

[67] Y. G. Sinai and K. M. Khanin, Mixing for some classes of special flows over rotations of the circle. *Funktsional. Anal. i Prilozhen.* **26** (1992), no. 3, 1–21.

[68] Y. G. Sinai and C. Ulcigrai, Weak mixing in interval exchange transformations of periodic type. *Lett. Math. Phys.* **74** (2005), no. 2, 111–133.

[69] J. Smillie and C. Ulcigrai, Geodesic flow on the Teichmüller disk of the regular octagon, cutting sequences and octagon continued fractions maps. In *Dynamical numbers—interplay between dynamical systems and number theory*, pp. 29–65, Contemp. Math. 532, Amer. Math. Soc., Providence, RI, 2010.

[70] C. Ulcigrai, *Ergodic properties of some area-preserving flows*. Princeton University, 2007.

[71] C. Ulcigrai, Mixing of asymmetric logarithmic suspension flows over interval exchange transformations. *Ergodic Theory Dynam. Systems* **27** (2007), no. 3, 991–1035.

[72] C. Ulcigrai, Weak mixing for logarithmic flows over interval exchange transformations. *J. Mod. Dyn.* **3** (2009), no. 1, 35–49.

[73] C. Ulcigrai, Absence of mixing in area-preserving flows on surfaces. *Ann. of Math. (2)* **173** (2011), no. 3, 1743–1778.

[74] C. Ulcigrai, Shearing and mixing in parabolic flows. In *European congress of mathematics*, pp. 691–705, Eur. Math. Soc., Zürich, 2013.

[75] C. Ulcigrai, Slow chaos in surface flows. *Boll. Unione Mat. Ital.* **14** (2021), no. 1, 231–255.

[76] W. A. Veech, Gauss measures for transformations on the space of interval exchange maps. *Ann. of Math.* **115** (1982), 201–242.

[77] J.-C. Yoccoz, Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition Diophantienne. *Ann. Sci. Éc. Norm. Supér. (4)* **17** (1984), no. 3, 333–359.

[78] J.-C. Yoccoz, Théorème de Siegel, nombres de Bruno et polynômes quadratiques. In *Petits diviseurs en dimension* 1, pp. 3–88, Astérisque 231, SMF, 1995.

[79] J.-C. Yoccoz, Analytic linearization of circle diffeomorphisms. In *Dynamical systems and small divisors (Cetraro, 1998)*, pp. 125–173, Lecture Notes in Math. 1784, Springer, Berlin, 2002.

[80] J.-C. Yoccoz, Échanges d'intervalles et surfaces de translation. In *Séminaire Bourbaki 2007/2008*, Astérisque, Société Mathématique De France, 2009.

[81] J.-C. Yoccoz, Interval exchange maps and translation surfaces. In *Homogeneous flows, moduli spaces and arithmetic*, pp. 1–69, Clay Math. Proc. 10, Amer. Math. Soc., Providence, RI, 2010.

[82] A. Zorich, Finite Gauss measure on the space of interval exchange transformation. Lyapunov exponents. *Ann. Inst. Fourier (Grenoble)* **46** (1996), 325–370.

[83] A. Zorich, Deviation for interval exchange transformations. *Ergodic Theory Dynam. Systems* **17** (1997), no. 6, 1477–1499.

[84] A. Zorich, How do the leaves of a closed 1-form wind around a surface? In *Pseudoperiodic topology*, pp. 135–178, Amer. Math. Soc. Transl. Ser. 2 197, Amer. Math. Soc., Providence, RI, 1999.

[85] A. Zorich, Flat surfaces. In *Frontiers in number theory, physics, and geometry. I*, pp. 437–583, Springer, Berlin, 2006.

CORINNA ULCIGRAI

Universität Zürich Institut für Mathematik, Y27-K36, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, corinna.ulcigrai@math.uzh.ch

SELF-SIMILAR SETS AND MEASURES ON THE LINE

PÉTER P. VARJÚ

ABSTRACT

We discuss the problem of determining the dimension of self-similar sets and measures on $\mathbf{R}$. We focus on the developments of the last four years. At the end of the paper, we survey recent results about other aspects of self-similar measures including their Fourier decay and absolute continuity.

MATHEMATICS SUBJECT CLASSIFICATION 2020

28A80

KEYWORDS

Self-similar set, self-similar measure, Bernoulli convolution, fractal, dimension theory

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3610–3634
DOI 10.4171/ICM2022/94

Published by EMS Press

A (self-similar) iterated function system, IFS for short, is a finite collection

$$\Phi = \{\varphi_i : i \in \Lambda\}$$

of contractive similarities of $\mathbf{R}^d$. A contractive similarity is a map $x \mapsto \lambda \cdot Ux + t$, where $\lambda \in (0,1)$, $U \in \mathrm{O}(d)$ is a rotation, and $t \in \mathbf{R}$. We call λ the contraction factor of the similarity. Given such an IFS, there is a unique self-similar set, that is, a compact set $K \subset \mathbf{R}^d$ such that

$$K = \bigcup_{i \in \Lambda} \varphi_i(K).$$

This set K is also known as the attractor of the IFS. Furthermore, given an IFS and a probability vector $\{p_i : i \in \Lambda\}$, there is a unique self-similar measure, that is, a probability measure μ on $\mathbf{R}^d$ such that

$$\mu = \sum_{i \in \Lambda} p_i \varphi_i(\mu).$$

Here $\varphi_i(\mu)$ denotes the push-forward of μ under φ_i. In other words, μ is the unique stationary measure for the Markov chain on $\mathbf{R}^d$ with transitions φ_i executed with probability p_i. The support of μ equals the self-similar set K provided $p_i > 0$ for all i.

Self-similar sets and measures are central objects of interest in fractal geometry and they include many classical examples of fractals. For example, the attractor of the IFS

$$\{x \mapsto \lambda x - 1, x \mapsto \lambda x + 1\}$$

is (a scaled copy of) the middle $1 - 2\lambda$ Cantor set for $\lambda \in (0, 1/2)$, while for $\lambda \geq 1/2$, the attractor is an interval. The self-similar measure associated to the same IFS with equal probability weights $p_i = 1/2$ is called the Bernoulli convolution and is denoted by ν_λ. They can also be defined as the distribution of the random variables

$$\sum_{n=0}^{\infty} \pm \lambda^n,$$

where the ± are independent fair coin tosses. The study of these measures go back at least to Wintner and his collaborators in the 1930s. See [38] for more on Bernoulli convolutions. Other classical self-similar sets include the Sierpiński triangle and (a side of) the Koch snowflake curve.

The systematic study of self-similar sets and measures was initiated by Hutchinson [25]. We refer to his paper and Falconer's book [15, CHAPTER 9] for thorough treatments of the fundamental properties of these objects.

Determining the dimension of self-similar sets and measures is a central problem in fractal geometry. While there are several competing notions of dimension for sets and measures, most of them coincide in the self-similar case. In this paper, for self-similar sets, by dimension we mean the common value of the Minkowski and Hausdorff dimensions.

The local dimension of a measure μ in $\mathbf{R}^d$ at a point x is

$$\lim_{r \to 0} \frac{\log \mu(B(x,r))}{\log r},$$

provided the limit exists, where $B(x,r)$ is the ball of radius r around x. We say that the measure is exact dimensional if its local dimension exists and is constant μ-almost everywhere. By the dimension of an exact dimensional measure μ we mean this μ-almost constant value of its local dimension. It is known that self-similar measures are exact dimensional (see [18]).

Before we state the main conjectures in the dimension theory of self-similar sets and measures on $\mathbf{R}$, which will be the main focus of this paper, we make some simple observations to motivate them. Let K be the attractor of a self-similar IFS $\{\varphi_i : i \in \Lambda\}$. We write H^s for the s-dimensional Hausdorff measure. Now suppose that the sets $\varphi_i(K)$ are pairwise disjoint for $i \in \Lambda$ and that $0 < H^s(K) < \infty$ for some s. Then we can write

$$H^s(K) = \sum_{i\in\Lambda} H^s\big(\varphi_i(K)\big) = H^s(K) \sum_{i\in\Lambda} \lambda_i^s,$$

where λ_i is the contraction factor of φ_i. It follows that s must be the unique solution of

$$1 = \sum_{i\in\Lambda} \lambda_i^s. \tag{1}$$

While an s with $0 < H^s(K) < \infty$ may not exist in general, if it does, then it must equal the Hausdorff dimension of K. Therefore, the above considerations suggest that a reasonable guess for $\dim(K)$ is the unique solution of the equation (1). It is a classical result going back to Moran [35] in some form, that this guess is correct when the IFS satisfies the so-called open set condition, which is a mild relaxation of requiring that the sets $\varphi_i(K)$ are pairwise disjoint. See [15, CHAPTER 9] for a precise definition.

It turns out that the unique solution of (1) is always an upper bound for $\dim(K)$ and it is natural to ask to what extent it is possible to drop the open set condition without turning this upper bound into a strict inequality. There are two immediate obstructions to this. First, the solution of (1) may be larger than d, but the dimension of K will never exceed d which is the dimension of the ambient space $\mathbf{R}^d$. Second, (1) depends on the IFS and not only on the set K. It may be possible to realize K as the attractor of another IFS such that the corresponding (1) has a smaller solution. This happens, for example, if the IFS contains exact overlaps, which we define now.

Definition 1. An IFS $\{\varphi_i : i \in \Lambda\}$ contains exact overlaps if there is some $n \in \mathbf{Z}_{\geq 1}$ and $(i_1, \dots, i_n) \neq (\tilde{i}_1, \dots, \tilde{i}_n) \in \Lambda^n$ such that

$$\varphi_{i_1} \circ \cdots \circ \varphi_{i_n} = \varphi_{\tilde{i}_1} \circ \cdots \circ \varphi_{\tilde{i}_n}. \tag{2}$$

In other words, the IFS contains no exact overlaps if and only if the semigroup generated by the maps in the IFS with respect to the composition operation is free. We note that it does not make a difference in the definition whether or not we require that we have the same number of composition factors on the two sides of (2).

The next conjecture due to Simon (see [47]) predicts that apart from the above two obstructions, $\dim(K)$ equals the unique solution of (1) in the $d = 1$ case.

Conjecture 2. *Let K be the attractor of a self-similar IFS $\{\varphi_i : i \in \Lambda\}$ on $\mathbf{R}$ that contains no exact overlaps. Let λ_i be the contraction factor of φ_i. Then*

$$\dim K = \min(1, s),$$

where s is the unique solution of the equation

$$\sum_i \lambda_i^s = 1.$$

The conjecture also has a counterpart for measures.

Conjecture 3. *Let μ be the self-similar measure on $\mathbf{R}$ associated to an IFS $\{\varphi_i : i \in \Lambda\}$ without exact overlaps and a probability vector $\{p_i\}$. Let λ_i be the contraction factor of φ_i. Then*

$$\dim \mu = \min\left(1, \frac{\sum_i p_i \log p_i^{-1}}{\sum_i p_i \log \lambda_i^{-1}}\right).$$

Self-similar measures are of interest in their own right, but a major motivation for Conjecture 3 is that it implies Conjecture 2. To see this, recall that if a set K supports an exact dimensional measure μ of dimension s, then the Hausdorff dimension of K is at least s, see [15, PRINCIPLE 4.2]. This is a common way of giving lower bounds on the Hausdorff dimension. Now let s be the solution of (1), and consider the probability weights $p_i = \lambda_i^s$. Observe that this choice yields

$$s = \frac{\sum_i p_i \log p_i^{-1}}{\sum_i p_i \log |\lambda_i|^{-1}},$$

showing that Conjecture 3 indeed implies Conjecture 2.

Almost all of this paper is concerned only with self-similar measures on $\mathbf{R}$. Some difficulties arise when one tries to formulate versions of Conjectures 2 and 3 for self-similar sets and measures in higher-dimensional ambient spaces due to the presence of affine subspaces of intermediate dimension. For a discussion of these issues and results in higher dimension, we refer to [22].

The purpose of this paper is to survey results towards Conjectures 2 and 3. Since this subject has already been exposed by Hochman in his ICM lecture in 2018 [24], we focus on the developments of the last four years and discuss earlier results only to the extent necessary to keep our presentation self-contained.

We will outline some ideas from the proofs of these results; however, we will not give full details, and some of our discussion will be imprecise. Our aim is to overview the theory and give insight into the role played by its components. For details and a rigorous discussion of the proofs we refer to the original papers.

In the final section, we briefly survey some further recent developments on Fourier decay and absolute continuity of self-similar measures.

1. EXPONENTIAL SEPARATION PROPERTY

The exponential separation property was introduced by Hochman [23] who showed that Conjectures 2 and 3 hold when the IFS satisfies this property. This property can be verified in many cases of interest. While these results have been already discussed in [24], we recall them now because they are of crucial importance to later developments both logically and for the motivation of ideas.

We begin with the definitions. We introduce a distance function on the group of similarities on $\mathbf{R}$. Let $\varphi_i : x \mapsto \lambda_i x + t_i$ be similarities for $i = 1, 2$. We define

$$\operatorname{dist}(\varphi_1, \varphi_2) = \begin{cases} |t_1 - t_2| & \text{if } \lambda_1 = \lambda_2, \\ \infty & \text{if } \lambda_1 \neq \lambda_2. \end{cases}$$

Given an IFS $\Phi := \{\varphi_i : i \in \Lambda\}$, we define its level n separation by

$$\Delta_n(\Phi) := \min_{(i_1,\dots,i_n)\neq(\tilde{i}_1,\dots,\tilde{i}_n)\in\Lambda^n} \operatorname{dist}(\varphi_{i_1} \circ \dots \circ \varphi_{i_n}, \varphi_{\tilde{i}_1} \circ \dots \circ \varphi_{\tilde{i}_n}).$$

We say that the IFS satisfies the exponential separation property if there is a constant $c > 0$ such that $\Delta_n(\Phi) > c^n$ for infinitely many n's.

We observe that the IFS contains exact overlaps if and only if $\Delta_n = 0$ for some and hence for all sufficiently large n. The exponential separation property is a quantitative strengthening of the condition that the IFS contains no exact overlaps. Hochman proved that Conjectures 2 and 3 hold under this strengthening of the hypothesis.

Theorem 4 (Hochman [23]). *Let $\{\varphi_i : i \in \Lambda\}$ be an IFS that satisfies the exponential separation property and let K be its attractor. Write λ_i for the contraction factor of φ_i. Then*

$$\dim K = \min(1, s),$$

where s is the unique solution of the equation

$$\sum_i \lambda_i^s = 1.$$

Let μ be the self-similar measure associated to the above IFS and a probability vector $\{p_i\}$. Then

$$\dim \mu = \min\left(1, \frac{\sum_i p_i \log p_i^{-1}}{\sum_i p_i \log \lambda_i^{-1}}\right).$$

It can be shown that the exponential separation property holds in parametric families of IFSs for all but possibly a (packing or Hausdorff) codimension 1 subset of exceptions. This shows that Conjectures 2 and 3 hold generically in a very strong sense. We refer to [23] for details and more precise results.

We also note that a stronger version of Conjecture 3 involving the L^q dimension instead of local dimension of measures was established subsequently by Shmerkin [46] under the exponential separation property. This result has very important and far reaching applications, see also [45] and Shmerkin's paper in this volume.

Our main focus here are explicit cases and families of IFSs for which the exponential separation property can be verified. We first observe that the exponential separation property holds always whenever all contraction and translation parameters in the IFS are rational and the IFS contains no exact overlaps. Indeed, writing Q for the least common denominator of all parameters, a simple calculation shows that the translation parameters of n-fold compositions of maps in the IFS have denominators that divide Q^n. This means that for all n, we have $\Delta_n \geq Q^{-n}$ or $\Delta_n = 0$. The second possibility is excluded by the absence of exact overlaps.

In fact, the above reasoning can be extended to the case when the parameters are algebraic numbers and not necessarily rational. To do this, one need to work with heights instead of denominators, or see [23, THEOREM 1.5] for a more elementary argument. This leads to the following result

Corollary 5 (Hochman). *Conjectures 2 and 3 hold for IFSs in which all contraction and translation parameters are algebraic numbers.*

The exponential separation property can be verified also for certain IFSs that involve transcendental parameters. One such example is the family IFSs

$$\left\{x \mapsto \frac{x}{3}, x \mapsto \frac{x}{3} + 1, x \mapsto \frac{x}{3} + t\right\}, \tag{3}$$

where $t \in \mathbf{R}$ is a parameter. It can be seen that the attractors of these IFSs are the linear projections of the Sierpiński triangle.

Another corollary of Theorem 4 is the following.

Corollary 6 (Hochman). *Conjectures 2 and 3 hold for the IFS (3) for all values of the parameter $t \in \mathbf{R}$.*

We sketch the proof of the exponential separation property for the family (3), as these ideas will recur later. For details, see [23, THEOREM 1.6], where this argument is attributed to Solomyak and Shmerkin. The translation component of an n-fold composition of maps from the above IFS is of the form

$$\sum_{j=0}^{n-1} \alpha_j 3^{-j},$$

where each α_j is equal to 0, 1, or t. Based on this observation, it can be seen that for each t and for each n, there are some integers $a_1, a_2 \in \mathbf{Z}$ not both 0 with $|a_1|, |a_2| \leq 3^{n-1}$ such that

$$\Delta_n = \frac{a_1}{3^{n-1}} - \frac{a_2}{3^{n-1}} t.$$

Assuming $a_2 \neq 0$, which holds whenever $\Delta_n \leq 3^{-n+1}$, we get

$$\left| t - \frac{a_1}{a_2} \right| \leq 3^n \Delta_n.$$

Now fix the value of the parameter t such that the IFS (3) contains no exact overlaps. Suppose $\Delta_n < 27^{-n-1}$ for some n. Then there is a rational number a_1/a_2 as above such that $|t - a_1/a_2| < 9^{-n-1}$. Let $\tilde{n}$ be such that $9^{-\tilde{n}-1} < |t - a_1/a_2| \leq 9^{-\tilde{n}}$. (Note that $t \neq a_1/a_2$, for otherwise we would have $\Delta_n = 0$ and the IFS would contain exact overlaps.) We observe that there is no rational $\tilde{a}_1/\tilde{a}_2$ with $|t - \tilde{a}_1/\tilde{a}_2| < 9^{-\tilde{n}-1}$ and $|\tilde{a}_1|, |\tilde{a}_2| \leq 3^{\tilde{n}-1}$. Indeed, if such a rational existed, we would have

$$\left| \frac{a_1\tilde{a}_2 - a_2\tilde{a}_1}{a_2\tilde{a}_2} \right| = \left| \frac{a_1}{a_2} - \frac{\tilde{a}_1}{\tilde{a}_2} \right| \leq \left| \frac{a_1}{a_2} - t \right| + \left| t - \frac{\tilde{a}_1}{\tilde{a}_2} \right| \leq 2 \cdot 9^{-\tilde{n}}.$$

Since $|a_2\tilde{a}_2| \leq 9^{\tilde{n}-1}$, this would yield $a_1\tilde{a}_2 - a_2\tilde{a}_1 = 0$, leading to $a_1/a_2 = \tilde{a}_1/\tilde{a}_2$ and contradicting

$$\left|t - \frac{\tilde{a}_1}{\tilde{a}_2}\right| < 9^{-\tilde{n}-1} < \left|t - \frac{a_1}{a_2}\right|.$$

This shows that $\Delta_{\tilde{n}} \geq 27^{-\tilde{n}-1}$, and the exponential separation property follows.

A key property of the IFS (3) exploited in the above argument is that exact overlaps occur for certain special values of the parameter t, in this case certain rational numbers, and these special values are very well separated from each other. This will be a recurrent concern for us in what follows.

A similar argument can be made when the contraction factor $1/3$ in (3) is replaced by another algebraic number. We omit the details.

2. BERNOULLI CONVOLUTIONS

In this section, we consider the one parameter family of IFSs

$$\Phi_\lambda := \{x \mapsto \lambda x, x \mapsto \lambda x + 1\},$$

where $\lambda \in (0, 1)$. Instead of 0 and 1 we could take any other pair of distinct real numbers as the translation parameters; we would get the same IFS up to a change of coordinates. In fact, it is more customary to take ±1 instead of 0 and 1, but the above choice will make notation more consistent with the rest of this note.

In this case, the resulting self-similar sets have a simple structure. For $\lambda < 1/2$, it is the middle $(1 - 2\lambda)$th Cantor set, while for $\lambda \geq 1/2$ it is an interval. In both cases, Conjecture 2 is easily verified. However, the associated self-similar measures called Bernoulli convolutions are more difficult to understand. The purpose of this section is to summarize the developments that lead to the following result.

Theorem 7. *Conjecture 3 holds for the IFS Φ_λ for any value of the parameter $\lambda \in (0, 1)$.*

For algebraic parameters, this result is due to Hochman as it falls under the scope of Corollary 5. For transcendental parameters, the result has been established in [54]. Strictly speaking, only the case of uniform $(1/2, 1/2)$ probability weights is treated there, but the arguments can be extended to the general case. Moreover, one can even allow more general IFSs with an arbitrary number of maps as long as the contraction factors are the same and the translation parameters are rational. This has been demonstrated in the Appendix of [41].

To simplify the exposition, we assume in our discussion that the probability weights are uniform. We write ν_λ for the self-similar measure associated to the IFS Φ_λ. We note that ν_λ is the law of the random variable $\sum_{n=0}^{\infty} \xi_n \lambda^n$, where (ξ_n) is a sequence of independent random variables taking the values 0 and 1 with equal probability.

In the algebraic case, Hochman's results yield more information, which allows computing the dimension even in the presence of exact overlaps. This is in terms of the entropy rate of the IFS Φ_λ, which we define now, and which will also play an important role later on. The entropy rate is defined as

$$h(\Phi_\lambda) := \lim_{n\to\infty} \frac{H(\sum_{j=0}^{n-1} \xi_j \lambda^j)}{n},$$

where $H(\cdot)$ stands for Shannon entropy of a discrete random variable. The numerator on the right can be shown to be a subadditive sequence, hence the limit exists and, moreover,

$$h(\Phi_\lambda) \le \frac{H(\sum_{j=0}^{n-1} \xi_j \lambda^j)}{n}$$

for each n.

See [9, SECTION 3.4] for the details of how the following follows from the main result of Hochman [23].

Theorem 8 (Hochman). *Let $\lambda \in (0,1)$ be an algebraic number. Then*

$$\dim \nu_\lambda = \min\left(1, \frac{h(\Phi_\lambda)}{\log \lambda^{-1}}\right). \tag{4}$$

This result, together with Theorem 7, gives an almost complete solution to the problem of determining the dimension of Bernoulli convolutions. In addition, there are numerical algorithms to compute $\dim \nu_\lambda$ with arbitrary precision for any given algebraic λ, see [1,17,21,29]. However, it is still not known precisely what is the set of algebraic parameters $\lambda \in (1/2, 1)$ for which $\dim \nu_\lambda < 1$.

We turn to the case of transcendental parameters in Theorem 7. If the IFS Φ_λ satisfied the exponential separation property whenever it does not contain exact overlaps, then Theorem 7 would follow at once from Theorem 4. This very well could be true; however, this is still an open problem, which seems to be beyond reach of existing methods.

In fact, the decay rate of $\Delta_n(\Phi_\lambda)$ is very closely related to a problem in Diophantine approximation, which is the separation between the elements of the set

$$\mathcal{E}^{(n)} := \{\eta : \ P(\eta) = 0 \text{ for some polynomial } P \in \mathcal{P}^{(n)}\},$$

where $\mathcal{P}^{(n)}$ is the set of polynomials of degree at most $n-1$ with coefficients $-1, 0, 1$. As it will be clear from what follows, the set

$$\mathcal{E} := \bigcup_n \mathcal{E}^{(n)} \cap (0,1)$$

is precisely the set of parameters for which Φ_λ contains exact overlaps.

We begin our discussion of the proof of Theorem 7 by explaining the connection between the behavior of $\Delta_n(\Phi_\lambda)$ and the separation properties of the sets $\mathcal{E}^{(n)}$ following Hochman [23, QUESTION 1.10]. This can be formalized as follows.

Lemma 9. *If it is true that the elements of $\mathcal{E}^{(n)}$ are separated by at least C^{-n} for some constant C for all n, then the exponential separation property holds for the IFS Φ_λ whenever it lacks exact overlaps.*

Sketch of proof. Fix some $\varepsilon > 0$ and assume $\lambda \in (\varepsilon, 1-\varepsilon)$. We first observe that if $\Delta_n(\Phi_\lambda) < C^{-n}$ for some $C = C(\varepsilon)$, then there is some $\eta \in \mathcal{E}^{(n)}$ with

$$|\lambda - \eta| < \Delta_n(\Phi_\lambda)^\alpha$$

for some $\alpha = \alpha(\varepsilon) > 0$. This follows from the fact that the translation component of an n-fold composition of the maps in Φ_λ in some order is a polynomial in λ of degree at most $n-1$ with coefficients $0, 1$. This means that $\Delta_n(\Phi_\lambda) = P(\lambda)$ for some $P \in \mathcal{P}^{(n)}$ that also depends on λ. To complete the proof of our observation, we need to argue that the only way $P(\lambda)$ can be very small is if λ is close to a root of P. For more details, see [52, LEMMA 5.2].

Now suppose that λ is such that $\Delta_n(\Phi_\lambda) < C_2^{-2n/\alpha}$ for some n, where α is as in the previous paragraph and C_2 is the constant C in the assumption about the separation between the elements of $\mathcal{E}_n$. Then there is η_n such that $|\lambda - \eta_n| < C_2^{-2n}$. If Φ_λ contains no exact overlaps, then $\lambda \notin \mathcal{E}$ so $\lambda \neq \eta_n$. Now we take the smallest integer $\tilde{n} > n$ such that $|\lambda - \eta_n| > C_2^{-2\tilde{n}}$. It follows by the assumed separation property on $\mathcal{E}^{(\tilde{n})}$ that there is no $\eta_{\tilde{n}} \in \mathcal{E}^{(\tilde{n})}$ with $|\lambda - \eta_{\tilde{n}}| < C_2^{-2\tilde{n}}$. This means that $\Delta_{\tilde{n}}(\Phi_\lambda) \geq C_2^{-2\tilde{n}/\alpha}$, and the exponential separation property follows. ■

It is not known whether or not the elements of $\mathcal{E}^{(n)}$ are exponentially separated. The best lower bound known for the minimal distance of the elements of $\mathcal{E}^{(n)}$ is $\exp(-Cn\log n)$ for some constant C (one could take, e.g., $C = 4$), which is due to Mahler [33]. This yields via the argument in the proof of Lemma 9 that for all λ such that Φ_λ contains no exact overlaps, there are infinitely many values of n with

$$\Delta_n(\Phi_\lambda) \geq \exp(-Cn\log n) \tag{5}$$

for some (other) constant C.

One may wonder if this weaker separation condition could be used in a refined form of Hochman's argument in place of exponential separation. This has been done in [8], however, the argument requires that there are several values of n sufficiently close to each other such that the separation (5) holds. Such a condition can be satisfied if we assume that λ is not approximated too closely by elements of $\mathcal{E}^{(n)}$. Indeed, in the above argument the size of $\tilde{n}$ is controlled by the distance between λ and $\mathcal{E}^{(n)}$. More precisely, the following was proved in [8].

Theorem 10 (Beruillard, Varjú). *Let $\lambda \in (1/2, 1)$ be such that Conjecture 3 does not hold for Φ_λ. Then there is $\delta > 0$ and there are infinitely many values of n such that there is $\eta_n \in \mathcal{E}^{(n)} \cap (1/2, 1)$ with*

$$\begin{aligned} &|\lambda - \eta_n| < \exp(-n^{100}), \\ &\dim \nu_{\eta_n} < 1 - \delta. \end{aligned}$$

The exponent 100 can be replaced by any other number, or even by a slowly growing function of n, see [8] for details. This result along with Theorem 4 are major ingredients in the proof of Theorem 7. Given some $\lambda \in (1/2, 1)$ such that Φ_λ lacks exact overlaps, it can be shown that λ has only finitely many approximants η_n as in the conclusion of Theorem 10 or else Φ_λ satisfies the exponential separation property. In either case, Conjecture 3 follows for Φ_λ from one of Theorems 4 or 10.

Before we discuss the details of how this can be done, a further remark about Theorem 10 is in order. We have seen that if $\Delta_n(\Phi_\lambda) < C^{-n}$ for some n and λ with an appropriate constant C, then λ is approximated by some $\eta \in \mathscr{E}^{(n)}$. However, we claim some additional properties of this η in Theorem 10, most importantly that $\dim \nu_\eta < 1 - \delta$. Now we indicate how this can be deduced. This leads us to a somewhat lengthy digression; however, it also gives us the opportunity to introduce several concepts and ideas that will be needed later on.

Already in Theorem 4, the exponential separation property can be relaxed (see [23, THEOREMS 1.3 AND 1.4]). Instead of assuming $\Delta_n(\Phi) > C^{-n}$, it is enough to know that there are not too many pairs of n-fold compositions of maps in Φ whose translation components are closer than C^{-n}. Likewise in the proof of Theorem 10, we work with a similarly relaxed version of (5).

To properly quantify this, we use entropy. Let X be a bounded real valued random variable and let $r \in \mathbf{R}_{>0}$. The entropy of X at scale r is defined as

$$H(X;r) = H\big(\lfloor r^{-1}X \rfloor\big),$$

where $H(\cdot)$ on the right is Shannon entropy. This is the entropy of X with respect to a partition of $\mathbf{R}$ into consecutive intervals of length r. The choice of this partition is not canonical, and we obtain different values of $H(X;r)$ by translating X. There are advantages of averaging over translations of X in the definition of $H(X;r)$, as it is done, e.g., in [8, 53] and subsequent papers; however, we ignore this point here for the sake of simplicity.

By definition, $\Delta_n(\Phi_\lambda) > r$ implies that the points in the support of $\sum_{j=0}^{n-1} \xi_j \lambda^j$ are separated by a distance of at least r, hence

$$H\left(\sum_{j=0}^{n-1} \xi_j \lambda^j; r\right) = \log(2) \cdot n.$$

In the proof of Theorem 10, instead of working with lower bounds on $\Delta_n(\Phi_\lambda)$ like (5), we work with bounds of the type

$$H\left(\sum_{j=0}^{n-1} \xi_j \lambda^j; r\right) \geq \beta n \tag{6}$$

with suitable β and r.

Now consider some $\lambda > 1/2$ that lacks the approximations η_n as described in the conclusion in Theorem 10. We discuss how this assumption can be used to show that bounds of the type (6) hold for suitably many different values of n. Using such bounds and arguments based on Hochman's proof of Theorem 4, which we do not discuss in this paper, it can be shown that $\dim \nu_\lambda = 1$ proving (the contrapositive of) Theorem 10.

In short, the failure of (6) with a suitably small r implies that λ can be approximated by some $\eta_n \in \mathscr{E}^{(n)}$ such that Φ_{η_n} has enough exact overlaps to force $\dim \nu_{\eta_n} \leq \beta / \log \eta_n^{-1}$.

We give some more details. For every pair of numbers x_1, x_2 in the support of $\sum_{j=0}^{n-1} \xi_j \lambda^j$ such that $|x_1 - x_2| \leq r$, there is a polynomial $P \in \mathscr{P}^{(n)}$ such that

$$|x_1 - x_2| = \big|P(\lambda)\big| \leq r.$$

As we have already seen, all such polynomials have a root near λ provided $r < C^{-n}$ for a suitable constant C. If $r < \exp(-Cn \log n)$ for another suitable C, then all the roots obtained this way as (x_1, x_2) goes over all pairs of points in the support of $\sum_{j=0}^{n-1} \xi_j \lambda^j$ that are at distance not more than r can be shown to coincide. This follows from Mahler's aforementioned bound on the separation of elements in $\mathcal{E}^{(n)}$. For an alternative argument, see [8, SECTION 3].

Now it follows that if

$$H\left(\sum_{j=0}^{n-1} \xi_j \lambda^j ; r\right) < \log(2) \cdot n$$

for some $r < \exp(-Cn \log n)$, then there is some $\eta_n \in \mathcal{E}^{(n)}$ close to λ (the common root of the polynomials discussed in the previous paragraph) such that

$$H\left(\sum_{j=0}^{n-1} \xi_j \eta_n^j\right) \leq H\left(\sum_{j=0}^{n-1} \xi_j \lambda^j ; r\right). \tag{7}$$

Notice that on the left there is no designated scale, so $H(\cdot)$ stands for Shannon entropy there. Provided $H(\sum_{j=0}^{n-1} \xi_j \lambda^j ; r)$ is sufficiently small, this can be turned into a bound on $\dim \nu_{\eta_n}$ with the help of Theorem 8. Indeed, combining our observations, we see that

$$H\left(\sum_{j=0}^{n-1} \xi_j \lambda^j ; r\right) \leq \beta n$$

implies

$$\dim \nu_{\eta_n} \leq \frac{h(\Phi_{\eta_n})}{\log \eta_n^{-1}} \leq \frac{H(\sum_{j=0}^{n-1} \xi_j \eta_n^j)}{n \log \eta_n^{-1}} \leq \frac{H(\sum_{j=0}^{n-1} \xi_j \lambda^j ; r)}{n \log \eta_n^{-1}} \leq \frac{\beta}{\log \eta_n^{-1}}.$$

By the assumption that λ lacks the approximations as in the conclusion of Theorem 10, we conclude $|\lambda - \eta_n| > \exp(-n^{100})$. As we have already discussed, this implies that we can find an $\tilde{n}$ not larger than n^{100} such that even (5) holds with $\tilde{n}$ in place of n. This provides a sufficiently plentiful supply of numbers n such that at least a bound of the type (6) holds.

We return to the proof of Theorem 7. We suppose to the contrary that $\lambda \in (1/2, 1)$ is a counterexample to Conjecture 3. By Theorem 10, there are infinitely many approximants η_n to λ satisfying the conclusion of that theorem. We fix such an η_n corresponding to a suitably large n.

By virtue of (4), we have $h(\Phi_{\eta_n}) \leq (1 - \delta) \log \eta_n$. Our next step is to convert this information to something that is easier to exploit with the methods of Diophantine Approximation. We introduce a definition for this purpose. The Mahler measure of an algebraic number η with minimal polynomial $a_d(x - \eta^{(1)}) \cdots (x - \eta^{(d)}) \in \mathbf{Z}[x]$ is defined as

$$M(\eta) = |a_d| \prod_{j=1}^{d} \max\bigl(1, \bigl|\eta^{(j)}\bigr|\bigr),$$

i.e., it is the product of the absolute values of the leading coefficient and the roots outside the unit disk. This quantity is widely used in number theory as a measure of the "complexity" of η. Notice that if $\eta \in \mathbf{Q}$, then $M(\eta)$ is the maximum of the absolute values of the numerator and the denominator of η.

Breuillard and Varjú [9] found a connection between the entropy rate and the Mahler measure. A form of this most suited for the proof of Theorem 7 is the following.

Theorem 11 (Breuillard, Varjú). *For any $h \in (0, \log 2)$, there is a number $C(h)$ such that $h(\Phi_\eta) \leq h$ implies $M(\eta) < C(h)$ for all algebraic numbers η.*

See [54, THEOREM 9] for the details of how this follows from the technical results of [9].

Using this theorem, we conclude that $M(\eta_n) < C$ for a constant C that only depends on λ, but not on n. Furthermore, recall that we have $|\lambda - \eta_n| < \exp(-n^{100})$. Now we use the following, which follows easily from a more general result of Mignotte [34].

Theorem 12 (Mignotte). *Let η be an algebraic number of degree at most n. Let $\tilde{n} > n(\log n)^2$ be an integer, and let $\widetilde{\eta} \neq \eta \in \mathcal{E}^{(\tilde{n})}$. Then there is an absolute constant C, such that*

$$|\eta - \widetilde{\eta}| \geq C^{-\tilde{n}} M(\eta)^{-2\tilde{n}}.$$

We finish our discussion of the proof of Theorem 7. Thanks to the approximation of λ by η_n this theorem acts as a substitute for the separation condition between elements of $\mathcal{E}^{(\tilde{n})}$ in the proof of Lemma 9, and we can conclude that $\Delta_{\tilde{n}}(\Phi_\lambda) > C^{-\tilde{n}}$ for a suitable choice of $\tilde{n}$ for some C independent of n. Now we are in a position to apply Theorem 4 to show that Conjecture 3 holds for λ, which is our desired contradiction proving Theorem 7.

The original argument in [54] used an alternative variant of Theorem 12, which was deduced from an observation of Garsia [20] and a transversality argument of Solomyak [49]. It was pointed out by Vesselin Dimitrov that the transversality argument can be replaced by a simpler version based on Jensen's formula. This has the advantage that it is applicable in greater generality. See [41, LEMMATA 2.3 AND 4.6] for details.

3. FAILURE OF EXPONENTIAL SEPARATION

As we discussed in the previous section, it is not known whether Bernoulli convolutions without exact overlaps satisfy the exponential separation property. However, they are known to satisfy a slightly weaker lower bound on Δ_n, and this played an important role in the proof of Conjecture 3 for this class of IFS's.

On the other hand, there are some IFS's without exact overlaps for which it is known that the exponential separation property fails, and moreover, Δ_n converges to 0 in an arbitrarily fast prescribed way.

Theorem 13 (Baker; Bárány, Käenmäki). *Let $(\eta_n) \subset \mathbf{R}_{>0}$. Then there is an IFS Φ without exact overlaps such that $\Delta_n(\Phi) \leq \eta_n$ for all n.*

The first examples of such IFSs were given by Baker [4] in the form

$$\left\{x \mapsto \frac{x}{2}, x \mapsto \frac{x+1}{2}, x \mapsto \frac{x+s}{2}, x \mapsto \frac{x+t}{2}, x \mapsto \frac{x+1+s}{2}, x \mapsto \frac{x+1+t}{2}\right\}$$

for suitable choices of the parameters t, s, and by Bárány, Käenmäki [5] in the form

$$\{x \mapsto \lambda x, x \mapsto \lambda x + 1, x \mapsto \lambda x + t\}$$

for suitable choices of λ, t. Baker's example was modified by Chen [11], who disposed of the last two maps and replaced the denominator 2 by an arbitrary real algebraic number not smaller than 2. These constructions were further extended by Baker [3].

In what follows we give a heuristic argument to show why such IFSs with very small separation may be expected to exist. Our purpose (due to limitation of space) is not to give insight to the proofs of Theorem 13, which are based on a variety of tools, such as continued fraction expansions in [4] and the transversality method in [5]. Instead, we just aim to highlight the difference between families of IFSs depending on a single parameter, such as Bernoulli convolutions, or the examples covered by Corollary 6, and families depending on more than one parameter, which will be discussed in the next two sections.

Let

$$\Phi_{x,y} = \{\varphi_{i,x,y} : i \in \Lambda\}$$

be a family of IFS's (smoothly) depending on two parameters. Let $n \in \mathbf{Z}_{>0}$, and we write $\Gamma^{(n)}$ for the collection of curves in the parameter space, which arise as the solution sets of equations of the form

$$\varphi_{i_1,x,y} \circ \cdots \circ \varphi_{i_n,x,y} = \varphi_{\tilde{i}_1,x,y} \circ \cdots \circ \varphi_{\tilde{i}_n,x,y}$$

in (x, y) where $i_1, \dots, i_n$ and $\tilde{i}_1, \dots, \tilde{i}_n$ are two distinct sequences of indices in Λ. Note that the union of all these curves is the set of all parameter points for which the IFS contains exact overlaps.

The key difference between this setting and a family depending on a single parameter is that exact overlaps occur along curves in the parameter space rather than at isolated points. These curves may intersect each other, and then there is no separation between them, which rules out the arguments presented for the proof of Corollary 6 and later in Section 2.

We now give the heuristic suggesting the existence of the IFS's claimed in Theorem 13. We give a recursive construction. After the kth step, we will have a sequence $n_1, \dots, n_k \in \mathbf{Z}_{\geq 1}$, a sequence $\gamma_1, \dots, \gamma_k$, where γ_j is a segment of a curve in $\Gamma^{(n_j)}$, and a sequence $\delta_1, \dots, \delta_{k-1} \in \mathbf{R}_{>0}$. These will satisfy the property that γ_k is contained in the δ_j neighborhood of γ_j for all $j < k$.

We begin the process by setting γ_1 to be any segment (of positive length) of a curve in $\Gamma^{(1)}$. Suppose now that $\gamma_1, \dots, \gamma_k$ and $\delta_1, \dots, \delta_{k-1}$ are given for some $k \geq 1$. We choose a curve $\widetilde{\gamma}_{k+1} \in \Gamma^{(n_{k+1})}$ for some $n_{k+1} > n_k$ that intersects γ_k. The existence of such a curve is plausible, but requires proof, and this is why this construction is only a heuristic. We observe that $\Delta_n(\Phi_{x,y}) = 0$ for all $n \geq n_k$ and $(x, y) \in \gamma_k$. By continuity, there is a choice of δ_k so that $\Delta_n(\Phi_{x,y}) \leq \eta_n$ holds for all $n \in [n_k, n_{k+1})$ and (x, y) in the δ_k neighborhood of γ_k. Finally, we set γ_{k+1} to be a suitable segment of $\widetilde{\gamma}_{k+1}$ contained in the δ_j neighborhood of γ_j for all $j \leq k$.

It is immediate from the construction that there is a point (x, y) which is contained in the (closed) δ_k neighborhood of γ_k for all k, and that $\Delta_n(\Phi_{x,y}) \leq \eta_n$ for all n.

With a small modification of the construction, we can ensure that $\Phi_{x,y}$ contains no exact overlaps for the resulting parameter point (x, y). Indeed, observe that $\bigcup \Gamma^{(n)}$ is a countable set, and let $\gamma_1^*, \gamma_2^*, \dots$ be an enumeration of it. In the construction, we have considerable liberty in choosing the curve segment γ_k so we can make sure that it does not intersect γ_k^*. (This requires, in particular, that we choose $\widetilde{\gamma}_k$ not to coincide with γ_k^*. The possibility of this is again plausible, but requires proof.) Then in the next step of the construction, we can ensure that δ_k is chosen to be sufficiently small so that γ_k^* is entirely outside the δ_k neighborhood of γ_k. This way we can ensure that the resulting parameter point (x, y) at the end of the process is not contained in γ_k^* for any k, and hence $\Phi_{x,y}$ is without exact overlaps.

4. IFSS WITH ALGEBRAIC CONTRACTION FACTORS

In this section we discuss the following result of Rapaport [39].

Theorem 14 (Rapaport). *Conjectures 2 and 3 hold for all IFSs in which all contraction parameters are algebraic numbers.*

This is a far reaching common generalization of Hochman's Corollaries 5 and 6. We discuss some of the main ideas in the special case of the family of IFSs

$$\Phi_{s,t} = \left\{ x \mapsto \frac{x}{3}, x \mapsto \frac{x}{3} + 1, x \mapsto \frac{x}{3} + s, x \mapsto \frac{x}{3} + t \right\}$$

with uniform probability weights. This is perhaps the simplest family not contained in the results of Hochman, and as was shown by Chen (see Section 3), this family contains IFSs without exact overlaps that fail the exponential separation property (in a very strong sense).

Let $\xi_1, \xi_2, \dots$ be a sequence of independent random variables taking the values $0, 1, s, t$ with equal probabilities. As we discussed in Section 2, the exponential separation property can be relaxed in Hochman's results. Instead of a lower bound on Δ_n, it suffices to have bounds of the form

$$H\left(\sum_{j=0}^{n-1} \xi_j \cdot 3^{-j}; C^{-n}\right) \geq (\log 3 - \varepsilon_n)n \tag{8}$$

for infinitely many values of n with some constant C and a sequence $\varepsilon_n \to 0$. (See Section 2 for the definition of this notation.)

Theorem 14 is proved by verifying condition (8). With this aim in mind, we examine what happens when (8) fails for some n, C and ε_n. We write $\mathcal{L}^{(n)}$ for the family of (inhomogeneous) linear forms of the form $a_1 \cdot 1 + a_2 Y_1 + a_3 Y_2$, where each a_i is a sum of a subset of the numbers $1, 3^{-1}, \dots, 3^{-n+1}$ and each term 3^j is allowed in at most one of the a_i. This definition is designed so that the values taken by the random variable $\sum_{j=0}^{n-1} \xi_j \cdot 3^{-j}$ are precisely the values of the linear forms in $\mathcal{L}^{(n)}$ evaluated at s and t.

We write $\mathcal{L}^{(n)} - \mathcal{L}^{(n)}$ for the set of linear forms that can be written as the difference of two elements of $\mathcal{L}^{(n)}$. We also fix some parameter point (s_0, t_0) such that the IFS lacks exact overlaps. We consider pairs of elements in the support of $\sum_{j=0}^{n-1} \xi_j \cdot 3^{-j}$ that are at distance no more than C^{-n}. Then for any such pair, there corresponds a linear form $L \in \mathcal{L}^{(n)} - \mathcal{L}^{(n)}$ such that $|L(s_0, t_0)| \leq C^{-n}$. We write $\mathcal{A}^{(n)}$ for the collection of linear forms in $\mathcal{L}^{(n)} - \mathcal{L}^{(n)}$ that arise in this way. (This definition depends on C, s_0 and t_0, which we suppress in our notation.)

Let n be such that (8) fails (for some choice of ε_n and C). We distinguish two cases depending on the rank of $\mathcal{A}^{(n)}$. The first case arises when there are at least two linearly independent forms in $\mathcal{A}^{(n)}$, and the second case is when the elements of $\mathcal{A}^{(n)}$ are all scalar multiples of each other.

In the first case, we take two linearly independent $L_1, L_2 \in \mathcal{L}^{(n)} - \mathcal{L}^{(n)}$. Provided C is sufficiently large, the lines determined by L_1 and L_2 cannot be parallel. Indeed, if that was the case, their distance would be a rational number with denominator bounded by an exponential in n, which we can force to be 0 by taking C sufficiently large. Since the lines are not parallel, we can solve the equations

$$\begin{aligned} L_1(s_n, t_n) &= 0, \\ L_2(s_n, t_n) &= 0, \end{aligned}$$

and find that their solution (s_n, t_n) is a pair of rational numbers with denominators bounded by an exponential in n. Moreover, the distance of (s_n, t_n) from (s_0, t_0) will be an arbitrarily small exponential in n if C is chosen sufficiently large.

The points (s_n, t_n) have the same repellency property as those in the proof of Corollary 6. We discuss next how to show that the second case, that is when the elements of $\mathcal{A}^{(n)}$ are proportional, arises for only finitely many values of n. Then the argument for Corollary 6 can be carried over to prove (8).

We begin by extending the definition of entropy rates. Let ℓ be a line in $\mathbf{R}^2$ (that does not necessarily contain 0). We denote by $Y_\ell^{(n)}$ the random $\ell \to \mathbf{R}$ function $(s,t) \mapsto \sum_{j=0}^{n-1} \xi_j(s,t) \cdot 3^{-j}$. We define the entropy rate of the line ℓ by

$$h(\ell) := \lim_{n\to\infty} \frac{H(Y_\ell^{(n)})}{n}.$$

Here $H(Y_\ell^{(n)})$ stands for the Shannon entropy of $Y_\ell^{(n)}$, which is a random element taking finitely many values. It can be shown that $H(Y_\ell^{(n)})$ is subadditive, hence the limit exists and is equal to the infimum. The quantity $h(\ell)$ measures the amount of exact overlaps that occur simultaneously for all parameter points $(s,t) \in \ell$.

Now suppose that the second case occurs for some n in our above discussion, that is the linear forms in $\mathcal{A}^{(n)}$ are proportional. Let ℓ be the line on which all elements of $\mathcal{A}^{(n)}$ vanish. It is immediate from the definition of $\mathcal{A}^{(n)}$ that

$$H(Y_\ell^{(n)}) \leq H\left(\sum_{j=0}^{n} \xi_j 3^{-j}; C^{-n}\right).$$

Supposing

$$H\left(\sum_{j=0}^{n} \xi_j 3^{-j}; C^{-n}\right) \leq (\log 3 - \varepsilon)n \tag{9}$$

for some $\varepsilon > 0$, we can conclude

$$h(\ell) \leq \log 3 - \varepsilon.$$

In light of all this, the next proposition—implicit in [39]—implies that the second case and (9) for some fixed $\varepsilon > 0$ may occur for only finitely many n's.

Proposition 15. *Let (s_0, t_0) be some parameters such that the IFS Φ_{s_0,t_0} contains no exact overlaps. Fix some $\varepsilon > 0$. Then there is a neighborhood of (s_0, t_0) that is not intersected by any lines ℓ with $h(\ell) \leq \log 3 - \varepsilon$.*

We end this section by discussing the proof of this result. Suppose to the contrary that the result is false, that is, there is a sequence $\ell_1, \ell_2, \dots$ of lines passing closer and closer to (s_0, t_0) with $h(\ell_n) < \log 3 - \varepsilon$. We suppose as we may that the lines ℓ_n converge (in any reasonable topology) to a line ℓ_∞. We also suppose for simplicity that none of $\ell_1, \ell_2, \dots, \ell_\infty$ is parallel to either of the s or t axes, and none of them goes through the origin.

We associate a self-similar measure in $\mathbf{R}^2$ to each line ℓ_j. For $j = 1, 2, \dots, \infty$, let σ_j and τ_j be the unique numbers such that ℓ_j is spanned by $(\sigma_j, 0)$ and $(0, \tau_j)$. For $\sigma, \tau \in \mathbf{R}$, we define the IFS

$$\Psi_{\sigma,\tau} := \left\{(x, y) \mapsto \left(\frac{x}{3}, \frac{y}{3}\right), (x, y) \mapsto \left(\frac{x}{3} + 1, \frac{y}{3} + 1\right),\right.$$
$$\left.(x, y) \mapsto \left(\frac{x}{3} + \sigma, \frac{y}{3}\right), (x, y) \mapsto \left(\frac{x}{3}, \frac{y}{3} + \tau\right)\right\},$$

and write $\nu_{\sigma,\tau}$ for the associated self-similar measure (with equal probability weights).

It is immediate from the definitions that the same exact overlaps occur for the random variables $Y_{\ell_j}^{(n)}$ as for the IFS $\Psi(\sigma_j, \tau_j)$. It follows that

$$h(\Psi_{\sigma_j,\tau_j}) = h(\ell_j) \leq \log 3 - \varepsilon$$

for $j < \infty$. Using this, it can be shown that

$$\dim \nu_{\sigma_j,\tau_j} \leq \frac{\log 3 - \varepsilon}{\log 3} = 1 - \varepsilon / \log 3.$$

It is a general phenomenon that the dimension of self-similar measures depends lower semicontinuously on the parameters, see, e.g., [16] for results of this type covering even self-affine measures. Using this, it follows that

$$\dim \nu_{\sigma_\infty,\tau_\infty} \leq 1 - \varepsilon / \log 3.$$

The proof of Proposition 15 is now finished by establishing a suitable analogue of Conjecture 3 for the IFSs $\Psi_{\sigma,\tau}$, which shows that Ψ_{σ_0,θ_0} and hence $\Phi_{s,t}$ for all $(s, t) \in \ell$ including (s_0, t_0) contains exact overlaps. This can be done along the lines of the proof of Corollary 5 discussed in Section 1 using a higher dimensional version of Hochman's theorem, which can be found in [22]. The crucial difference between the IFSs $\Phi_{s,t}$ and $\Psi_{\sigma,\tau}$ is that the ambient space is 2-dimensional for the latter and this matches the number of parameters. This means that exact overlaps occur at single points (as opposed to along lines), which have the required repellency property.

5. HOMOGENEOUS IFSS OF THREE MAPS

In this section, we discuss the IFSs

$$\Phi_{\lambda,t} = \big\{(x \mapsto \lambda x, x \mapsto \lambda x + 1, x \mapsto \lambda x + t)\big\}.$$

Rapaport and Varjú [41] made some partial progress towards extending the results for Bernoulli convolutions discussed in Section 2 to this setting and to some more general IFSs (see [41, SECTION 3]).

Before we can state these results, we need to introduce some relevant notation and terminology. We write $\mu_{\lambda,t}$ for the self-similar measure associated to the IFS $\Phi_{\lambda,t}$ and uniform probability weights. Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random $\mathbf{R} \to \mathbf{R}$ functions taking the values $t \mapsto 0$, $t \mapsto 1$ and $t \mapsto t$ with equal probability. Let $U \subset (0,1) \times \mathbf{R}$, $n \in \mathbf{Z}_{\geq 0}$, and write $A_U^{(n)}$ for the random $U \to \mathbf{R}$ function

$$(\lambda, t) \mapsto \sum_{j=1}^{n} \xi_j(t)\lambda^j.$$

We define the entropy rate

$$h(U) := \lim_{n\to\infty} \frac{H(A_U^{(n)})}{n} = \inf \frac{H(A_U^{(n)})}{n}.$$

We abbreviate $A^{(n)}_{\{\lambda,t\}}$ as $A^{(n)}_{\lambda,t}$, and $h(\{\lambda,t\})$ as $h(\lambda,t)$. One should think about $h(\lambda,t)$ as a quantity expressing the amount of exact overlaps contained in the IFS $\Phi_{\lambda,t}$ and $h(U)$ aims to quantify the amount of exact overlaps occurring simultaneously for the parameter points in U.

We write $\mathscr{R}$ for the set of meromorphic functions on the unit disc that can be written as ratios of two power series with coefficients $-1, 0, 1$. We denote by Γ the set of curves $\gamma \subset (0,1) \times \mathbf{R}$ that are either of the following two forms:

- $\gamma = \{(\lambda, t) \in (0,1) \times \mathbf{R} : t = R(\lambda)\}$ for some $R \in \mathscr{R}$,
- $\gamma = \{(\lambda_0, t) : t \in \mathbf{R}\}$ for some fixed $\lambda_0 \in (0,1)$.

It can be shown that exact overlaps occur in the family of IFSs $\Phi_{\lambda,t}$ along finite unions of curves in Γ, but not all elements of Γ arises in this way.

The next result is an analogue of Theorem 10 in the setting of the IFS $\Phi_{\lambda,t}$.

Theorem 16 (Rapaport, Varjú). *Suppose that Conjecture 3 does not hold for the IFS $\Phi_{\lambda,t}$ for some choice of parameters λ and t. Then for every $\varepsilon > 0$ and $N \geq 1$, there exist $n \geq N$ and $(\eta, s) \in (0,1) \times \mathbf{R}$ such that*

(1) $|\lambda - \eta|, |t - s| \leq \exp(-n^{\varepsilon^{-1}})$,

(2) $\frac{1}{n \log \eta^{-1}} H(A^{(n)}_{\eta,s}) \leq \dim \mu_{\lambda,t} + \varepsilon$,

(3) $h(\gamma) \geq \min\{\log 3, \log \lambda^{-1}\} - \varepsilon$ *for all* $\gamma \in \Gamma$ *with* $(\eta, s) \in \gamma$.

Item (2) in the conclusion means that the IFS $\Phi_{\eta,s}$ contains enough overlaps after n iteration to force the dimension of $\mu_{\eta,s}$ below $\dim \mu_{\lambda,t} + \varepsilon$. Item (3) in the conclusion implies that not all of these exact overlaps occur along the same curve γ. From these properties it can be deduced in particular that η and s are algebraic numbers and roots of polynomials of low degree with small integer coefficients. (For a precise statement, see [41, THEOREM 1.3].) This yields a bound on the number of possible points that can arise as (η, s) in the conclusion and together with Item (1), this shows that the Hausdorff dimension of the set of exceptional parameters for which Conjecture 3 fails is 0. This improves Hochman's bound, which is 1, albeit that bound is given for the stronger notion of packing dimension, which may exceed the Hausdorff dimension.

It is still an open problem whether an analogue of Theorem 11 holds for the IFS $\Phi_{\lambda,t}$. One possible formulation is the following.

Question 17. Is it true that for all $\varepsilon > 0$, there is M such that the following holds? Let $(\lambda, t) \in (\varepsilon, 1-\varepsilon) \times \mathbf{R}$ be such that $h(\lambda, t) \le \min(\log 3, \log \lambda^{-1}) - \varepsilon$ and $h(\gamma) \ge \min(\log 3, \log \lambda^{-1}) - M^{-1}$ for all $\gamma \in \Gamma$ with $(\lambda, t) \in \gamma$. Then $M(\lambda) \le M$.

We note that a condition about the entropy rate of curves passing through (λ, t) is necessary. Indeed, we have, for example, $h(\gamma) = \log 3 - (2/3) \log 2$ for the curve $\gamma = \{(\lambda, 1) : \lambda \in (0, 1)\}$, and hence $h(\lambda, 1) \le \log 3 - (2/3) \log 2$ for all $\lambda \in (0, 1)$.

We also have the following conditional result towards Conjecture 3.

Theorem 18 (Rapaport, Varjú). *Suppose that the answer to Question 17 is affirmative. Then Conjecture 3 holds for the IFS $\Phi_{\lambda,t}$ with equal probability weights for all $\lambda \in (0,1)$ and $t \in \mathbf{R}$.*

Using ideas from [9], one can answer Question 17 affirmatively if we restrict λ to be near 1. This allows for the following unconditional partial resolution of Conjecture 3.

Theorem 19 (Rapaport, Varjú). *Conjecture 3 holds for the IFS $\Phi_{\lambda,t}$ with equal probability weights for all $(\lambda, t) \in (2^{-2/3}, 1) \times \mathbf{R}$.*

The key new ingredient in the proof of Theorem 16 compared to that of Theorem 10 is the following result, whose role is similar to that of Proposition 15 in the proof of Theorem 14.

Proposition 20. *Let $(\lambda, t) \in (0,1) \times \mathbf{R}$ be such that the IFS $\Phi_{\lambda,t}$ contains no exact overlaps. Then for all $h < \min(\log \lambda^{-1}, \log 3)$, there is a neighborhood of (λ, t) that is not intersected by a curve $\gamma \in \Gamma$ with $h(\gamma) \le h$.*

The proof of this result like Proposition 15 is done by attaching suitable fractal objects to curves and relating their dimension to the entropy rates of the curves. Then the proposition is proved using lower semicontinuity of dimension and a limiting argument. The fractal measures used in the paper [41] are analogues of self-similar measures in function fields. A suitable notion of dimension is introduced for these objects and Hochman's theorem is generalized to this setting. The analogue of the exponential separation property is verified using an argument similar to that used in the proof of Corollary 6. An additional difficulty compared to the setting of Section 4 is caused by the fact that the curves in Γ are not necessarily lines and they may develop singularities, which complicates limiting arguments.

The proofs of Theorems 18 and 19 is complicated by the fact that like in the case of Bernoulli convolutions, the parameter points with exact overlaps have a weaker than exponential repellency property. To address this, an argument similar to that discussed at the end of Section 2 is used. This is the reason why we need to assume an affirmative answer to Question 17. The argument also requires a stronger form of Proposition 20 with a modified entropy rate. The precise statement requires some preparation. For this reason, we omit it and refer to [41, PROPOSITION 2.4].

6. OTHER DEVELOPMENTS

We survey some recent results about aspects of self-similar measures other than their dimensions. Due to limitation of space, our discussion will be very brief.

6.1. Fourier decay

We first discuss Fourier decay of self-similar measures. Specifically, we discuss the following three properties:

- A measure μ on $\mathbf{R}$ is Rajchman if its Fourier transform vanishes at infinity, that is,
$$\lim_{|\xi|\to\infty} \left|\widehat{\mu}(\xi)\right| = 0.$$
- A measure μ on $\mathbf{R}$ has polylogarithmic Fourier decay if there is a constant $a > 0$ such that for all sufficiently large ξ, we have
$$\left|\widehat{\mu}(\xi)\right| < \left|\log|\xi|\right|^{-a}.$$
- A measure μ on $\mathbf{R}$ has power Fourier decay if there is a constant $a > 0$ such that for all sufficiently large ξ, we have
$$\left|\widehat{\mu}(\xi)\right| < |\xi|^{-a}.$$

There are various motivations for studying these properties. The Rajchman property is closely related to an old subject in the theory of trigonometric series about the so-called sets of uniqueness and sets of multiplicity, see [27] for more. Fourier decay has also applications in metric Diophantine approximation. For example, polylogarithmic Fourier decay is sufficient to guarantee that almost all numbers with respect to the measure are normal in every bases. (In the case of self-similar measures on $\mathbf{R}$, even the Rajchman property is enough for this, see [2, THEOREM 1.4].) Power decay is very useful in proving absolute continuity of the measure, which we discuss more in the next section.

Results about these properties of self-similar measures come in two flavors. In the first category, properties are proved for most self-similar measures in a parametric family, in the second the properties are proved for explicit self-similar measures, that is, the hypotheses of the results are testable in concrete examples.

We begin by discussing results in the first category. Erdős [14] proved that Bernoulli convolutions (see Section 2) have power Fourier decay for almost all choices of the parameter $\lambda \in (0,1)$. His argument was revisited by Kahane [26] who showed that the exceptional set of parameters where the power decay fails is, in fact, of 0 Hausdorff dimension. This method was exposed in the survey [38], where the exponent a was also studied, and the term Erdős–Kahane argument was coined. Recently Solomyak [50] showed that nondegenerate self-similar measures on $\mathbf{R}$ have power Fourier decay if the vector of contraction parameters avoid an exceptional set of 0 Hausdorff dimension. See the references in [50, SECTION 1.1] and [51] for more recent applications of the Erdős–Kahane method.

The first results in the second category are also in the setting of Bernoulli convolutions. Erdős [13] proved that Bernoulli convolutions are not Rajchman when λ^{-1}, the reciprocal of the parameter, is a Pisot number, except when the probability weights are uniform and $\lambda = 1/2m$ for an odd integer m. Recall that a Pisot number is an algebraic integer all of whose Galois conjugates lie inside the complex unit disk. Salem [43] proved the converse of Erdős result by showing that Bernoulli convolutions are Rajchman when λ^{-1} is not Pisot.

The Rajchman property of general self-similar measures has been understood more recently. Sahlsten and Li [31] proved that self-similar measures are Rajchman whenever the semigroup generated by the contraction parameters is not lacunary, that is, it is not contained in $\{\lambda^n : n \in \mathbf{Z}_{\geq 0}\}$ for some n. Their work is based on a new method relying on renewal theory originating in [30]. See also [2], where this result is extended to self-conformal measures using a different method. The lacunary case was analyzed by Brémont [7], see also Varjú, Yu [55]. Finally, the problem was solved by Rapaport [40] for self-similar measures on $\mathbf{R}^d$.

For Bernoulli convolutions, polylogarithmic Fourier decay follows from a result of Bufetov and Solomyak [10, PROPOSITION 5.5] for algebraic parameters λ provided λ^{-1} is neither Salem nor Pisot, that is, it has another Galois conjugate outside the complex unit disk, see also [19]. Under a mild Diophantine condition for the contraction parameters, Sahlsten and Li [31] proved polylogarithmic Fourier decay for self-similar measures. Informally speaking, their condition requires that the semigroup generated by the contraction parameters is not approximated by lacunary semigroups in a suitable quantitative sense. See [2] for a similar result under a different Diophantine condition. Polylogarithmic Fourier decay was also established by Varjú and Yu [55] for certain self-similar measures in the lacunary case.

It is an important open problem to characterize which self-similar measures have power Fourier decay. Very little is known about this. See [12] for explicit examples of Bernoulli convolutions with power Fourier decay and [32] for results about self-similar measures on $\mathbf{R}^d$ for $d \geq 3$.

6.2. Absolute continuity

Let μ be a self-similar measure on $\mathbf{R}$ associated to an IFS with contraction factors $\{\lambda_i\}$ that contains no exact overlaps, and probability weights $\{p_i\}$. One may expect that μ is not only of dimension 1 if

$$\frac{\sum p_i \log p_i^{-1}}{\sum p_i \log \lambda_i^{-1}} > 1, \tag{10}$$

as predicted by Conjecture 3, but it is also absolutely continuous. When there is equality in (10), the self-similar measure is almost always singular, see [37, THEOREM 1.1].

In general, this expectation is false. Simon and Vágó [48] showed that in some families of IFSs, there is a dense G^d set of parameters, which violate the above statement. See [36] for earlier related results in a different setting. However, it could still be true that (10) and the lack of exact overlaps imply absolute continuity for some families of self-similar measures, for example for Bernoulli convolutions.

Nevertheless, it is expected that self-similar measures are absolutely continuous for almost all choices of the parameters in parametric families when (10) holds. For Bernoulli convolutions, this was proved by Erdős for λ near 1, as a consequence of power Fourier decay with parameter $a > 1$. The result has been extended to the optimal range $\lambda \in [1/2, 1]$ by Solomyak [49] using the transversality method. See [6,37,38] and their references for further developments. Shmerkin [44] proved that the set of exceptional parameters in $[1/2, 1]$ that make the Bernoulli convolution singular is of Hausdorff dimension 0. His method is based on a result of his that the convolution of a measure of dimension 1 and another one with power Fourier decay is absolutely continuous. He used this in conjunction with Hochman's theorem and the Erdős-Kahane method. See [42,46] and the references therein for further developments using this method.

Explicit examples of absolutely continuous self-similar measures are rare. The first examples were given by Garsia [20] as the Bernoulli convolutions with parameters of Mahler measure 2. See [12] for a generalization of this construction, and see [56] for an improvement on the regularity of the density function using Shmerkin's method. Varjú gave new examples of absolutely continuous Bernoulli convolutions in [53]. This paper relies on a similar method to Hochman's in a quantitatively refined form. A crucial point is that it requires the separation condition to hold at all sufficiently small scales rather than just at infinitely many of them. This restricts the method to algebraic parameters currently. A recent improvement was given by Kittle [28], who gave further new examples of absolutely continuous Bernoulli convolutions. While all the new examples in [53] are very close to 1, e.g., $1 - 10^{-50}$, this is not the case for [28], which includes, e.g., one near 0.799533... The paper [28] also introduces a new tool to quantify the smoothness of measures at scales.

See [32] for results about absolute continuity of self-similar measures on $\mathbf{R}^d$ for $d \geq 3$.

ACKNOWLEDGMENTS

I am grateful to Simon Baker, Balázs Bárány, Emmanuel Breuillard, Michael Hochman, Antti Käenmäki, Samuel Kittle, Ariel Rapaport, Tuomas Sahlsten, Pablo Shmerkin, Boris Solomyak, Laurtiz Streck, and Han Yu for reading and commenting on an earlier version of this paper.

FUNDING

The author has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 803711). The author was supported by the Royal Society.

REFERENCES

[1] S. Akiyama, D.-J. Feng, T. Kempton, and T. Persson, On the Hausdorff dimension of Bernoulli convolutions. *Int. Math. Res. Not. IMRN* **19** (2020), 6569–6595.

[2] A. Algom, F. Rodriguez Hertz, and Z. Wang, Pointwise normality and fourier decay for self-conformal measures. 2021, arXiv:2012.06529v2.

[3] S. Baker, Iterated function systems with super-exponentially close cylinders II. 2020, arXiv:2007.11291v1. To appear in *Proc. Amer. Math. Soc.*

[4] S. Baker, Iterated function systems with super-exponentially close cylinders. *Adv. Math.* **379** (2021), 107548, 13 pp.

[5] B. Bárány and A. Käenmäki, Super-exponential condensation without exact overlaps. *Adv. Math.* **379** (2021), 107549, 22 pp.

[6] B. Bárány, K. Simon, B. Solomyak, and A. Śpiewak, Typical absolute continuity for classes of dynamically defined measures. 2021, arXiv:2107.03692v1.

[7] J. Brémont, Self-similar measures and the Rajchman property. 2020, arXiv:1910.03463v8.

[8] E. Breuillard and P. P. Varjú, On the dimension of Bernoulli convolutions. *Ann. Probab.* **47** (2019), no. 4, 2582–2617.

[9] E. Breuillard and P. P. Varjú, Entropy of Bernoulli convolutions and uniform exponential growth for linear groups. *J. Anal. Math.* **140** (2020), no. 2, 443–481.

[10] A. I. Bufetov and B. Solomyak, On the modulus of continuity for spectral measures in substitution dynamics. *Adv. Math.* **260** (2014), 84–129.

[11] C. Chen, Self-similar sets with super-exponential close cylinders. 2020, arXiv:2004.14037v1.

[12] X.-R. Dai, D.-J. Feng, and Y. Wang, Refinable functions with non-integer dilations. *J. Funct. Anal.* **250** (2007), no. 1, 1–20.

[13] P. Erdős, On a family of symmetric Bernoulli convolutions. *Amer. J. Math.* **61** (1939), 974–976.

[14] P. Erdős, On the smoothness properties of a family of Bernoulli convolutions. *Amer. J. Math.* **62** (1940), 180–186.

[15] K. Falconer, *Fractal geometry. Third edn.* John Wiley & Sons, Ltd., Chichester, 2014.

[16] D.-J. Feng, Dimension of invariant measures for affine iterated function systems. 2020, arXiv:1901.01691v2.

[17] D.-J. Feng and Z. Feng, Estimates on the dimension of self-similar measures with overlaps. 2021, arXiv:2103.01700v2.

[18] D.-J. Feng and H. Hu, Dimension theory of iterated function systems. *Comm. Pure Appl. Math.* **62** (2009), no. 11, 1435–1500.

[19] X. Gao and J. Ma, Decay rate of Fourier transforms of some self-similar measures. *Acta Math. Sci. Ser. B Engl. Ed.* **37** (2017), no. 6, 1607–1618.

[20] A. M. Garsia, Arithmetic properties of Bernoulli convolutions. *Trans. Amer. Math. Soc.* **102** (1962), 409–432.

[21] K. G. Hare, T. Kempton, T. Persson, and N. Sidorov, Computing Garsia entropy for Bernoulli convolutions with algebraic parameters. *Nonlinearity* **34** (2021), no. 7, 4744–4763.

[22] M. Hochamn, On self-similar sets with overlaps and inverse theorems for entropy in $\mathbb{R}^d$. 2017, arXiv:1503.09043v2. To appear in *Mem. Amer. Math. Soc.*

[23] M. Hochman, On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)* **180** (2014), no. 2, 773–822.

[24] M. Hochman, Dimension theory of self-similar sets and measures. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. III. Invited lectures*, pp. 1949–1972, World Sci. Publ., Hackensack, NJ, 2018.

[25] J. E. Hutchinson, Fractals and self-similarity. *Indiana Univ. Math. J.* **30** (1981), no. 5, 713–747.

[26] J.-P. Kahane, Sur la distribution de certaines séries aléatoires. In *Colloque de Théorie des Nombres (Univ. Bordeaux, Bordeaux, 1969)*, pp. 119–122, Soc. Math. France, 1971.

[27] A. S. Kechris and A. Louveau, *Descriptive set theory and the structure of sets of uniqueness*. London Math. Soc. Lecture Note Ser. 128, Cambridge University Press, Cambridge, 1987.

[28] S. Kittle, Absolute continuity of self similar measures. 2021, arXiv:2103.12684v1.

[29] V. Kleptsyn, M. Pollicott, and P. Vytnova, Uniform lower bounds on the dimension of Bernoulli convolutions. 2021, arXiv:2102.07714v2.

[30] J. Li, Decrease of Fourier coefficients of stationary measures. *Math. Ann.* **372** (2018), no. 3–4, 1189–1238.

[31] J. Li and T. Sahlsten, Trigonometric series and self-similar sets. 2021, arXiv:1902.00426v3. To appear in *J. Eur. Math. Soc. (JEMS).*

[32] E. Lindenstrauss and P. P. Varjú, Random walks in the group of Euclidean isometries and self-similar measures. *Duke Math. J.* **165** (2016), no. 6, 1061–1127.

[33] K. Mahler, An inequality for the discriminant of a polynomial. *Michigan Math. J.* **11** (1964), 257–262.

[34] M. Mignotte, Approximation des nombres algébriques par des nombres algébriques de grand degré. *Ann. Fac. Sci. Toulouse Math. (5)* **1** (1979), no. 2, 165–170.

[35] P. A. P. Moran, Additive functions of intervals and Hausdorff measure. *Proc. Camb. Philos. Soc.* **42** (1946), 15–23.

[36] F. Nazarov, Y. Peres, and P. Shmerkin, Convolutions of Cantor measures without resonance. *Israel J. Math.* **187** (2012), 93–116.

[37] S.-M. Ngai and Y. Wang, Self-similar measures associated to IFS with non-uniform contraction ratios. *Asian J. Math.* **9** (2005), no. 2, 227–244.

[38] Y. Peres, W. Schlag, and B. Solomyak, Sixty years of Bernoulli convolutions. In *Fractal geometry and stochastics, II (Greifswald/Koserow, 1998)*, pp. 39–65, Progr. Probab. 46, Birkhäuser, Basel, 2000.

[39] A. Rapaport, Proof of the exact overlaps conjecture for systems with algebraic contractions. 2020, arXiv:2001.01332v2. To appear in *Ann. Sci. Éc. Norm. Supér.*

[40] A. Rapaport, On the Rajchman property for self-similar measures on $\mathbb{R}^d$. 2021, arXiv:2104.03955v2.

[41] A. Rapaport and P. P. Varjú, Self-similar measures associated to a homogeneous system of three maps. 2021, arXiv:2010.01022v2.

[42] S. Saglietti, P. Shmerkin, and B. Solomyak, Absolute continuity of non-homogeneous self-similar measures. *Adv. Math.* **335** (2018), 60–110.

[43] R. Salem, Sets of uniqueness and sets of multiplicity. *Trans. Amer. Math. Soc.* **54** (1943), 218–228.

[44] P. Shmerkin, On the exceptional set for absolute continuity of Bernoulli convolutions. *Geom. Funct. Anal.* **24** (2014), no. 3, 946–958.

[45] P. Shmerkin, L^q dimensions of self-similar measures, and applications: a survey. In *New trends in applied harmonic analysis, Vol. 2*, pp. 257–292, Birkhäuser, Basel, 2019.

[46] P. Shmerkin, On Furstenberg's intersection conjecture, self-similar measures, and the L^q norms of convolutions. *Ann. of Math. (2)* **189** (2019), no. 2, 319–391.

[47] K. Simon, Overlapping cylinders: the size of a dynamically defined Cantor-set. In *Ergodic theory of $\mathbf{Z}^d$ actions (Warwick, 1993–1994)*, pp. 259–272, London Math. Soc. Lecture Note Ser. 228, Cambridge Univ. Press, Cambridge, 1996.

[48] K. Simon and L. Vágó, Singularity versus exact overlaps for self-similar measures. *Proc. Amer. Math. Soc.* **147** (2019), no. 5, 1971–1986.

[49] B. Solomyak, On the random series $\sum \pm\lambda^n$ (an Erdős problem). *Ann. of Math. (2)* **142** (1995), no. 3, 611–625.

[50] B. Solomyak, Fourier decay for self-similar measures. *Proc. Amer. Math. Soc.* **149** (2021), no. 8, 3277–3291.

[51] B. Solomyak, Fourier decay for homogeneous self-affine measures. 2021, arXiv:2105.08129v2.

[52] P. P. Varjú, Recent progress on Bernoulli convolutions. In *European Congress of Mathematics*, pp. 847–867, Eur. Math. Soc., Zürich, 2018.

[53] P. P. Varjú, Absolute continuity of Bernoulli convolutions for algebraic parameters. *J. Amer. Math. Soc.* **32** (2019), no. 2, 351–397.

[54] P. P. Varjú, On the dimension of Bernoulli convolutions for all transcendental parameters. *Ann. of Math. (2)* **189** (2019), no. 3, 1001–1011.

[55] P. P. Varjú and H. Yu, Fourier decay of self-similar measures and self-similar sets of uniqueness. 2021, arXiv:2004.09358v2. To appear in *Ann. PDE.*

[56] H. Yu, Bernoulli convolutions with Garsia parameters in (1, $\sqrt{2}$ have continuous density functions. 2021, arXiv:2108.01008.

PÉTER P. VARJÚ

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK, pv270@dpmms.cam.ac.uk

10. PARTIAL DIFFERENTIAL EQUATIONS

FORMATION AND DEVELOPMENT OF SINGULARITIES FOR THE COMPRESSIBLE EULER EQUATIONS

TRISTAN BUCKMASTER, THEODORE D. DRIVAS, STEVE SHKOLLER, AND VLAD VICOL

ABSTRACT

In this paper we review the authors' recent work [1] which gives a complete description of the formation and development of singularities for the compressible Euler equations in two space dimensions, under azimuthal symmetry. This solves an open problem posed by Landau and Lifshitz, which was previously open even in one space dimension. Our proof applies mutatis mutandis in the drastically simpler situations of one-dimensional flows, or multidimensional flows with radial symmetry. We prove that for smooth and generic initial data with azimuthal symmetry, the 2D compressible Euler equations yield a local in time smooth solution, which in finite time forms a first gradient singularity, the so-called $C^{1/3}$ *preshock*. We then show that a discontinuous entropy producing *shock wave* instantaneously develops from the preshock. Simultaneous to the development of the shock, two other characteristic surfaces of higher-order cusp-type singularities emerge from the preshock. These surfaces have been termed *weak discontinuities* by Landau and Lifshitz [17, CHAPTER IX, §96], who conjectured their existence. We prove that along the characteristic surface moving with the fluid, a *weak contact discontinuity* is formed, while along the slowest surface in the problem, a *weak rarefaction wave* emerges. The constructed solution is the *unique* solution of the Euler equations in a certain class of entropy-producing weak solutions with azimuthal symmetry and with regularity determined by the fact that it arises from a generic preshock.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 76L05; Secondary 35L67, 35Q31

KEYWORDS

Euler equations, shocks, weak singularities

INTERNATIONAL CONGRESS OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3636–3659
DOI 10.4171/ICM2022/87

1. INTRODUCTION

The compressible Euler equations are the fundamental mathematical model of fluid dynamics. Their mathematical analysis has a very rich history, see, for instance, the classical books of Courant and Friedrichs [10], or Landau and Lifshitz [17]. The unknowns of the model are the velocity $u : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, the mass density $\rho : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_+$, the total energy $E : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_+$, where $d \geq 1$ is the spatial dimension. The quasilinear system of conservation laws describing their evolution is given by

$$\partial_t(\rho u) + \operatorname{div}(\rho u \otimes u + pI) = 0, \tag{1.1a}$$

$$\partial_t \rho + \operatorname{div}(\rho u) = 0, \tag{1.1b}$$

$$\partial_t E + \operatorname{div}\big((p + E)u\big) = 0, \tag{1.1c}$$

representing the conservation of momentum, mass, and energy. Here $p : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}_+$ is the pressure which may be computed in terms of (u, ρ, E) as

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho|u|^2\right), \tag{1.1d}$$

where $\gamma > 1$ denotes the adiabatic exponent. The pressure may alternatively be computed in terms of the (specific) entropy $S : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ via

$$p(\rho, S) = \frac{1}{\gamma}\rho^\gamma e^S. \tag{1.2}$$

Note that in regions of spacetime where the fields (u, ρ, E) are smooth, one may replace (1.1c) by the transport of specific entropy

$$\partial_t S + u \cdot \nabla S = 0. \tag{1.3}$$

The system (1.1) is supplemented with *smooth* Cauchy data (u_0, ρ_0, E_0).

At least since the middle of the 19th century and the work of Riemann [21], it is known that the compressible Euler equations exhibit solutions which have smooth initial data and develop a *finite-time singularity*. The nonlinear interactions in (1.1) cause a gradual steepening of the density and velocity profiles, eventually leading to a first spacetime point at which their slope becomes infinite (the *preshock*). A shock wave then forms and propagates through the fluid according to the so-called Rankine–Hugoniot jump conditions, which ensure that the evolution gives an entropy-producing weak solution of (1.1).

A rigorous mathematical understanding of the above described process of *shock formation* and *shock development*, from smooth initial data, is partially available only in one space dimension [10,11,17], or equivalently, in the presence of radial symmetry for $d \geq 2$. We emphasize, however, that even for $d = 1$ a complete understanding of these phenomena was not available as of 2019. Indeed, regarding the 1D shock formation process, a rigorous proof of the expectation (see Eggers and Fontelos [13]) that the first singularity is asymptotically self-similar, and a stability analysis of the associated self-similar profiles within the Euler evolution (1.1), was unavailable. This issue was settled in our work [3]. Regarding the shock development process, Landau and Lifshitz note in [17, CHAPTER IX, §96] that simultaneously to the development of the discontinuous shock wave, other surfaces of higher-order singularities are expected to form. Landau and Lifschitz termed these surfaces *weak discontinuities*, but stopped short of describing their nature: "*The irregularity may be of various kinds. For example, the first spatial derivatives of ρ, p, u, etc., may be discontinuous on a surface, or these derivatives may become infinite or higher derivatives may behave in the same manner*." In spite of the huge literature on compressible flows, we are not aware of any analysis of these weak discontinuities for the Euler system (1.1). Providing a resolution to the problem raised by Landau and Lifschitz is the purpose of our work [1].

We emphasize that the arguments in our works [3] and [1] are able to treat not just the case $d = 1$, or $d \geq 2$ with radial symmetry, but a more general situation: $d = 2$ for flows with *azimuthal symmetry* and nonzero vorticity. We view the analysis of solutions with azimuthal symmetry as a key step in our program of understanding shock formation and development for the full Euler system (1.1) in multiple space dimensions ($d \geq 2$), from smooth initial data, in the absence of any symmetry assumptions, which is considered to be *the* outstanding open problem in the field.

2. PRIOR RESULTS FOR EULER SHOCK FORMATION AND DEVELOPMENT

The mathematical literature on the compressible Euler equations is too vast to review here. The majority of results have been focused on either the one-dimensional problem, or on the theory of weak solutions, or on the Riemann problem. See, for instance, the book of Dafermos [11] for an extensive modern review. In spite of this, there are very few results devoted to the mathematical analysis of shock formation for smooth initial data, and even less so to the shock development problem.

For the one-dimensional p-system (which models 1D isentropic Euler), Lebaud [18] was the first to prove shock formation and development. Chen and Dong [5], and also Kong [16], revisited the proof of Lebaud and established the formation and development of shocks for the 1D p-system with slightly more general initial data. However, as explained in Remark 3.3 below, the use an isentropic system cannot produce weak solutions to the Euler equations, even for $d = 1$. The first work to address the formation and development problem for the nonisentropic Euler equations was Yin [22], who considered the 3×3 system under spherical symmetry (which makes the problem one-dimensional). Independently of Yin, shock development for the barotropic Euler equations under spherical symmetry was established by Christodoulou and Lisbach [8]. Since isentropic dynamics cannot yield weak solutions to the Euler equations (see Remark 3.3), the analysis in [8] has been termed the *restricted shock development*. Christodoulou [7] has established restricted shock development for *irrotational and isentropic* 3D Euler equations, outside of symmetry assumptions. We note, however, that besides the inability of the isentropic model to capture the correct shock jump conditions, outside of radial symmetry the usage of an irrotational model can also not be justified; regular shock solutions produce entropy and generically create vorticity (see Remark 4.1 below).

As noted above, Landau and Lifshitz conjectured in [17, CHAPTER IX, §96] that at the same time that the discontinuous shock wave develops, other surfaces of weak singularities are expected to simultaneously form. For the full Euler system (1.1), with or without symmetry, even in one space dimension, the analysis of these surfaces of weak singularities has been heretofore nonexistent. In [1] we have proven that for the Euler equations in azimuthal symmetry, two such surfaces emerge from the preshock and move with the slower sound-speed characteristic ($\mathscr{s}_1$), and respectively with the fluid velocity ($\mathscr{s}_2$). We shall refer to this $\mathscr{s}_2$ surface as a *weak contact* because it moves with the fluid velocity, and both the normal velocity and pressure are one degree smoother than the density and entropy across this surface. The shall also refer to $\mathscr{s}_1$ as a *weak rarefaction* because the normal velocity to this curve is decreasing in the direction of its motion.

The precise analysis of the shock development problem in [1] is made possible by a very detailed understanding of the preshock which arises from smooth generic initial data. In multiple space dimensions and in the absence of symmetries, such a comprehensive description of the first singularity is currently unavailable. The constructive proofs of shock formation by Christodoulou [6], Christodoulou and Miao [9], and by Luk and Speck [19,20] yield the existence of at least one point in spacetime where a shock must form, and a bound is given for this blow up time; however, since the construction of the shock solution is a perturbation of a simple plane wave, there are numerous possibilities for the type of singularities that actually form; the blowup could potentially occur at one point, at multiple points, on a curve, or along a surface. The first step towards the precise characterization of the preshock in three space dimensions, without symmetries and for the full Euler equations, has been obtained recently by the first and last two authors [2,4]. We prove in [2,4] that the first singularity which arises from smooth and nondegenerate initial data develops at a single point in spacetime, it forms in an asymptotically self-similar way, and the corresponding similarity profiles are stable. This first singularity has been termed a *point-shock*, and it is given by the intersection of the preshock surface with the time slice $\{t = T_1\}$, where T_1 is the first time a gradient blowup occurs.

3. CLASSICAL VS REGULAR SHOCK SOLUTIONS

Given a sufficiently smooth initial datum (u_0, ρ_0, E_0) defined on $\mathbb{R}^d \times \{T_0\}$, the existence of a unique local in time smooth solution to the Euler system (1.1) defined on $\mathbb{R}^d \times [T_0, T_0 + \delta)$ for some $\delta > 0$ is classical. For a proof, see, for instance, the H^s energy estimates of Kato [14]. This solution may be continued uniquely on a maximal time interval $[T_0, T_1)$, characterized by the fact that T_1 is the first time at which the solution has an infinite gradient. Thus, there is no ambiguity in the notion of solution to (1.1) on $\mathbb{R}^d \times [T_0, T_1)$ since all the fields are differentiable in space and time, and so the solution is *classical*. The evolution on the time interval $[T_0, T_1)$ is called *shock formation*, leading to a first singularity at time T_1, the so-called *preshock*,[1] which we shall prove is generically of cusp-type, with the solution retaining Hölder $1/3$ regularity.

The evolution (1.1) may be continued past the time of the first singularity, say on an interval $(T_1, T_2]$, in what is known as *shock development*. The preshock instantaneously evolves into a discontinuous entropy producing shock wave, and we shall prove that in addition two other families of weak characteristic singularities simultaneously emerge from the preshock. In order to discuss shock development, we first need a suitable notion of solution to (1.1) on $\mathbb{R}^d \times (T_1, T_2]$, which in turn requires the introduction of the *Rankine–Hugoniot jump conditions* and of the *entropy condition*.

The Rankine–Hugoniot jump conditions are a manifestation of the fact (u, ρ, E) is a weak solution of (1.1), and thus the shock speed is related to the jumps of various quantities across the shock surface. More precisely, suppose that the shock front $\mathcal{S} \subset \mathbb{R}^d \times (T_1, T_2]$ is an orientable spacetime hypersurface across which the velocity, density, and energy jump. For $t \in (T_1, T_2]$, the shock front at time t locally separates space into two sets $\Omega^\pm(t)$, and we denote the values of the fields in these sets by $(u^\pm, \rho^\pm, E^\pm)$. We consider the case where this surface is parametrized as $\mathcal{S} := \{\mathfrak{s}(x,t) = 0\}$, and denote the spacetime normal to this surface as $-(\nabla_x \mathfrak{s}, \partial_t \mathfrak{s})|_{\mathcal{S}} =: (\mathfrak{n}, -\dot{\mathfrak{s}})$. We let $\mathfrak{n}(\cdot, t)$ point from $\Omega^-(t)$ to $\Omega^+(t)$, which is the direction of propagation of the shock front. We denote by $\dot{\mathfrak{s}}$ the shock speed, while the *jump* of a quantity f across the shock is written as $[\![f]\!] = f^- - f^+$, where $f^\pm$ are the traces of f along $\mathcal{S}$ in the regions $\Omega^\pm$. Let $u_{\mathfrak{n}} = u \cdot \mathfrak{n}|\mathfrak{n}|^{-1}$ be the projection of the velocity field in the direction of the normal vector $\mathfrak{n}$. The tangential components of the velocity are continuous across the shock, i.e., $[\![u - u_{\mathfrak{n}}\mathfrak{n}|\mathfrak{n}|^{-1}]\!] = 0$. The Rankine–Hugoniot jump conditions state that

$$\dot{\mathfrak{s}}|\mathfrak{n}|^{-1}[\![\rho u_{\mathfrak{n}}]\!] = \left[\!\!\left[\rho u_{\mathfrak{n}}^2 + pI\right]\!\!\right], \tag{3.1a}$$

$$\dot{\mathfrak{s}}|\mathfrak{n}|^{-1}[\![\rho]\!] = [\![\rho u_{\mathfrak{n}}]\!], \tag{3.1b}$$

$$\dot{\mathfrak{s}}|\mathfrak{n}|^{-1}[\![E]\!] = \left[\!\!\left[(p + E)u_{\mathfrak{n}}\right]\!\!\right]. \tag{3.1c}$$

Note that only one of the equations in (3.1) are used to compute the shock speed, while the remaining equations yield two constraints for the variables $(u_{\mathfrak{n}}^+, \rho^+, E^+)|_{\mathcal{S}}$ and $(u_{\mathfrak{n}}^-, \rho^-, E^-)|_{\mathcal{S}}$.

The entropy condition is nothing but the second law of thermodynamics, and states that the entropy ρS, which in view of (1.3) satisfies the conservation law $\partial_t(\rho S) + \nabla \cdot (\rho u S) = 0$ as long as the solution is smooth, must *increase* in the presence of a shock singularity. With the above choice of orientation of the normal vector $\mathfrak{n}$, the mass flux $j = \rho(u_{\mathfrak{n}} - \dot{\mathfrak{s}}|\mathfrak{n}|^{-1})$ is negative, mass is passing across the shock from $\Omega^+(t)$ into $\Omega^-(t)$, and so the physical entropy condition becomes

$$[\![S]\!] > 0. \tag{3.2}$$

[1] To be precise, this first singularity is called a *preshock* only for one-dimensional problems, or in the presence of azimuthal symmetry, discussed here. For $d \geq 2$, in the absence of any symmetry, this first singularity occurs at a single point in spacetime, the *point-shock*. The point-shock is the intersection of the preshock with the time slice $\{t = T_1\}$.

Remark 3.1 (The physical entropy condition and the geometric Lax entropy conditions). The negativity of the mass flux $j = \rho(u_{\mathfrak{n}} - \dot{\mathfrak{s}}|\mathfrak{n}|^{-1})$ immediately gives

$$u^- \cdot \mathfrak{n} < \dot{\mathfrak{s}}, \quad u^+ \cdot \mathfrak{n} < \dot{\mathfrak{s}}. \tag{3.3}$$

The *Lax geometric entropy conditions* are given by (3.3) along with

$$u^+ \cdot \mathfrak{n} + c^+ < \dot{\mathfrak{s}} < u^- \cdot \mathfrak{n} + c^-, \tag{3.4}$$

where c^- and c^+ are the sound speeds behind and in front of the shock. Condition (3.4) states that the shock discontinuity is supersonic relative to the state in front (the "+" phase) and subsonic relative to the state behind (the "−" phase) the shock. It turns out that for an ideal gas, and under the assumption that (u, ρ, E) has a *weak shock*, i.e.,

$$\sup_{t\in[T_1,T_2]} \big|[\![u(t)]\!]\big| + \big|[\![\rho(t)]\!]\big| + \big|[\![E(t)]\!]\big| \ll 1,$$

the physical entropy condition (3.2) is *equivalent* to the Lax geometric entropy conditions. Moreover, in this setting one may show that the Rankine–Hugoniot jump conditions imply

$$[\![S]\!] = \mathcal{O}\big([\![p]\!]^3\big), \tag{3.5}$$

with a positive prefactor; it follows that the entropy production postulated in (3.2) implies the positivity of the jumps $[\![p]\!] > 0$, $[\![\rho]\!] > 0$, and$[\![u_{\mathfrak{n}}]\!] > 0$. See Landau and Lifshitz [17, CHAPTER IX] or [1, SECTION 2] for details.

Having defined the Rankine–Hugoniot conditions (3.1) and the entropy condition (3.2), we are now ready to define the physically relevant notion of solution to the development problem for (1.1), evolving from the preshock data.

Definition 3.2 (Regular shock solution). We say that (u, ρ, E) and a shock front $\mathcal{S}$ is a *regular shock solution* on $\mathbb{R}^d \times [T_1, T_2]$ if the following conditions hold:

- (u, ρ, E) is a weak solution of (1.1) and $\rho \geq \rho_{\min} > 0$;
- the shock front $\mathcal{S} \subset \mathbb{R}^d \times [T_1, T_2]$ is an orientable codimension 1 hypersurface;
- (u, ρ, E) are Lipschitz continuous in space and time on the complement of the shock surface $(\mathbb{R}^d \times [T_1, T_2]) \setminus \mathcal{S}$;
- (u, ρ, E) have discontinuities across the shock which satisfy the Rankine–Hugoniot jump conditions (3.1);
- entropy is produced at the shock, so that (3.2) holds.

Remark 3.3 (Regular shock solutions cannot be isentropic). Definition 3.2 shows that one *cannot* study the physical shock development problem within the isentropic Euler model ($S \equiv 0$). Indeed, while the isentropic Euler system is perfectly justifiable prior to the first singularity since $S|_{t=T_0} = 0$ implies by (1.3) that $S(\cdot, t) = 0$ for all $t \in [T_0, T_1]$, as soon as a shock front develops entropy must be generated according to (3.5). That is, the flow becomes nonisentropic in order to satisfy the Rankine–Hugoniot jump conditions, or equivalently, in order for (u, ρ, E) to be a weak solution of the Euler system (1.1). Consistency with the production of entropy (3.2) is a secondary condition, which is meant to rule out the physically incorrect weak solutions.

4. AZIMUTHAL SYMMETRY

In the regions of spacetime where the fields (u, ρ, E) are differentiable, the divergence form of the Euler equations (1.1) is equivalent to a more symmetric version, in which the conservation of the energy is replaced by the transport of specific entropy S, and the conservation of mass is replaced by the evolution of the rescaled sound speed σ, defined as

$$\sigma = \frac{1}{\alpha}\sqrt{\partial p/\partial \rho} = \frac{1}{\alpha} e^{\frac{S}{2}} \rho^{\alpha}, \quad \text{where } \alpha = \frac{\gamma - 1}{2}. \tag{4.1}$$

With this notation, the ideal gas equation of state (1.2) becomes $p = \frac{\alpha^2}{\gamma}\rho\sigma^2$, while the Euler equations (1.1), as a system for (u, σ, S), are given by

$$\partial_t u + (u \cdot \nabla)u + \alpha\sigma\nabla\sigma = \frac{\alpha}{2\gamma}\sigma^2\nabla S, \tag{4.2a}$$

$$\partial_t \sigma + (u \cdot \nabla)\sigma + \alpha\sigma \operatorname{div} u = 0, \tag{4.2b}$$

$$\partial_t S + (u \cdot \nabla)S = 0. \tag{4.2c}$$

Note that the system (4.2) is valid away from the shock surface, and that the Rankine–Hugoniot conditions need to be determined from the conservation law form of the Euler equations (1.1). Additionally, we note that the Rankine–Hugoniot jump conditions, defined in terms of the jumps of normal velocity, density, and energy (3.1), may be translated into jump conditions for the variables (u, σ, E), by appealing to (4.1) and $E = \frac{1}{2}\rho|u|^2 + \frac{\alpha}{2\gamma}\rho\sigma^2$.

A fundamental quantity to the analysis of (4.2) is the vorticity, defined as $\omega = \nabla^{\perp} \cdot u$ for $d = 2$ and $\omega = \nabla \times u$ for $d = 3$. Then, the *specific vorticity* $\zeta = \frac{\omega}{\rho}$ solves

$$\partial_t \zeta + (u \cdot \nabla)\zeta = \begin{cases} \frac{\alpha}{\gamma}\frac{\sigma}{\rho}\nabla^{\perp}\sigma \cdot \nabla S, & d = 2, \\ (\zeta \cdot \nabla u) + \frac{\alpha}{\gamma}\frac{\sigma}{\rho}\nabla\sigma \times \nabla S, & d = 3, \end{cases} \tag{4.3}$$

and the analysis of (4.3) is of fundamental importance to our works [1–4].

Remark 4.1 (Regular shock solutions generically create vorticity). The baroclinic torque term on the right side of (4.3) shows that a misalignment of density and entropy gradients creates vorticity. Combining this observation with Remark 3.3, it is thus expected that even when one starts the shock formation process with isentropic irrotational flow, as soon as the shock surface is formed, *generically* not just entropy is created, but vorticity is created as well. Thus, for generic smooth initial data, the shock development problem cannot be studied in the class of irrotational flows. The only two exceptions we are aware of are $d = 1$ or the conceptually equivalent situation $d \geq 2$ under the reduction of *radial symmetry*, when there is no vorticity to speak of in the first place.

The above remark motivates our introduction of the class of solutions to the Euler equations with *azimuthal symmetry*. This class of solutions may be defined for $d = 2$ by the requirement that the velocity and sound speed are linear functions of r with nonlinear dependence of (θ, t), while the entropy is only a function of (θ, t). Here (r, θ) are the polar coordinates on $\mathbb{R}^2$. This class of solutions is formally maintained under the Euler evolution (1.1). These solutions have nonzero vorticity, both velocity components are nontrivial and strongly affect the shock formation and development, and the system has three distinct wave-speed families. As such, we view azimuthal symmetry as a multidimensional intermediary case between one-dimensional problems, and multidimensional problems without any symmetry. More precisely, by introducing the unknowns (a, b, c, k) via

$$(u_r, u_\theta, \sigma, S)(r, \theta, t) =: \big(ra(\theta, t), rb(\theta, t), rc(\theta, t), k(\theta, t)\big), \tag{4.4}$$

and canceling all powers of r, the Euler system (4.2) becomes

$$(\partial_t + b\partial_\theta)a + a^2 - b^2 + \alpha c^2 = 0, \tag{4.5a}$$

$$(\partial_t + b\partial_\theta)b + \alpha c\partial_\theta c + 2ab = \frac{\alpha}{2\gamma} c^2 \partial_\theta k, \tag{4.5b}$$

$$(\partial_t + b\partial_\theta)c + \alpha c\partial_\theta b + \gamma ac = 0, \tag{4.5c}$$

$$(\partial_t + b\partial_\theta)k = 0. \tag{4.5d}$$

For smooth initial data (u_0, ρ_0, E_0) or (u_0, σ_0, S_0) at $t = T_0$ which has azimuthal symmetry, one may define via (4.4) suitable initial data (a_0, b_0, c_0, k_0) for the system (4.5). Then, solving (4.5) gives a unique solution (a, b, c, k) on a maximal time interval $[T_0, T_1)$ on which the solution remains smooth. On this time interval, the unique solution (u, σ, S) to (4.2) is then given by the identification (4.4). That is, as long as solutions remain smooth, the azimuthal symmetry of the data is preserved, and systems (1.1), (4.4), and (4.5) are all equivalent. As we shall see below, we may in fact continue the solution (a, b, c, k) of (4.5) past $t = T_1$ in a *unique* way as a physical shock solution by translating the Rankine–Hugoniot jump conditions (3.1) and the entropy condition (3.2) into corresponding azimuthal jump/entropy conditions. The resulting solution (u, σ, S) (or equivalently (u, ρ, E)) obtained via the identification (4.4) can be shown to be a regular weak solution of the full Euler system (4.4) (equivalently (1.1)) in the sense of Definition 3.2. The uniqueness of this regular weak solution to (1.1) is only known to hold if we assume that the solution has azimuthal symmetry.

4.1. Riemann-like variables in azimuthal symmetry

For simplicity of presentation, for the remainder of this review, as was done in [1], we shall work with the adiabatic exponent

$$\gamma = 2, \quad \text{or equivalently} \quad \alpha = \frac{1}{2}. \tag{4.6}$$

We also note that it is convenient to rescale time, letting

$$t = \frac{3}{4}\tilde{t}, \quad \text{so that} \quad \partial_t \mapsto \frac{4}{3}\partial_{\tilde{t}}, \tag{4.7}$$

and for notational simplicity, we continue to write t for $\tilde{t}$. More importantly, it is convenient for the subsequent analysis to work with Riemann-like variables w and z which symmetrize (in a certain sense) the b and c evolutions (4.5). These Riemann variables are defined by

$$w = b + c, \quad z = b - c, \tag{4.8}$$

so that $b = \frac{1}{2}(w+z)$ and $c = \frac{1}{2}(w-z)$. We shall refer to w as the *dominant Riemann variable*, and to z as the *subdominant Riemann variable*.

With the adiabatic exponent from (4.6), the temporal rescaling (4.7), and using the Riemann variables from (4.8), the system (4.5) can be equivalently written as

$$\partial_t w + \lambda_3 \partial_\theta w = -\frac{8}{3} a w + \frac{1}{24}(w-z)^2 \partial_\theta k, \tag{4.9a}$$

$$\partial_t z + \lambda_1 \partial_\theta z = -\frac{8}{3} a z + \frac{1}{24}(w-z)^2 \partial_\theta k, \tag{4.9b}$$

$$\partial_t k + \lambda_2 \partial_\theta k = 0, \tag{4.9c}$$

$$\partial_t a + \lambda_2 \partial_\theta a = -\frac{4}{3} a^2 + \frac{1}{3}(w+z)^2 - \frac{1}{6}(w-z)^2. \tag{4.9d}$$

where the three distinct wave speeds are given by

$$\lambda_1 = \frac{1}{3} w + z, \quad \lambda_2 = \frac{2}{3} w + \frac{2}{3} z, \quad \lambda_3 = w + \frac{1}{3} z. \tag{4.10}$$

The Cauchy problem for (4.9) is considered with initial conditions given by $(w_0, z_0, a_0, k_0)(\theta) = (w, z, a, k)(\theta, T_0)$. We shall henceforth refer to (4.9)–(4.10) as the *azimuthal Euler system*.

Remark 4.2 (Specific vorticity in azimuthal symmetry). Using the azimuthal symmetry ansatz (4.4), the specific vorticity ζ may be written as

$$\zeta(r, \theta, t) = \varpi(\theta, t) = \big(4(w + z - \partial_\theta a) c^{-2} e^k\big)(\theta, t), \tag{4.11}$$

and we may show that it solves the evolution equation

$$\partial_t \varpi + \lambda_2 \partial_\theta \varpi = \frac{8}{3} a \varpi + \frac{4}{3} e^k \partial_\theta k. \tag{4.12}$$

Remark 4.3 (Motivation for the choice of γ in (4.6)). The choice of adiabatic exponent $\gamma = 2$ was made in order to emphasize that the shock wave produces not just entropy, but it also generates the subdominant Riemann variable z. In order to clearly emphasize this, for the shock formation process we choose initial data at time $t = T_0$ which satisfies

$$k(\theta, T_0) = 0, \quad \text{and} \quad z(\theta, T_0) = 0. \tag{4.13}$$

The entropy transport (4.9c) ensures that for any $t \in [T_0, T_1]$, where T_1 is the time of the first singularity, we have $k(\cdot, t) = 0$. The Rankine–Hugoniot conditions (cf. (4.16) below) guarantee that entropy *must be produced* at the shock, resulting in $k(\theta, t) > 0$ in a certain region of points $(\theta, t) \in \mathbb{T} \times (T_1, T_2]$. The choice of $k_0 = 0$ in (4.13) emphasizes the production of entropy in the clearest possible way. The choice $\gamma = 2$ ($\alpha = \frac{1}{2}$) is related to the evolution of the subdominant Riemann variable z. Since we have that $k \equiv 0$, the right-hand side of (4.9b) simplifies to $-\frac{8}{3} a z$, but we note that for general values of γ, this term would simplify to $-\frac{3+2\alpha}{1+\alpha} a z - \frac{1-2\alpha}{1+\alpha} a w$. As such, even if $z_0 = 0$, the term $-\frac{1-2\alpha}{1+\alpha} a w$ would ensure that $z \not\equiv 0$ for $t > T_0$. For $\alpha = \frac{1}{2}$, this term, however, does not exist, and so the choice of $k_0 = 0$ in (4.13) ensures that $z(\cdot, t) = 0$ for all $t \in [T_0, T_1]$. The remarkable fact is that the Rankine–Hugoniot conditions (cf. (4.16) below) imply that we must have $z < 0$ for a certain region of points $(\theta, t) \in \mathbb{T} \times (T_1, T_2]$. Thus, the choice $z_0 = 0$ is made in order to most clearly emphasize the breaking of the symmetry $b = c$ at the shock.

As noted in Remark 4.3, the choice of initial datum in (4.13) implies that during the shock formation process, we have that $k \equiv 0$ and $z \equiv 0$, so that the system (4.9) becomes

$$\partial_t w + w\partial_\theta w = -\frac{8}{3}aw, \tag{4.14a}$$

$$\partial_t a + \frac{2}{3}w\partial_\theta a = -\frac{4}{3}a^2 + \frac{1}{6}w^2. \tag{4.14b}$$

The preshock, which will be shown to be smooth away from a unique blowup point $\theta_* \in \mathbb{T}$, inherits the property that $k(\theta, T_1)$ and $z(\theta, T_1)$ vanish on $\mathbb{T}$, but these symmetries are broken instantaneously during the shock development process. The presence of a shock necessitates that we supplement the system (4.9) with Rankine–Hugoniot jump and entropy conditions.

4.2. Rankine–Hugoniot jump and entropy conditions

In azimuthal symmetry, with the adiabatic exponent from (4.6) and the temporal rescaling (4.7), the shock hypersurface is given as

$$\mathcal{S} = \big\{(r, \theta, t) : \mathfrak{s}(t) - \theta = 0\big\}.$$

The spatial normal to this hypersurface is $\mathfrak{n} = \frac{1}{r}\vec{e}_\theta$. We have that $\dot{\mathfrak{s}} > 0$ and so the shock is moving from left to right when the angular variable θ is viewed as being defined on $[-\pi, \pi)$. To see this, note that since $z = 0$ by (4.8) we have that $w = 2c$, and since we wish to stay away from vacuum, we must have $c \geq c_{\min} > 0$ on $\mathbb{T}$; therefore, w is strictly positive on $\mathbb{T}$, which implies that the three wave speeds defined in (4.10) are all strictly positive, and ordered as $\lambda_1 < \lambda_2 < \lambda_3$ on $\mathbb{T} \times [T_0, T_1]$ (by continuity this also holds on $\mathbb{T} \times (T_1, T_2]$ if $T_2 - T_1 \ll 1$). The negativity of the mass flux in (3.3) then yields $\dot{\mathfrak{s}} > 0$. According to the orientation of $\mathfrak{n}$, we denote by $(w_+, z_+, a_+, k_+)(t)$ the limiting values on the shock curve $\mathfrak{s}(t)$ from the right (or front) of the shock, and by $(w_-, z_-, a_-, k_-)(t)$ the limiting values from the left (or back) of the shock. As discussed in [1, REMARK 2.5], the Lax geometric entropy inequalities (3.3)–(3.4) imply that the characteristics of the three wave speeds $\{\lambda_i\}_{i=1}^3$ in front of the shock (the "+" phase) impinge on the shock front, carrying with them the data from the $\{t = T_1\}$ Cauchy hypersurface. In particular, since $k(\cdot, T_1) = z(\cdot, T_1) = 0$, this implies that during the development process we have

$$k_+(t) = z_+(t) = 0, \quad \text{for all } t \in (T_1, T_2], \tag{4.15}$$

so that $[\![k]\!] = k_-$ and $[\![z]\!] = z_-$. Using (4.15) and the observation that $u_{\mathfrak{n}}^{\pm} = rb_\pm(\mathfrak{s}(t), t)$ the Rankine–Hugoniot jump conditions (3.1) may be shown to be equivalent to a system of two equations which are used to determine the values of z_- and k_- in terms of w_+ and w_-

$$
\begin{aligned}
&(e^{k_-}-1)(w_- - z_-)^4\big(3w_+^2e^{k_-} - (w_- - z_-)^2\big)\\
&\quad = \big((w_- - z_-)^2 - e^{k_-}w_+^2\big)^3, \qquad (4.16a)\\
&\left((w_- - z_-)^2(w_- + z_-)^2 + \frac{1}{8}(w_- - z_-)^4 - \frac{9}{8}e^{k_-}w_+^4\right)\big((w_- - z_-)^2 - e^{k_-}w_+^2\big)\\
&\quad = \big((w_- - z_-)^2(w_- + z_-) - e^{k_-}w_+^3\big)^2, \qquad (4.16b)
\end{aligned}
$$

and an evolution equation for $\dot{\mathfrak{s}}$ given by

$$
\dot{\mathfrak{s}}(t) = \frac{2}{3}\frac{e^{-k_-}(w_- - z_-)^2(w_- + z_-) - w_+^3}{e^{-k_-}(w_- - z_-)^2 - w_+^2}. \qquad (4.16c)
$$

To summarize, the values of the dominant Riemann variable, w_+ in the front and w_- in the back of the shock, determine the values of z_- and k_- via (4.16a)–(4.16b), which in turn allows one to compute the location of the evolving shock front. We note that the dominant Riemann variable w travels according to the fastest wave-speed in the system (4.9), namely λ_3. Thus, the values of w_+ and w_- are carried from the $\{t = T_1\}$ Cauchy hypersurface via the characteristics of λ_3, which impinge on the shock front from the left and right.

Remark 4.4 (The entropy condition in azimuthal symmetry). The system of three equations (4.16) is in one-to-one correspondence with the Rankine–Hugoniot jump conditions (3.1). So the natural question is: What is the equivalent of the physical entropy condition (3.2) in azimuthal symmetry? To answer this question, we first note that (4.16a)–(4.16b) are a coupled system of sixth-order polynomials in the variables w_+, w_-, z_-, e^{k_-}. The second observation is that at the preshock we have $w_+(T_1) = w_-(T_1)$ and $z_-(T_1) = k_-(T_1) = 0$, which solves (4.16a)–(4.16b). The natural question then is whether in the weak shock regime $0 < [\![w]\!] = w_- - w_+ \ll 1$, with $\langle\!\langle w\rangle\!\rangle = \frac{1}{2}(w_- + w_+) > 0$, the system (4.16a)–(4.16b) has a *unique* solution or not. For the sixth-order equations with real coefficients, the presence of one real solution implies the presence of at least one more solution. Indeed, one may verify that in the weak shock regime the system (4.16a)–(4.16b) has exactly two real solutions with $|z_-| + |k_-| \ll 1$, the other roots being complex. The remarkable fact is that only one of these two solutions is entropy producing, $k_- > 0$. Thus, the role of the physical entropy condition (3.2), which is equivalent in view of (4.15) to $k_- > 0$, is to *select the unique physically relevant root* of the system of equations (4.16a)–(4.16b).

We conclude this section by revisiting the notion of a regular shock solution, as defined in Definition 3.2, in the context of the azimuthal Euler equations. During the *formation part* of our result, i.e., for $t \in [T_0, T_1)$, we have that the solution (w, z, k, a) of (4.9)–(4.10) is smooth, so that the notion of solution is the classical one: the system (4.9) is satisfied in the sense of C^1-functions of space and time. On the time interval $[T_1, T_2]$, which covers the *development part* of our result, the notion of *regular shock solution* becomes:

Definition 4.5 (Regular azimuthal shock solution). We say that (w, z, k, a) and a shock front parametrized as $\mathcal{S} = \{\mathfrak{s}(t) = \theta\}$ is a *regular azimuthal shock solution* on $\mathbb{T} \times [T_1, T_2]$ if

- (w, z, k, a) are $C^1_{\theta,t}$ smooth, and ϖ is $C^0_{\theta,t}$ smooth, on the complement of $\mathcal{S}$;
- on the complement of the shock curve, (w, z, k, a) solve the equations (4.9)–(4.10) pointwise, and ϖ solves (4.12) pointwise;
- (w, z, k) have jump discontinuities across the shock curve which satisfy the algebraic equations (4.16a)–(4.16b);
- the shock location $\mathfrak{s} : [T_1, T_2] \to \mathbb{T}$ is C^1_t smooth and solves (4.16c);
- entropy is produced at the shock so that $[\![k]\!](t) > 0$ for $t \in (T_1, T_2]$.

5. MAIN RESULTS

The main result of [1] is stated first in terms of the azimuthal variables (w, z, k, a). The result may be best visualized by inspecting Figures 1, 2, 3, 4. A condensed statement is as follows; for details, see [1, THEOREMS 3.2, 5.5, 6.1].

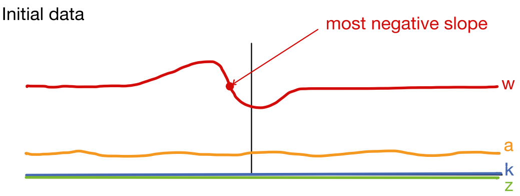


FIGURE 1
The initial conditions $(w, z, k, a)|_{t=T_0}$ satisfying (4.13) are represented in (red, green, blue, orange) as functions of the angular variable $\theta \in [-\pi, \pi)$. The function $w(\cdot, T_0)$ is strictly positive and has has a nondegenerate most negative slope of size $\approx -\frac{1}{\varepsilon}$ at a unique point in $\mathbb{T}$. The function $a(\cdot, T_0)$ is $\mathcal{O}(1)$ in $C^4(\mathbb{T})$.

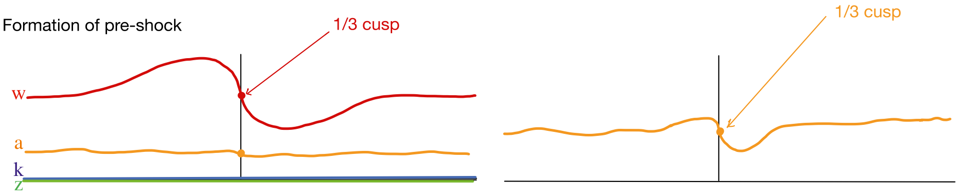


FIGURE 2
At the time of the first singularity, the functions $(w, z, k, a)|_{t=T_1}$ are sketched in the figure on the left, using the same color scheme as in Figure 1. In the image on the right, we have plotted the function $\partial_\theta a$, which also develops a singularity at $t = T_1$. More precisely, the shock formation process for the system (4.14) results in the formation of the preshock at time T_1, manifested as a $C^{\frac{1}{3}}$ cusp at a unique distinguished angle $\theta_* \in \mathbb{T}$ for the functions w and $\partial_\theta a$. At T_1 we have that z and k remain equal to 0.

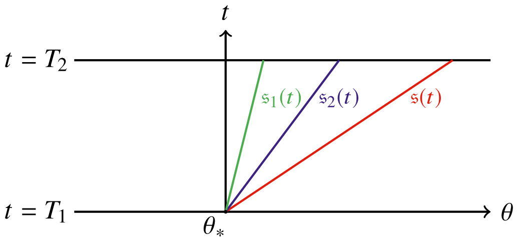


FIGURE 3

Three distinct families of singularities instantaneously emerge from the the preshock located at (θ_*, T_1). Across the classical shock curve $\mathfrak{s}$ the fields $(w, z, k, \partial_\theta a)$ jump, and the Rankine–Hugoniot conditions are satisfied. A weak rarefaction singularity develops across the curve $\mathfrak{s}_2$ which travels along characteristics of λ_2. Here the quantities (w, z, k) have regularity $C^{1,1/2}$ and no better. A weak contact singularity forms across the curve $\mathfrak{s}_1$ which travels with the characteristics of λ_1. Here the function z has regularity $C^{1,1/2}$ and no better. The functions z and k are equal to 0 on the left-hand side of $\mathfrak{s}_1$ and on the right-hand side of $\mathfrak{s}$.

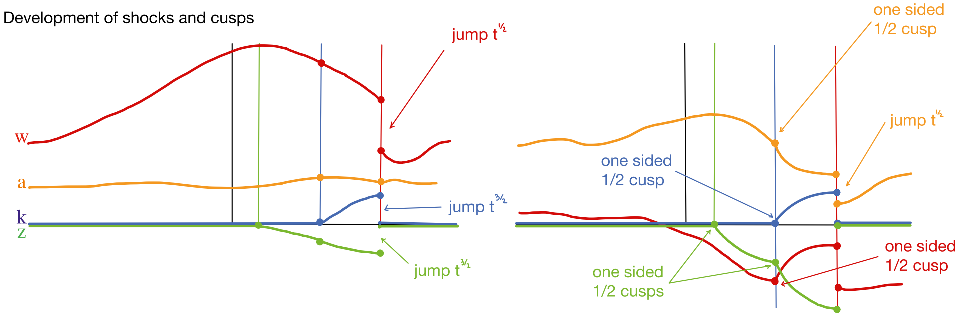


FIGURE 4

On the left-hand side, we have a schematic representation of the functions $(w, z, k, a)|_{t=T_2}$ using the color scheme from Figure 1. On the right-hand side, a schematic representation of the functions $(\partial_\theta w, \partial_\theta z, \partial_\theta k, \partial_\theta a)|_{t=T_2}$ is given. In both images, the vertical lines represent the location of $\mathfrak{s}_1(T_2) < \mathfrak{s}_2(T_2) < \mathfrak{s}(T_2)$ using the color scheme from Figure 3. The image on the left emphasizes that all quantities except for a jump across the shock, and that z and k remain equal to 0 on $\mathbb{T} \setminus [\mathfrak{s}_1(T_2), \mathfrak{s}(T_2)]$. The image on the right emphasizes that the one-sided cusps form at the weak contact and weak rarefaction, and that $\partial_\theta a$ jumps across the shock.

Theorem 5.1 (Main result in azimuthal symmetry). *From smooth isentropic initial data at time T_0 with vanishing subdominant Riemann variable, as described in the first paragraph of Section 6, there exist smooth solutions to the azimuthal Euler system (4.9) that form a preshock singularity, at a time $T_1 > T_0$. The first singularity occurs at a single point in space, θ_*, and this first singularity is shown to have an asymptotically self-similar shock profile exhibiting a $C^{1/3}$ cusp in the dominant Riemann variable and a $C^{1,1/3}$ cusp in the radial velocity. A series expansion for $w(\cdot, T_1)$ in terms of $(\theta - \theta_*)^{1/3}$ may be computed explicitly.*

After the preshock is formed, the solution to (4.9)–(4.10) *is continued* uniquely *for a short time* $(T_1, T_2]$ *as a regular azimuthal shock solution (cf. Definition* 4.5*) with the following properties:*

- *Across the shock curve* $\mathfrak{s}$*, for all* $t \in (T_1, T_2]$*, the state variables jump*

$$[\![w]\!] \sim (t - T_1)^{\frac{1}{2}}, \quad [\![\partial_\theta a]\!] \sim (t - T_1)^{\frac{1}{2}}, \quad [\![z]\!] \sim (t - T_1)^{\frac{3}{2}}, \quad [\![k]\!] \sim (t - T_1)^{\frac{3}{2}}.$$

- *Across the characteristic* $\mathfrak{s}_2$ *emanating from the preshock and moving with the fluid velocity, the Riemann variables and the entropy make* $C^{1,1/2}$ *cusps approaching from the right. Limiting from the left, these variables are* C^2 *smooth.*
- *Across the characteristic* $\mathfrak{s}_1$ *emanating from the preshock and moving with the sound speed minus the fluid velocity, the entropy is zero while the subdominant Riemann variable makes a* $C^{1,1/2}$ *cusp from the right. Limiting from the left, all fields are* C^2 *smooth.*

We note that the proof of Theorem 5.1, which is the bulk of our paper [1], applies with minor modifications to the case of the Euler equations for $d = 1$, or in the case of radial symmetry $d \geq 2$. In fact, as mentioned already in Remark 4.1, these two cases are simpler than the azimuthal symmetry considered here, since the vorticity vanishes identically.

Via the identification (4.4), Theorem 5.1 implies the following result for the Euler system in terms of hydrodynamic variables. We only state a condensed result here, and refer the interested reader to [1, THEOREMS 1.2, 7.1, 7.2] for details. The pictorial representation of this result is given in Figure 5 below.

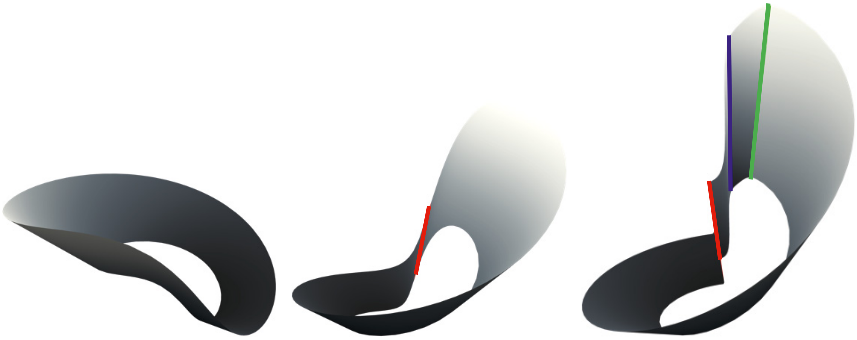


FIGURE 5

Values of the density written in polar coordinates $\rho(r, \theta, t)$, and plotted for $r \in [1, 2]$. The image on the left represents the smooth data at time T_0. The center image shows the preshock formed at time T_1, at one specific value of the angular coordinate; we marked the corresponding line in red. The image on the right represents the density at time T_2, where we have represented in red the line along which the shock discontinuity occurs, in blue the line containing the weak contact, and in green the line corresponding to the weak rarefaction.

Theorem 5.2 (Main result for 2D Euler). *For smooth isentropic initial data at time T_0 with azimuthal symmetry, there exist smooth solutions to the 2D Euler equations* (1.1) *that form a preshock singularity at a time $T_1 > T_0$. The first singularity occurs along a half-infinite ray and the blowup is asymptotically self-similar, exhibiting a $C^{1/3}$ cusp in the angular velocity and mass density, and a $C^{1,1/3}$ cusp in the radial velocity. Moreover, the blowup is given by a series expansion whose coefficients are computed as a function of the initial data.*

Past the preshock, the solution is continued on $(T_1, T_2]$, as an entropy-producing regular shock solution (cf. Definition 3.2*) of the full 2D Euler equations* (1.1)*. The solution is unique in the class of entropy producing weak solutions with azimuthal symmetry, with a certain weak shock structure and suitable regularity off the shock (see the space $\mathcal{X}_{\overline{\varepsilon}}$ defined in* (7.8) *below). The following properties are established for $t \in (T_1, T_2]$:*

- *Across the classical shock hypersurface, all the state variables jump:*

$$[\![u_\theta]\!] \sim (t - T_1)^{\frac{1}{2}}, \quad [\![\rho]\!] \sim (t - T_1)^{\frac{1}{2}},$$
$$[\![\partial_\theta u_r]\!] \sim (t - T_1)^{\frac{1}{2}}, \quad [\![S]\!] \sim (t - T_1)^{\frac{3}{2}}.$$

- *Across the characteristic emanating from the preshock and moving with the fluid velocity, the entropy, density, and radial velocity all have a $C^{1,1/2}$ one-sided cusp from the right, while from the left, they are all C^2 smooth. The second derivatives of the angular velocity and pressure are bounded across this curve, justifying the name weak rarefaction.*

- *Across the characteristic emanating from the preshock and moving with sound speed minus the fluid velocity, the entropy is zero while the angular velocity and density have $C^{1,1/2}$ one-sided cusps from the right, while from the left, they are C^2 smooth. The second derivative of the radial velocity is bounded across this curve, justifying the name weak contact singularity.*

Theorem 5.2 yields a full propagation of singularities result for regular shock solutions of the Euler equations, capturing both the jump discontinuity and the weak singularities emanating from the initial cusp in the preshock. This gives an answer to the problem raised by Landau and Lifschitz in [17, CHAPTER IX, §96], at least in the context of flows with azimuthal symmetry (or one-dimensional flows).

Remark 5.3 (Anomalous entropy production). Theorem 5.2 provides an example of an entropy producing weak solution $(u, \rho, E) \in L^\infty_t(BV \cap L^\infty)_{\mathrm{loc}} \subset L^\infty_t(B^{1/p}_{p,\infty})_{\mathrm{loc}}$, for all $p \geq 1$. This regularity class encodes the emergence of a regular shock, obtained by continuing the past the first singularity. This proves that the Onsager-criterion proven by the second author and Eyink in [12, THEOREM 3], which states that if $(u, \rho, E) \in L^\infty_t(B^{1/3+}_{3,\infty} \cap L^\infty)_{\mathrm{loc}}$ then there is no entropy production, is in fact sharp.

Remark 5.4 (Uniqueness and entropy). Theorem 5.2 establishes the uniqueness of solutions in a class of weak solutions with azimuthal symmetry, with *weak shock structure*, and which have regularity consistent with the fact that they emanate from a $C^{1/3}$ preshock (cf. (7.8) below), which in turn is the generic regularity that should be expected to arise at the first singularity from a smooth initial datum. The role of the entropy condition in establishing this uniqueness was explained in Remark 4.4. We contrast our uniqueness statement to the ill-posedness of the Euler system within the class of bounded, entropy-producing weak solutions emanating from 1D Riemann data, cf. Klingenberg et al. [15] and references therein.

6. OUTLINE: THE FORMATION OF THE PRESHOCK

Fix a constant $\kappa_0 > 1$ sufficiently large and let $\varepsilon > 0$ be sufficiently small. Consider the azimuthal Euler system (4.9)–(4.10) with initial data given at time $T_0 = -\varepsilon$, satisfying (4.13), and with $w(\cdot, T_0)$ and $a(\cdot, T_0)$ which lie in a certain open subset of $C^4(\mathbb{T})$ described roughly as follows. The initial data for the radial velocity is taken to satisfy $\|a(\cdot,-\varepsilon)\|_{L^\infty} \le \varepsilon$, $\|\partial_\theta a(\cdot,-\varepsilon)\|_{L^\infty} \lesssim \frac{1}{20}\kappa_0$, and $\|\partial_\theta^n a(\cdot,-\varepsilon)\|_{L^\infty} \lesssim 1$ for $2 \le n \le 4$. The initial data for the dominant Riemann variable is described in detail in [1, **EQUATIONS (4.17)–(4.25)**]. The most important property is that $w(\cdot,-\varepsilon) \in C^4(\mathbb{T})$ has a nondegenerate global minimum at a single point of $\mathbb{T}$, labeled for convenience by 0, where it holds that

$$w(0,-\varepsilon) = \kappa_0, \quad \partial_\theta w(0,-\varepsilon) = -\varepsilon^{-1}, \quad \partial_\theta^2 w(0,-\varepsilon) = 0, \quad \partial_\theta^3 w(0,-\varepsilon) = 6\varepsilon^{-4}. \tag{6.1}$$

Other conditions are that $\frac{7}{8}\kappa_0 \le w(\cdot,-\varepsilon) \le \frac{9}{8}\kappa_0$ which ensures that the density is bounded away from vacuum, that $w(\cdot,-\varepsilon) - \kappa_0$ is compactly supported $B_{\varepsilon^{1/2}}(0)$, and that the function $W(y) := \varepsilon^{-1/2}(w(y\varepsilon^{3/2},-\varepsilon) - \kappa_0)$ lies in a certain ε-dependent open ball in the C^4 topology centered at the stable global self-similar solution of the 1D Burgers equation, $\overline{W}$, which is defined implicitly as the analytic solution of $\overline{W}(y) + \overline{W}(y)^3 + y = 0$.

For such datum, the formation of the first gradient singularity for (4.9)–(4.10) was previously established in [3]. This singularity is characterized as a *stable asymptotically self-similar* $C_\theta^{1/3}$ cusp for the dominant Riemann variable w, the so-called *preshock*, which occurs at a precisely computable spacetime location (θ_*, T_1), with $\theta_* \approx \kappa_0\varepsilon$ and $T_1 = \mathcal{O}(\varepsilon^3)$. The subdominant Riemann variable z and entropy κ remain identically equal to 0 on $\mathbb{T} \times [-\varepsilon, T_1]$, while radial velocity and specific vorticity satisfy $a \in L^\infty(-\varepsilon, T_1; C^{1,1/3}(\mathbb{T}))$ and $\varpi \in L^\infty(-\varepsilon, T_1; C^{0,1}(\mathbb{T}))$. From here, one may show that asymptotically as $\theta \to \theta_*$:

$$w(\theta, T_1) = \kappa - \mathsf{b}(\theta-\theta_*)^{\frac{1}{3}} + o\big((\theta-\theta_*)^{\frac{1}{3}}\big), \tag{6.2a}$$

$$a(\theta, T_1) = \mathsf{a}_0 + \mathsf{a}_1(\theta-\theta_*) + \mathsf{a}_2(\theta-\theta_*)^{\frac{4}{3}} + o\big((\theta-\theta_*)^{\frac{4}{3}}\big), \tag{6.2b}$$

for suitable constants computable constants $\mathsf{b} \approx 1$, a_i, and κ such that $|\kappa - \kappa_0| \lesssim \varepsilon^2$.

While the description of the preshock given by (6.2) would be likely sufficient to describe the classical shock singularity $\mathfrak{s}$ emerging from the preshock, in order to rigorously capture the formation of higher order characteristic singularities emerging along the curves $\mathfrak{s}_1$ and $\mathfrak{s}_2$ in Figure 3, a much finer understanding of the dominant Riemann variable w at the preshock is required. This information is not available in [3], and it is the subject of the analysis in [1, **SECTION 4**]. In particular, [1, **THEOREM 4.1**] proves that

$$w(\theta, T_1) = \kappa - \mathsf{b}(\theta-\theta_*)^{\frac{1}{3}} + \mathsf{c}_1(\theta-\theta_*)^{\frac{2}{3}} + \mathsf{c}_2(\theta-\theta_*) + \mathcal{O}\big((\theta-\theta_*)^{\frac{4}{3}}\big) \tag{6.3}$$

holds for all θ in an ε-dependent ball around θ_*, for explicitly computable constants c_i. More importantly, we prove that *the fractional series expansion* (6.3) *holds in a C^3 sense*, meaning that the first three derivatives of the left-hand side in (6.3) equal to the first three derivatives of the expansion on the right-hand side, with error bounds stable under differentiation.

The proof of (6.3) is based on a fully-Lagrangian characterization of the preshock, and a subtle interplay between the characteristics of the speeds $\lambda_3 = w$ and $\lambda_2 = \frac{2}{3}w$ present in (4.14), and which are defined by

$$\begin{aligned} \partial_t \eta &= \lambda_3\big(\eta(x,t),t\big) = w\big(\eta(x,t),t\big), \quad \eta(x,-\varepsilon) = x, \\ \partial_t \phi &= \lambda_2\big(\phi(x,t),t\big) = \frac{2}{3} w\big(\phi(x,t),t\big), \quad \phi(x,-\varepsilon) = x. \end{aligned}$$

By (4.14a), it is clear that η is the natural flow of the w evolution, while (4.14b) and (4.12), which simplify here to $\partial_t \varpi + \frac{2}{3} w \partial_\theta \varpi = \frac{8}{3} a \varpi$, show that ϕ is the natural flow for a and ϖ.

The first and most important observation is that the spacetime location of the first singularity (θ_*, T_1) is characterized by $\theta_* = \eta(x_*, T_1)$, where (x_*, T_1) are the unique Lagrangian label and the first time, respectively, which simultaneously solve the system

$$\partial_x \eta(x_*, T_1) = \partial_{xx} \eta(x_*, T_1) = 0. \tag{6.4}$$

In fact, as part of the proof it is crucial that we establish

$$\begin{aligned} \partial_x \eta(x,t) &= \big(1 + \mathcal{O}(\varepsilon^{\frac{1}{2}})\big)\varepsilon^{-1}(T_* - t) + \big(3 + \mathcal{O}(\varepsilon^{\frac{1}{8}})\big)\varepsilon^{-3}(x - x_*)^2, \\ \partial_{xx} \eta(x,t) &= (T_* - t)\mathcal{O}(\varepsilon^{-2}) + \big(6 + \mathcal{O}(\varepsilon^{\frac{1}{8}})\big)\varepsilon^{-3}(x - x_*), \\ \partial_{xxx} \eta(x,t) &= \big(6 + \mathcal{O}(\varepsilon^{\frac{1}{8}})\big)\varepsilon^{-3}, \end{aligned}$$

for all labels $|x - x_*| \le \varepsilon^2$ and all $t \in [-\varepsilon, T_1]$. This asymptotic description of the Lagrangian flow may be traced back to the initial datum assumption (6.1).

The second ingredient in the proof is that the fields η, $w \circ \eta$, $a \circ \eta$, $\varpi \circ \eta$ remain C^4 smooth as functions of the Lagrangian label x, *uniformly* in time on the interval $[-\varepsilon, T_1]$. Roughly speaking, this is achieved by appealing to the identities

$$\eta(x,t) = x + \int_{-\varepsilon}^{t} w \circ \eta(x,s) ds, \tag{6.5a}$$

$$w \circ \eta(x,t) = w(x,-\varepsilon) e^{-\frac{8}{3}\int_{-\varepsilon}^{t} a \circ \eta(x,s) ds}, \tag{6.5b}$$

which show that the regularity of $a \circ \eta$ implies the regularity of η and $w \circ \eta$, and to the one-derivative gains provided by the relations $\partial_\theta a = w - \frac{1}{16} w^2 \varpi$ and

$$\begin{aligned} \partial_x \phi(x,t) &= \left(\frac{w(x,-\varepsilon)}{w \circ \phi(x,t)}\right)^2 e^{-\frac{16}{3}\int_{-\varepsilon}^{t} a \circ \phi(x,s) ds}, \\ \varpi \circ \phi(x,t) &= \varpi_0(x,-\varepsilon) e^{\frac{8}{3}\int_{-\varepsilon}^{t} a \circ \phi(x,s) ds}, \end{aligned}$$

which in turn allows us to establish the desired higher order regularity of a and ϖ.

The third ingredient in the proof concerns the invertibility of the map $x \mapsto \eta(x, T_1)$. Using (6.4) and a Taylor series expansion justified by the regularity of η, we have that

$$\theta = \eta(x, T_1) = \theta_* + \frac{1}{6}\partial_{xxx}\eta(x_*, T_1)(x - x_*)^3 + \frac{1}{24}\partial_{xxxx}\eta(\bar{x}, T_1)(x - x_*)^4,$$

where $\theta_* = \eta(x_*, T_1)$, and $\bar{x}$ is a point between x_* and x. As such, with $\Theta = \theta - \theta_*$ and $X = x - x_*$, we are left to invert the quartic polynomial $\Theta = \mathrm{g}_1 X^3 + \mathrm{g}_2 X^4$, where $\mathrm{g}_1 \approx \varepsilon^{-3} > 0$ and $|\mathrm{g}_2| = \mathcal{O}(\varepsilon^{-4})$. This inversion via a Newton iteration results in a fractional power series $X = \mathrm{f}_1 \Theta^{1/3} + \mathrm{f}_2 \Theta^{2/3} + \mathrm{f}_3 \Theta + \mathcal{O}(\Theta^{4/3})$, with explicitly computable real coefficients f_i. This fractional power series is then directly translated into a power series expansion for the inverse map $\eta^{-1}(\theta, T_1)$ in powers of $(\theta - \theta_*)^{1/3}$, valid for θ sufficiently close to θ_*. At last, we insert this expansion into (6.5b), to obtain

$$w(\theta, T_1) = w\big(\eta^{-1}(\theta, T_1), -\varepsilon\big) e^{-\frac{8}{3} \int_{-\varepsilon}^{T_1} a \circ \eta(\eta^{-1}(\theta, T_1), s) ds}.$$

Using the known expansion for $\eta^{-1}(\cdot, T_1)$ and the regularity of $a \circ \eta$, we deduce (6.3).

7. OUTLINE: THE DEVELOPMENT OF SHOCKS AND WEAK SINGULARITIES

We next turn to the development problem, within the class of regular azimuthal shock solutions, cf. Definition 4.5. The initial datum for this development problem are the functions (w, z, k, a) at which we have arrived in the formation process at time T_1. For simplicity of the presentation, let us shift the preshock location (θ_*, T_1) to $(0, 0)$, and let us denote the values of the azimuthal fields at the preshock by (w_0, z_0, k_0, a_0). By the analysis in Section 6, we have that $z_0 \equiv k_0 \equiv 0$ on $\mathbb{T}$, $a_0 \in C^{1,1/3}(\mathbb{T})$ with $\|a_0\|_{W^{1,\infty}} \lesssim \kappa_0$, $\varpi_0 \in \mathrm{Lip}(\mathbb{T})$ with $1 < \kappa_0 \varpi_0(\theta) \lesssim 1$, and the dominant Riemann variable is given by

$$w_0(\theta) = \kappa - \mathsf{b}\theta^{\frac{1}{3}} + \mathsf{c}_1 \theta^{\frac{2}{3}} + \mathsf{c}_2 \theta + \mathcal{O}(\theta^{\frac{4}{3}}), \tag{7.1}$$

equality which holds in a C^3 sense, with $\kappa \approx \kappa_0 > 1$, $\mathsf{b} \approx 1$, and $\mathsf{c} = \mathcal{O}(\varepsilon^{1/2})$. The shock development problem from this initial data is solved on the interval $[0, \bar{\varepsilon}]$, i.e., $T_2 = T_1 + \bar{\varepsilon}$ in the language of Theorem 5.1, for a $\bar{\varepsilon}$ which is sufficiently small in terms of the data. The detailed analysis is carried out in [1, SECTIONS 5 AND 6], and here we only give the main ideas.

Given a smooth shock curve $\mathfrak{s}\colon [0, \bar{\varepsilon}] \to \mathbb{T}$, we shall denote the spacetime complement of the shock as $\mathcal{D}_{\bar{\varepsilon}} = (\mathbb{T} \times [0, \bar{\varepsilon}]) \setminus (\mathfrak{s}(t), t)_{t \in [0, \bar{\varepsilon}]}$, and for any function $f\colon \mathcal{D}_{\bar{\varepsilon}} \to \mathbb{R}$ we denote the left and right traces at the shock by $f_\pm(t) = \lim_{\theta \to \mathfrak{s}(t)^\pm} f(\theta, t)$, and the jump and mean across the shock as $[\![f]\!](t) = f_-(t) - f_+(t)$ and $\langle\!\langle f \rangle\!\rangle(t) = \frac{1}{2}(f_-(t) + f_+(t))$, respectively. Note that since $\bar{\varepsilon}$ is chosen to be sufficiently small, we have that $t \ll 1$ is a small parameter.

To leading order in $0 < t \ll 1$ and for $|\theta| \ll 1$, the *intuition* behind the shock development problem is as follows. First, from the Rankine–Hugoniot jump conditions one has that to leading order the speed of propagation of weak shock waves (relative to the fluid) is equal to the sound speed, which in the context of azimuthal symmetry means that

$$\begin{aligned} \dot{\mathfrak{s}} \approx b + c = w &\approx w_0 + (\text{small error for } t \ll 1) \\ &\approx \kappa + (\text{small error for } |\theta| \ll 1) + (\text{small error for } t \ll 1). \end{aligned}$$

Thus, to leading order we may expect that $\mathfrak{s}(t) \approx \kappa t$.

Second, we note that although entropy k and the subdominant Riemann variable z are strictly positive for $t > 0$, for short time they are expected to be small. As such, to leading order one may expect that the evolution of the dominant Riemann variable w (cf. (4.9a)) may be approximated as

$$\partial_t w + (w + \text{small error})\partial_\theta w = (\text{small errors involving entropy gradients}),$$
$$w_0 = \kappa - \mathsf{b}\theta^{\frac{1}{3}} + (\text{small error near for } |\theta| \ll 1).$$

Thus, we may hope to view the dominant Riemann variable w as being a perturbation of an inviscid Burgers solution w_{B} with associated Lagrangian η_{B}, namely

$$w_{\mathsf{B}}(\theta, t) = w_0\big({\eta_{\mathsf{B}}}^{-1}(\theta, t)\big), \quad \eta_{\mathsf{B}}(x, t) = x + t w_0(x). \tag{7.2}$$

Here we denote Eulerian space variable by θ and the Lagrangian label by x. There is an important caveat in the standard-looking definition (7.2). Since the initial data w_0 is a preshock (recall (7.1)), the map ${\eta_{\mathsf{B}}}^{-1}(\theta, t)$ is not well defined for θ which is very close to $\mathfrak{s}(t)$; indeed, in this region the map is two-valued. This is natural since these characteristics are expected to impinge upon the shock from either the left or the right, which ensures that the Lax entropy conditions (3.4) are satisfied. To overcome this, given any $t \in (0, \overline{\varepsilon}]$, and *given a shock curve* $\mathfrak{s}(t)$, we compute two Lagrangian labels $x_\pm(t) = {\eta_{\mathsf{B}}}^{-1}(\mathfrak{s}(t)^\pm, t)$ such that the associated particle trajectories $\eta_{\mathsf{B}}(x_\pm(t), s)$ fall into the shock exactly at time $s = t$. This allows us to define ${\eta_{\mathsf{B}}}^{-1}(\cdot, t)\colon \mathbb{T} \setminus \{\mathfrak{s}(t)\} \to \mathbb{T} \setminus [x_-(t), x_+(t)]$ as a bijective map, giving a meaning to (7.2). Note that to leading order one may compute $\eta_{\mathsf{B}}(x, t) \approx x + \kappa t - (\mathsf{b}t)x^{1/3}$, and since to leading order $\mathfrak{s}(t) \approx \kappa t$, we deduce that $x_\pm(t) \approx (\mathsf{b}t)^{3/2}$. It follows that we may expect the jump of the dominant Riemann variable across the shock curve to be given, to leading order in t, by

$$[\![w]\!](t) \approx [\![w_{\mathsf{B}}]\!](t) = w_0\big(x_-(t)\big) - w_0\big(x_+(t)\big) \approx 2\mathsf{b}^{\frac{3}{2}} t^{\frac{1}{2}}. \tag{7.3}$$

Third, in analogy to how (3.5) was derived, we may show that in the weak shock regime $|[\![w]\!]| \ll 1$ (justified in view of (7.3)) the smallest root (in absolute value) of the system of equations (4.16a)–(4.16b) (which were derived from the azimuthal form of the Rankine–Hugoniot conditions) is given to leading order by

$$[\![z]\!](t) \approx -\frac{9[\![w]\!](t)^3}{16\langle\!\langle w \rangle\!\rangle(t)^2} \approx -\frac{9\mathsf{b}^{\frac{9}{2}}}{2\kappa^2} t^{\frac{3}{2}} \quad \text{and} \quad [\![k]\!](t) \approx \frac{4[\![w]\!](t)^3}{\langle\!\langle w \rangle\!\rangle(t)^3} \approx \frac{32\mathsf{b}^{\frac{9}{2}}}{\kappa^3} t^{\frac{3}{2}}. \tag{7.4}$$

Just as (7.3), (7.4) may be shown to hold in a C^2_t sense. The jump relations show that positive entropy and negative subdominant Riemann variable must be *produced instantaneously* along the shock in order for mass, momentum, and energy not to be lost.

Fourth, we need to carefully analyze the three characteristic families present in the azimuthal Euler equations (4.9)–(4.10). These flows are defined naturally as

$$\partial_t \eta = \lambda_3 \circ \eta, \quad \partial_t \phi = \lambda_2 \circ \phi, \quad \partial_t \psi = \lambda_1 \circ \psi, \quad (\eta, \phi, \psi)(x, 0) = x.$$

Our heuristics indicate that to leading order in $t \ll 1$ and $|x| \ll 1$ we have that

$$\eta(x, t) \approx \eta_{\mathsf{B}}(x, t) \approx x + \kappa t - (\mathsf{b}t)x^{\frac{1}{3}}, \quad \phi(x, t) \approx x + \frac{2\kappa}{3} t, \quad \psi(x, t) \approx x + \frac{\kappa}{3} t, \tag{7.5}$$

which confirms our intuition that the λ_3 characteristic η impinges on the shock curve $\mathfrak{s}(t) \approx \kappa t$ only after we look at the next order term in t and x, and also that the λ_2 and λ_1 characteristics ϕ and ψ are transversal to the shock. Note that the two characteristic surfaces of weak singularities are nothing but the images under these slow flows of the point-shock

$$\mathfrak{s}_2(t) = \phi(0,t) \approx \frac{2\kappa}{3}t, \quad \mathfrak{s}_1(t) = \psi(0,t) \approx \frac{\kappa}{3}t.$$

The transversality of characteristic families mentioned above plays a crucial role in our analysis: it may be combined with the fact that we stay away from the vacuum state in order to interchange a space derivative with a time derivative in terms which are composed with ϕ or ψ. For example, it allows us to heuristically replace the statements $[\![z]\!] \sim -t^{3/2}$ and $[\![k]\!] \sim t^{3/2}$ from (7.4), with asymptotic descriptions $z(\theta,t) \sim -(\theta - \mathfrak{s}_1(t))^{3/2}$ and $k(\theta,t) \sim (\theta - \mathfrak{s}_2(t))^{3/2}$ asymptotically as $\theta \to \mathfrak{s}_1(t)^+$ and $\theta \to \mathfrak{s}_2(t)^+$, respectively. Thus, the jump relations (7.4) and transversality imply that the fields z and k form $C^{1,1/2}$ cusps at $\mathfrak{s}_1$ and $\mathfrak{s}_2$, when approaching from the right.

Besides determining the location of the weak singularities, the flows η, ϕ, ψ also paint a detailed picture as to how information is carried from the $\{t = 0\}$ initial data surface, respectively how information about the jumps at the shock are propagated through the fluid in spacetime. A schematic description is provided by Figure 6 below.

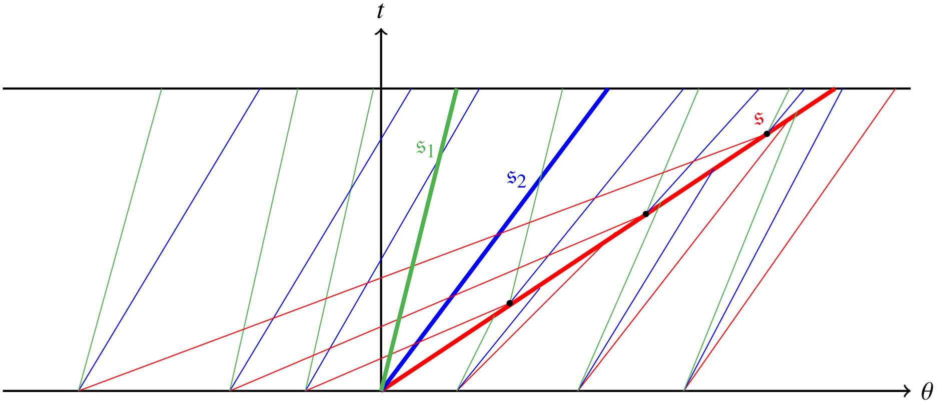


FIGURE 6

The three distinct wave families η, ϕ, and ψ are represented in red, blue, and respectively green, for various initial labels. The most interesting such labels are marked with black dots: these do not lie on the time-slice $\{t = 0\}$, but instead they lie on the shock curve $\mathfrak{s}$ at various values of time; at these points the values of k_- and z_- are computed according to (7.4). To leading order, the entropy k is propagated off the shock curve along the λ_2 characteristics ϕ, while the subdominant Riemann variable z is also propagated off the shock curve $\mathfrak{s}$, but along the λ_1 characteristics ψ. The λ_3 characteristics η initiated at $\{t = 0\}$, represented in red, impinge on the shock curve from the left side, determining w in terms of w_0 on both sides of the shock.

Fifth, we note that according to (4.9d) and (4.12), the fluid velocity λ_2 and its associated characteristic ϕ are the natural ones for carrying information about the radial velocity a and the specific entropy ϖ. In particular, since ϕ is transversal to $\mathfrak{s}$, we are able to use (4.12) in order to show that the specific vorticity is continuous across the shock curve. As such, the relation (4.11) implies that it is $\partial_\theta a$ and not a which has a jump discontinuity at $\mathfrak{s}$, and, moreover, to leading order we have

$$[\![\partial_\theta a]\!](t) \approx [\![w]\!](t) \approx 2\mathsf{b}^{\frac{3}{2}} t^{\frac{1}{2}}. \tag{7.6}$$

Sixth, concerning the characterization of the higher order singularities across the curves $\mathfrak{s}_1$ and $\mathfrak{s}_2$, the intuition regarding the precise regularity of the fields (w, z, k, a) stems from the jump relations (7.3), (7.4), (7.6), a detailed description of the Lagrangian flows ϕ and ψ similar to (7.5), and the structure of the forcing terms in (4.9) and (4.12). For instance, we have already mentioned in the paragraph below (7.5) that the transversality of ϕ and ψ to $\mathfrak{s}$, along with the jump relations (7.4) allow us to precisely compute the regularity of z and k approaching $\mathfrak{s}$ from the left. This matter is, however, more subtle near $\mathfrak{s}_1$ and $\mathfrak{s}_2$. To see this, we may inspect Figure 6 and note that an Eulerian point (θ, t) with $0 < \theta - \mathfrak{s}_2(t) \ll 1$ is traced backwards in time along the blue characteristics ϕ to a point which lies on the shock curve at some time $\mathsf{T}(\theta, t) \sim \theta - \mathfrak{s}_2(t) \ll 1$ (*shock-intersection times* are defined precisely in [1, DEFINITIONS 5.15 AND 5.16]). Thus, singular information about the derivatives of the jumps of k at a time $\mathsf{T}(\theta, t) \ll t$ is carried via the ϕ characteristics to the point (θ, t), resulting in infinite terms as $\theta \to \mathfrak{s}_2(t)^+$.

An additional difficulty in analyzing the higher-order singularities is that, if we naively consider the evolution equations for $\partial_\theta w$ or $\partial_\theta z$, cf. (4.9a) and respectively (4.9b), we note the emergence of the forcing term $\frac{1}{24}(w - z)^2 \partial_{\theta\theta} k$, resulting in what seems to be a derivative loss. In order to overcome this issue, we introduce the *good unknowns*

$$q^w := \partial_\theta w - \frac{1}{4} c \partial_\theta k, \quad q^z := \partial_\theta z + \frac{1}{4} c \partial_\theta k,$$

which satisfy the evolution equations

$$(\partial_t + \lambda_3 \partial_\theta) q^w + \left(\partial_\theta \lambda_3 + \frac{8}{3} a\right) q^w = -\frac{8}{3} \partial_\theta a w + \left(\frac{4}{3} ac + \frac{1}{6} c \partial_\theta \lambda_2\right) \partial_\theta k, \tag{7.7a}$$

$$(\partial_t + \lambda_1 \partial_\theta) q^z + \left(\partial_\theta \lambda_1 + \frac{8}{3} a\right) q^z = -\frac{8}{3} \partial_\theta a z - \left(\frac{4}{3} ac + \frac{1}{6} c \partial_\theta \lambda_2\right) \partial_\theta k. \tag{7.7b}$$

The remarkable feature of the system (7.7) is that the second derivatives of k do not appear in the equations, allowing us to close estimates. The unknowns q^w and q^z are useful because they involve only the first derivative of the entropy, $\partial_\theta k$, and this term makes a $C^{\frac{1}{2}}$ cusp along the curve $\mathfrak{s}_2$. On the other hand, the natural flows in the system (7.7) are η and ψ, respectively, which are transversal to the flow ϕ along which the singularities of k are carried through the flow. This geometric structure of (7.7) and of the good unknowns q^w and q^z analytically result in a one-derivative regularization effect, which is not apparent if we were to inspect (4.9)–(4.12) directly. Another outcome of this derivative gain is that $q^w + q^z = \partial_\theta z + \partial_\theta w = \frac{2}{3} \partial_\theta u_\theta$ is smoother than the naive expectation $C^{\frac{1}{2}}$ because the $\partial_\theta k$ terms cancel. This translates into at least C^2 regularity for the angular velocity u_θ along the curve $\mathfrak{s}_2$; in contrast, the entropy S, the density ρ and the radial velocity u_r are precisely $C^{1,1/2}$ across $\mathfrak{s}_2$, which justifies the name *weak contact singularity*.

In closing, we note that making the six-step heuristic outlined in this section rigorous requires a good functional framework and a number of analytical tricks for the analysis of Lagrangian flows. In broad terms, we proceed as follows. We build an iteration scheme in which we start with a C^2 smooth shock curve $\mathfrak{s}$ with $|\mathfrak{s}(t) - \kappa t| \lesssim t^2$, use it to construct a Burgers solution w_{B} adapted to this particular shock curve (as described in Step 2), and then use a contraction mapping principle to build a solution (w, z, k, a) of the azimuthal Euler equations (4.9)–(4.12) which has jump discontinuities across $\mathfrak{s}$ that satisfy the algebraic system (4.16a)–(4.16b) resulting from the Rankine–Hugoniot jump conditions, and such that the regularity of the solution is consistent with the fact that the solution emanates from a $C^{1/3}$ preshock. More precisely, there exists a sufficiently small $\bar{\varepsilon}$ such that the solution lies in the functional space

$$\mathcal{X}_{\bar{\varepsilon}} = \Big\{(w, z, k, a) \in C^1_{\theta,t}(\mathcal{D}_{\bar{\varepsilon}}) : (w, z, k, a)|_{t=0} = (w_0, 0, 0, a_0), \\ \big\| (w - w_{\mathsf{B}}, z, k, a) \big\|_{\bar{\varepsilon}} \le 1 \Big\} \tag{7.8}$$

where the norm $\|(v, z, k, a)\|_{\bar{\varepsilon}}$ is defined by

$$\begin{aligned} \big\|(v, z, k, a)\big\|_{\bar{\varepsilon}} = \sup_{(\theta,t)\in\mathcal{D}_{\bar{\varepsilon}}} \max\Big\{ & \mathsf{m}_1 t^{-1} \big|v(\theta, t)\big|, \mathsf{m}_2 \big(\mathsf{b}^3 t^3 + \big(\theta - \mathfrak{s}(t)\big)^2\big)^{\frac{1}{6}} \big|\partial_\theta v(\theta, t)\big|, \\ & \mathsf{m}_3 t^{-\frac{3}{2}} \big|z(\theta, t)\big|, \mathsf{m}_3 t^{-\frac{1}{2}} \big|\partial_\theta z(\theta, t)\big|, \mathsf{m}_4 t^{-\frac{3}{2}} \big|k(\theta, t)\big|, \\ & \mathsf{m}_4 t^{-\frac{1}{2}} \big|\partial_\theta k(\theta, t)\big|, \mathsf{m}_5 \big|a(\theta, t)\big|, \mathsf{m}_5 \big|\partial_\theta a(\theta, t)\big| \Big\} \end{aligned}$$

where m_i are sufficiently large constants. In particular, we note that the space $\mathcal{X}_{\bar{\varepsilon}}$ encodes precisely how close w is to the Burgers solution w_{B}.

So far, we have thus defined a map $\mathfrak{s} \mapsto (w, z, k, a)$, but we are missing one key ingredient: the shock curve was just a given curve with $|\mathfrak{s}(t) - \kappa t| \lesssim t^2$, it did not satisfy the evolution equation (4.16c) imposed by the Rankine–Hugoniot jump conditions. This, however, gives us a natural way of updating the shock curve: we solve for $\tilde{\mathfrak{s}}$ the ODE (4.16c) with data $\tilde{\mathfrak{s}}(0) = 0$ and fields (w_+, w_-, z_-, k_-) given by the restrictions of (w, z, k, a) on the old curve $\mathfrak{s}$. Then, we prove that $\tilde{\mathfrak{s}}$ is C^2 smooth and satisfies $|\tilde{\mathfrak{s}}(t) - \kappa t| \lesssim t^2$. Lastly, we prove that above described iteration $\mathfrak{s} \mapsto \tilde{\mathfrak{s}}$ is in fact a contraction in C^2, resulting in a unique fixed point which is the desired shock curve. Associated to this curve, we also prove that there is a unique regular azimuthal shock solution $(w, z, k, a) \in \mathcal{X}_{\bar{\varepsilon}}$, as soon as $\bar{\varepsilon} > 0$ is sufficiently small. This completes the proof of Theorem 5.1.

FUNDING

T. B. was supported by the NSF grant DMS-1900149 and a Simons Foundation Mathematical and Physical Sciences Collaborative Grant. T. D. was supported by NSF grant DMS-2106233. S. S. was supported by NSF grant DMS-2007606 and the Department of Energy Advanced Simulation and Computing (ASC) Program. V. V. was supported by the NSF CAREER grant DMS-1911413.

REFERENCES

[1] T. Buckmaster, T. D. Drivas, S. Shkoller, and V. Vicol, Simultaneous development of shocks and cusps for 2D Euler with azimuthal symmetry from smooth data. 2021, arXiv:2106.02143.

[2] T. Buckmaster, S. Shkoller, and V. Vicol, Formation of point shocks for 3D compressible Euler, 2019, arXiv:1912.04429. *Comm. Pure Appl. Math.* (to appear).

[3] T. Buckmaster, S. Shkoller, and V. Vicol, Formation of shocks for 2D isentropic compressible Euler, 2019, arXiv:1907.03784. *Comm. Pure Appl. Math.*, DOI 10.1002/cpa.21956, in print.

[4] T. Buckmaster, S. Shkoller, and V. Vicol, Shock formation and vorticity creation for 3d Euler, 2020, arXiv:2006.14789. *Comm. Pure Appl. Math.* (to appear).

[5] S. Chen and L. Dong, Formation and construction of shock for p-system. *Sci. China Ser. A* **44** (2001), no. 9, 1139–1147.

[6] D. Christodoulou, *The formation of shocks in 3-dimensional fluids*. EMS Monogr. Math., European Mathematical Society (EMS), Zürich, 2007.

[7] D. Christodoulou, *The shock development problem*. EMS Monogr. Math., European Mathematical Society (EMS), Zürich, 2019.

[8] D. Christodoulou and A. Lisibach, Shock development in spherical symmetry. *Ann. PDE* **2** (2016), no. 1, Art. 3, 246.

[9] D. Christodoulou and S. Miao, *Compressible flow and Euler's equations*. Surv. Mod. Math. 9, International Press, Somerville, MA; Higher Education Press, Beijing, 2014.

[10] R. Courant and K. O. Friedrichs, *Supersonic flow and shock waves*. Interscience, New York, 1948.

[11] C. M. Dafermos, *Hyperbolic conservation laws in continuum physics*. Grundlehren Math. Wiss. 325, Springer, Berlin, 2010.

[12] T. D. Drivas and G. L. Eyink, An Onsager singularity theorem for turbulent solutions of compressible Euler equations. *Comm. Math. Phys.* **359** (2018), no. 2, 733–763.

[13] J. Eggers and M. A. Fontelos, *Singularities: formation, structure, and propagation*. Cambridge Texts Appl. Math., Cambridge University Press, 2015.

[14] T. Kato, The Cauchy problem for quasi-linear symmetric hyperbolic systems. *Arch. Ration. Mech. Anal.* **58** (1975), no. 3, 181–205.

[15] C. Klingenberg, O. Kreml, V. Mácha, and S. Markfelder, Shocks make the Riemann problem for the full Euler system in multiple space dimensions ill-posed. *Nonlinearity* **33** (2020), no. 12, 6517.

[16] D.-X. Kong, Formation and propagation of singularities for 2×2 quasilinear hyperbolic systems. *Trans. Amer. Math. Soc.* **354** (2002), no. 8, 3155–3179.

[17] L. D. Landau and E. M. Lifshitz, *Fluid mechanics*. Pergamon Press, Oxford, 1987.

[18] M. P. Lebaud, Description de la formation d'un choc dans le p-systéme. *J. Math. Pures Appl.* **73** (1994), no. 6, 523–565.

[19] J. Luk and J. Speck, Shock formation in solutions to the 2D compressible Euler equations in the presence of non-zero vorticity. *Invent. Math.* **214** (2018), no. 1, 1–169.

[20] J. Luk and J. Speck, The stability of simple plane-symmetric shock formation for 3D compressible Euler flow with vorticity and entropy. 2021, arXiv:2107.03426.

[21] B. Riemann, Über die Fortpflanzung ebener Luftwellen von endlicher Schwingungsweite. *Abh. Königlichen Ges. Wiss. Göttingen* **8** (1860), 43–66.

[22] H. Yin, Formation and construction of a shock wave for 3-D compressible Euler equations with the spherical initial data. *Nagoya Math. J.* **175** (2004), 125–164.

TRISTAN BUCKMASTER
Department of Mathematics, Princeton University, Princeton, NJ 08544, USA, tjb4@math.princeton.edu

THEODORE D. DRIVAS
Department of Mathematics, Stony Brook University, Stony Brook, NY 11794, USA, tdrivas@math.stonybrook.edu

STEVE SHKOLLER
Department of Mathematics, University of California Davis, Davis, CA 95616, USA, shkoller@math.ucdavis.edu

VLAD VICOL
Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA, vicol@cims.nyu.edu

SELECTED TOPICS IN MEAN FIELD GAMES

PIERRE CARDALIAGUET AND FRANÇOIS DELARUE

ABSTRACT

Mean field game theory was initiated a little more than 15 years ago with the aim of simplifying the search for Nash equilibria in games with a large number of weakly interacting players. Since then, a lot has been done. Numerous equilibrium existence results have been obtained, using different characterizations and in various contexts. The analysis of the master equation, which describes the evolution of the value of the game, has also seen significant progress, which has, for example, allowed establishing in certain cases the convergence of games with a finite number of players. However, mean field games remain of a complex nature. For instance, the typical lack of uniqueness of solutions raises selection issues that are still poorly understood. The objective of the note is to present some of the latest advances, as well as some avenues to address further challenging questions.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 91A13; Secondary 91A15, 60H30

KEYWORDS

Mean field games, master equation, convergence problem, common noise, long time behaviour

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3660–3703
DOI 10.4171/ICM2022/17

The theory of mean field games (MFG) aims at providing an asymptotic description of differential games with a large number of interacting players. The number of applications of the theory is huge, ranging from macroeconomics to crowd motions, and from finance to power grid models. In all these models, each player controls his/her own dynamical state which evolves in time according to a deterministic or stochastic differential (or difference if the state space is at most countable) equation. The individual goal is to minimize some cost depending on his/her own control but also on the behavior of the whole population of agents, which is described through the empirical distribution of their states. In this setting, the central concept is the notion of Nash equilibria, which explains how agents play in an optimal way by taking into account the others' strategies. The MFG theory is precisely intended to simplify the search for these Nash equilibria. In this respect, the key idea is to postulate that, asymptotically, the single theoretical (and not empirical) statistical distribution of the states is sufficient to compute the individual goal of each player.

The MFG theory was introduced and largely developed by Lasry and Lions through a series of papers around 2005 and during the famous lectures of Lions at the Collège de France [85–88]. At about the same time, Caines, Malhamé, and Huang discussed similar models under the terminology of "Nash certainty equivalence principle" [73,74]. The MFG theory is also reminiscent of the so-called heterogenous agent models developed in economics at the end of the 1990s by Aiyagari [7] and by Krusell and Smith [79] or, more recently, by Lucas and Moll [91]. One of the main achievements of the MFG theory—though not discussed here—is a better formulation and understanding of these models (see, for instance, Achdou et al. [1]). After a decade and a half of research, the theory has answered—at least partially—several important questions and has developed a number of mathematical techniques and tools for this purpose. A large part of the material can be found in the monographs or in the surveys [6,15,25,35,36,71].

From a mathematical perspective, the MFG theory lies at the intersection of probability and partial differential equations (PDEs). The connection between games with finitely many players and MFGs is addressed by means of statistical averaging arguments, which are made possible by the symmetric structure of the interactions. This approach is, of course, reminiscent of the very typical issues and techniques underpinning the standard mean field theory and the related propagation of chaos properties for large weakly interacting particle systems (see [76,92] for the earliest papers in the field and [101] for a review). However, unlike the standard mean field theory, in which the interacting particles obey a given dynamics, the dynamics of the agents is not given a priori in the MFG theory but rather is obtained after an optimization procedure. This seemingly innocuous difference dramatically increases the level of complexity of the problem, as it introduces several nonlinearities in the equations describing the mean field models. These nonlinearities manifest themselves in several ways, depending on the formulation used to characterize the equilibria and, implicitly, on the approach chosen to manage the optimization step in the definition of these equilibria. In this respect, let us say that both probabilistic and PDE arguments have been successfully developed. In short, the probabilistic approach aims at following the dynamics of a reference player in the population, while the PDE one aims at following the dynamics of the statistical state of the whole population. The key feature is that both approaches lead to the study of a form of forward–backward system that couples either two stochastic differential equations or two PDEs: the probabilistic system is usually referred to as a forward–backward McKean–Vlasov system and the PDE system is usually known as the "MFG system." Regardless of the system, the strong coupling between the forward and backward components therein raises many issues. Obviously, one knows in general how to pass from one approach to the other and, generally speaking, PDE tools are useful to obtain better regularity of the solutions. Very importantly, these two systems can be regarded as the characteristics of a common infinite-dimensional PDE of hyperbolic type set on the space of probability measures. It is called the "master equation." This master equation has become a challenging object in the field and has attracted much attention in analysis, probability, and calculus of variation. At present, it is only well-understood in certain cases where the solutions are known to be regular. A theory allowing less regular solutions and thus covering a wider scope is totally lacking. Needless to say, this is a very exciting area of research.

In addition to the analysis of mean field games themselves, the study of the convergence problem, namely the convergence of games with a finite number of players to a mean field game, is another challenge, which has also required the development of appropriate arguments. As already mentioned, this asks for a nontrivial adaptation of the existing results on the convergence of weakly interacting particle systems. Among others, a key idea is to test classical solutions of the master equation onto the equilibria of the games with finitely many players. The main contributions in this direction are presented in the notes, but many questions remain open. To wit, solutions to mean field games are typically nonunique and identifying those that are selected by taking the limit in large games is a fascinating, but really difficult question.

Before presenting the rest of the contents of these notes in a more exhaustive way, we insist on the fact that the MFG theory provides a concept that has proven to be effective in the analysis of some typical examples of game theory. However, the same concept can be applied to many other cases. We give an overview of some of them at the very end of the notes. For example, mean field games with common noise is an extension of the original concept that has stimulated many recent works. In short, this corresponds to the case where the state of the population itself is random. Understanding the precise impact of noise on equilibria is another challenge in the field. To emphasize the importance of this research direction, we have therefore decided to write these notes by systematically including common noise in the models we present. We hope that this will help the reader to grasp the essence of it.

Contents. After a short presentation of the PDE formulation of MFGs in Section 1, we concentrate ourselves on the following three fundamental aspects of the theory:

1) The analysis of the convergence problem, which, as we have said, investigates how Nash equilibria in differential games with finitely many players converge to MFG equilibria. This point is essential to justify the MFG models and is one of the main mathematical achievement of the MFG theory. We provide an overview in Section 2, which includes a presentation of the master equation.

2) The long time behavior of the MFG equilibria. Since time-dependent models are difficult to handle and to approximate numerically, the analysis of "stationary models" and their robustness is essential in both theory and application. For instance, economists often concentrate on these stationary solutions. We present the main results in Section 3.

3) The regularizing aspect of the common noise in MFG. Since the MFG equilibria are in general not unique, it is crucial to understand the extent to which a common noise can force uniqueness. This question is addressed in Section 4.

We complete the notes by providing in the final Section 5 a general overview of other topics from MFG theory that are not discussed in the first four sections. We give some references that may be useful for the reader and we provide some open problems.

Notation. We denote by $\mathcal{P}_2(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$ with a finite second order moment, endowed with the Wasserstein distance (see, for instance, [9]). If $x \in \mathbb{R}^d$, we denote by δ_x the Dirac mass at x. For X a random variable, we denote by $\mathcal{L}(X)$ the law of X.

1. THE MFG EQUILIBRIA

In this section we introduce the main problems of the MFG theory. The simplest for this is to start with a game with a large number of players and then to pass (at least formally) to the limit as the number of players tends to infinity.

1.1. The N-player problem

The N-player game. Let $N \in \mathbb{N}$, with $N \geq 1$ being the (large) number of players. Player i (where $i \in \{1, \dots, N\}$) controls her own state X_t^i, which is an element of $\mathbb{R}^d$ and evolves in time according to the stochastic differential equation (SDE)

$$dX_t^i = \alpha_t^i + \sqrt{2} dB_t^i + \sqrt{2\epsilon} dW_t,$$

for prescribed initial conditions $(X_0^i)_{i=1,\dots,N}$. Here the processes $((B_t^i)_{t\geq 0})_{i=1,\dots,N}$ and $(W_t)_{t\geq 0}$ are independent d-dimensional Brownian motions. The noise $(B_t^i)_t$, which affects only the dynamics of player i, is called the idiosyncratic (or the individual) noise. The Brownian motion $(W_t)_t$, on the contrary, impacts all the dynamics and is called the common noise; the nonnegative real ϵ denotes (up to the square root) the intensity of the effective common noise that is felt by all the players. The initial conditions $(X_0^i)_{i\geq 1}$ are independent and identically distributed (i.i.d.) random variables with common distribution $\tilde{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$. We assume that the random variables $(X_0^i)_{i=1,\dots,N}$ and the Brownian motions $((B_t^i)_t)_{i=1,\dots,N}$ and W are independent. Player i chooses a bounded control $(\alpha_t^i)_t$ that takes values in $\mathbb{R}^d$ and that is adapted to the filtration $(\mathbb{F}_t = \sigma\{X_0^j, B_s^j, W_s, s \leq t, j = 1, \dots, N\})$.

The cost of player i is given by

$$\mathcal{J}_i^N\big(\alpha^i,(\alpha^j)_{j\neq i}\big)=\mathbb{E}\bigg[\int_0^T\Big(\frac{1}{2}|\alpha_t^i|^2+F(X_t^i,m_{\boldsymbol{X}_t}^{N,i})\Big)dt+G(X_T^i,m_{\boldsymbol{X}_T}^{N,i})\bigg],$$

where $\boldsymbol{X}_t=(X_t^1,\dots,X_t^N)$ and $m_{\boldsymbol{X}_t}^{N,i}=\frac{1}{N-1}\sum_{j=1,\dots,N,j\neq i}\delta_{X_t^j}$. To fix the ideas, we work here with a finite-horizon problem (where $T>0$ is the horizon) and we assume the maps $F:\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ and $G:\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\to\mathbb{R}$ to be continuous and bounded. Here we make the assumption that the running cost of player i depends only on her own control, her own position, and on the distribution of the other players' positions, while the terminal cost depends only on her position and on the distribution of the other players' positions at terminal time. The important point is the symmetry of the problem: for a player, the other players play exactly the same role. The specific form of the cost and dynamics is made here for simplicity.

Nash equilibria. In that setting, a natural notion of equilibrium is the the notion of *Nash equilibrium*. We say that a family $(\bar\alpha^1,\dots,\bar\alpha^N)$ of (time-dependent stochastic) controls is a Nash equilibrium of the N-player game if, for any $i\in\{1,\dots,N\}$ and any control α^i,

$$\mathcal{J}_i^N\big(\bar\alpha^i,(\bar\alpha^j)_{j\neq i}\big)\le\mathcal{J}_i^N\big(\alpha^i,(\bar\alpha^j)_{j\neq i}\big).$$

We are intentionally fuzzy in the definition of what a control is. There are actually many possibilities and we feel better to restrict ourselves to two of them. The controls can be either (i) open-loop, which means that they are regarded as adapted functions of the initial conditions $(X_0^i)_i$ and of the noises $((B_t^i)_t)_i$ and W, or (ii) closed-loop controls, in which case they are considered as adapted functions of the trajectories $((X^i)_t)_i$ (when the closed-loop structure is Markov, the dependence just occurs through the current states of the players). The main difference between the two notions is as follows: when one player deviates, the function underpinning the definition is kept fixed. As such, the controls played by the other players remain the same in the open-loop case while they change in the closed-loop case. In the rest of the note, we always mean Markov closed-loop control when speaking about a closed-loop control.

The Nash system. A key fact with games involving closed-loop controls is that they have a PDE interpretation, in the form of a system of equations for the equilibrium value of the game. In our setting, one can show that if $v^N:[0,T]\times(\mathbb{R}^d)^N\to\mathbb{R}^N$ is the classical solution to the following backward parabolic system (called here the Nash system)

$$\begin{cases}-\partial_t v_t^{N,i}-\displaystyle\sum_{j=1}^N\Delta_{x_j}v_t^{N,i}-\epsilon\sum_{j,k=1}^N\mathrm{Tr}(D^2_{x_jx_k}v_t^{N,i})+\frac{1}{2}|D_{x_i}v_t^{N,i}|^2\\ \qquad+\displaystyle\sum_{j\neq i}D_{x_i}v_t^{N,i}\cdot D_{x_j}v_t^{N,j}=F(x_i,m_{\boldsymbol{x}}^{N,i})\quad\text{in }(0,T)\times(\mathbb{R}^d)^N,\quad i\in\{1,\dots,N\},\\ v_T^{N,i}(\boldsymbol{x})=G(x_i,m_{\boldsymbol{x}}^{N,i})\qquad\qquad\text{in }(\mathbb{R}^d)^N,\quad i\in\{1,\dots,N\},\end{cases}\tag{1.1}$$

then $(\bar{\alpha}^i(t, \boldsymbol{x}) := -D_{x_i} v_t^{N,i}(\boldsymbol{x}))_{i=1,\dots,N}$ is a Nash equilibrium of the N-player game in closed-loop form. Here, the notation $\boldsymbol{x}$ stands for an N-tuple $(x_1, \dots, x_N)$, in which case x_i is the entry number i in $\boldsymbol{x}$. The existence and uniqueness of the solution to the above system, called the Nash system, is classical under suitable assumptions on F and G and discussed, for instance, in [84]. Under similar assumptions on F and G, this equilibrium can be shown to be unique (within the class of bounded Markov closed-loop controls); see, for instance, [36, CHAPTER 6].

The main question raised by MFG theory is the characterization of the limit, as N tends to infinity, of the Nash equilibria of the game (or of the Nash system) and the analysis of the resulting limit.

1.2. The MFG equilibria

In this part we derive from the N-player problem several (equivalent) formulations of an MFG equilibrium. The derivation is formal at this stage, but will be justified more rigorously in Section 2.

MFG equilibria without common noise ($\epsilon = 0$). The are several ways to guess and write the limit of the Nash equilibrium or of the Nash system as N tends to infinity. We start with the problem without common noise, which is easier to grasp. As players are symmetric, one can expect, using classical ideas of mean field theory [101], that the in-equilibrium trajectories $((\overline{X}_t^{N,i})_t)_i$ associated with the Nash equilibrium identified right above become more and more decorrelated as N increases and eventually become asymptotically independent. In this case the empirical measure $m^{N,i}_{\overline{\boldsymbol{X}}_t^N}$ should become asymptotically deterministic and, as N gets larger and larger, the impact of the deviation of a player over $m^{N,i}_{\overline{\boldsymbol{X}}_t^N}$ should be negligible. Therefore players can solve their own optimization problem as if $m^{N,i}_{\overline{\boldsymbol{X}}_t^N}$ were given and independent of i. Implementing this idea, one finds the notion of MFG equilibrium in its probabilistic formulation:

Probabilistic formulation of the MFG equilibrium ($\epsilon = 0$). One searches for a pair (m, α), where $m = (m_t)_t \in C^0([0, T], \mathcal{P}_2(\mathbb{R}^d))$, and $\alpha = (\alpha_t)_t$ is a control such that

(i) α is optimal for the control problem

$$\inf_{\beta} \mathbb{E}\left[\int_0^T \left(\frac{1}{2}|\beta_t|^2 + F(X_t^\beta, m_t)\right) dt + G(X_T^\beta, m_T)\right], \tag{1.2}$$

where the infimum is taken over the controls $\beta = (\beta_t)_t$ (that are $(X_0^1, (B_t^1)_t)$-progressively measurable) and where X^β is the solution to

$$dX_t^\beta = \beta_t dt + \sqrt{2} dB_t^1, \quad X_0^\beta = X_0^1. \tag{1.3}$$

(ii) For any $t \in [0, T]$, the law of X_t^α is m_t.

Other probabilistic formulations of MFG equilibria are possible: Carmona and Delarue discuss in [34] a formulation involving the stochastic maximum principle. Mainly, the optimizers in item (i) are described by means of a forward–backward stochastic differential equation depending on the input $(m_t)_t$. Under the fixed point condition (ii), this input is identified with the marginal law of the solution of the forward equation, which gives rise to a so-called forward–backward system of McKean–Vlasov type. While Pontryagin's principle provides the dynamics of an equilibrium feedback along a corresponding equilibrium trajectory, an alternative approach is to provide a representation of the equilibrium value. This approach is usually known as the weak formulation as it may rely on a convenient change of noise in the dynamics. In short, it provides another form of the forward–backward system of McKean–Vlasov type, see Carmona and Lacker [39] and [35, CHAPTER 3]. The latter is useful for proving existence results. In comparison, the stochastic Pontryagin principle provides, in general, only sufficient conditions satisfied by an arbitrary equilibrium.

PDE formulation of the MFG equilibria: the MFG system ($\epsilon = 0$). Another characterization of the MFG equilibria goes through a forward–backward system of PDEs known as the MFG system: the unknown are (u, m) where u corresponds to the value function associated with the optimal control problem described in the probabilistic formulation while m solves the Kolmogorov equation satisfied by the marginal law of the equilibrium. It reads therefore

$$\begin{cases} -\partial_t u_t(x) - \Delta u_t(x) + \frac{1}{2}\left|Du_t(x)\right|^2 = F(x, m_t) & \text{in } (0,T) \times \mathbb{R}^d, \\ \partial_t m_t(x) - \Delta m_t(x) - \operatorname{div}(m_t(x) Du_t(x)) = 0 & \text{in } (0,T) \times \mathbb{R}^d, \\ m_0(x) = \tilde{m}_0, \quad u_T(x) = G(x, m_T) & \text{in } \mathbb{R}^d. \end{cases} \tag{1.4}$$

This system is unusual: the first equation (a Hamilton–Jacobi equation) is backward in time, while the Kolmogorov equation is forward in time. The main issue is that both equations are strongly coupled, in the sense that each of the two unknowns shows up in the other equation. Since the two equations are set in opposite time directions, this creates a conflict which makes the spice of the analysis. The existence of a solution has been proved by Lasry and Lions [85–87] under suitable assumptions on the coupling functions F and G (regularity and growth conditions). In general, there is no uniqueness: this is a typical feature of equilibria in game theory (in contrast, uniqueness holds in the finite game because of the smoothing effect of the Laplacians in the related Nash system (1.1); we will come back to this observation in Section 4). However, the solution of (1.4) is unique if the following monotonicity condition, introduced in [85–87], is satisfied:

$$\begin{aligned} &\int_{\mathbb{R}^d} \big(F(x,m) - F(x,m')\big)(m - m')(dx) \geq 0, \\ &\int_{\mathbb{R}^d} \big(G(x,m) - G(x,m')\big)(m - m')(dx) \geq 0. \end{aligned} \tag{1.5}$$

There is by now a huge literature on the MFG system, including different types of coupling functions, different types of boundary conditions, etc. We briefly present some aspects of this literature in Section 5.1.

The MFG equilibria with common noise ($\epsilon > 0$). In the presence of common noise, the heuristic analysis of the limit problem is more subtle. Indeed, even if the players do not take into account the idiosyncratic noises of the other players, their dynamics are perturbed by the common noise $(W_t)_t$. Therefore, the limit $(m_t)_t$ (if it exists) of the marginal empirical measure $(m^{N,i}_{\mathbf{X}^N_t})_t$ associated with the equilibrium trajectories $((\overline{X}^{N,j}_t)_t)_{j\neq i}$ becomes random and is typically expected to be adapted to the Brownian motion $(W_t)_t$ (very much as before, this limit is expected to be independent of i).

Probabilistic formulation of the MFG equilibrium with common noise ($\epsilon > 0$). One searches for a pair (m, α), where the stochastic process $(m_t)_t$ is adapted to $(W_t)_t$ and takes values in $C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$ and $\alpha = (\alpha_t)_t$ is a control such that

(i) α is optimal for the control problem

$$\inf_{\beta} \mathbb{E}\left[\int_0^T \left(\frac{1}{2}|\beta_t|^2 + F(X_t^\beta, m_t)\right)dt + G(X_T^\beta, m_T)\right], \tag{1.6}$$

where the infimum is taken over the controls $\beta = (\beta_t)_t$ (that are $(X_0^1, (B_t^1, W_t)_t)$-progressively measurable) and where X^β is the solution to

$$dX_t^\beta = \beta_t dt + \sqrt{2} dB_t^1 + \sqrt{2\epsilon} W_t, \quad X_0^\beta = X_0^1. \tag{1.7}$$

(ii) For any $t \in [0,T]$, the (conditional) law of X_t^α given $(W_s)_s$ is m_t.

In general, it is difficult to prove the existence of MFG equilibria because the fixed-point condition (ii) is defined, in the presence of common noise, on a very wide space. To overcome this issue, a possible path is to discretize the common noise into a noise with finitely many outcomes (see [38]). In that case, it is much easier to adapt the arguments used when $\epsilon = 0$. However, much may be lost when passing to the limit over the discretization of the common noise. Very similar to weak solutions to stochastic differential equations, equilibria that are obtained in this way may no longer be adapted with respect to the original common noise $(W_t)_t$. This requires a relevant notion of weak MFG equilibria, in which the flow of measures $(m_t)_t$ is adapted to a larger filtration than that generated by $(W_t)_t$. When the monotonicity property (1.5) is in force, it can be proved that these weak solutions are in fact strong, i.e., they are adapted with respect to $(W_t)_t$.

PDE formulation of the MFG equilibria with common noise: the stochastic MFG system ($\epsilon > 0$). In the probabilistic formulation of the MFG equilibria with a common noise, the optimal control problem (1.6)–(1.7) (which is solved by a reference player in the population) is driven by random coefficients (because $(m_t)_t$ is random). The associated value function is no longer deterministic. Following Peng [96], it should be regarded as the solution of a backward stochastic Hamilton–Jacobi equation. Moreover, the Kolmogorov equation satisfied by the random flow $(m_t)_t$ is stochastic. The resulting MFG system therefore reads:

$$\begin{cases} du_t = \Big[-(1+\epsilon)\Delta u_t + \frac{1}{2}|Du_t|^2 - F(x,m_t) - 2\epsilon\,\mathrm{div}(v_t)\Big]dt \\ \qquad\quad + v_t \cdot \sqrt{2\epsilon}\,dW_t & \text{in } (0,T)\times\mathbb{R}^d, \\ dm_t = \big[(1+\epsilon)\Delta m_t + \mathrm{div}(m_t Du_t)\big]dt - \mathrm{div}(m_t\sqrt{2\epsilon}\,dW_t) & \text{in } (0,T)\times\mathbb{R}^d, \\ m_0(x) = \tilde{m}_0, \quad u_T(x) = G(x,m_T) & \text{in } \mathbb{R}^d. \end{cases} \tag{1.8}$$

Note that now the unknown is the triplet (u,v,m). As explained in Peng [96], the role of the random field v is to ensure the solution u to the backward Hamilton–Jacobi equation to be adapted to the common noise $(W_t)_t$. The existence of a solution for (1.8) is subtle and has been achieved, under suitable conditions on F and G including monotonicity, in [25] (see also [36]).

2. THE MASTER EQUATION AND THE CONVERGENCE OF THE NASH SYSTEM

In this part we address the rigorous derivation of the MFG equilibria and the convergence of the Nash system. This analysis requires the introduction of a new equation, the master equation, which is a nonlinear equation stated on the infinite-dimensional space $\mathcal{P}_2(\mathbb{R}^d)$. In order to restrict the technicality of the exposition, we will often be fuzzy in the assumption and in the statement of the results and refer to [25,36], that we follow closely, for details.

2.1. Derivatives of maps defined on the space of probability measures

There are several notions of derivatives for a map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$: we refer, for instance, to [8,9,25,35] and the references therein for several possible notions together with an overview of the connections between all of them. Here we mostly discuss an idea of Lions which consists in lifting the map U to a suitable space of random variables.

Let us consider the space $L^2 := L^2((\Omega,\mathbb{F},\mathbb{P}),\mathbb{R}^d)$ of square-integrable random variables on $\mathbb{R}^d$, with Ω being a Polish space, $\mathbb{F}$ its Borel σ-algebra, and $\mathbb{P}$ an atomless probability measure. The space L^2 is endowed with the usual Hilbert scalar product. It is known that, for any $m \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a random variable X with law m.

Given a map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$, we lift U to L^2 by setting

$$\tilde{U}(X) = U\big(\mathcal{L}(X)\big) \quad \forall X \in L^2.$$

Definition 2.1 (The L-derivative). We say that U is L-differentiable at $m \in \mathcal{P}_2(\mathbb{R}^d)$ if there exists a random variable $X \in L^2$ with law m such that $\tilde{U}$ is Fréchet differentiable at X (we denote by $\nabla\tilde{U}(X)$ its gradient).

Theorem 2.1 (Structure of the L-derivative). *Assume that U is L-differentiable at $m \in \mathcal{P}_2(\mathbb{R}^d)$. Then there exists a map $D_m U(m,\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ which is Borel measurable* *and such that*

$$\nabla \tilde{U}(X) = D_m U(m, X)$$

for any random variable X with $\mathcal{L}(X) = m$. We call the map $D_m U(m, \cdot)$ the L-derivative of U at m.

The first version of this result goes back to Lions [88]. The version given here is due to Gangbo and Tudorascu [65], who also explain the connection with the notion of subdifferential introduced in [9].

Finite dimensional projection. A key principle to establish the link between the Nash system and the master equation is to associate with any function defined on $\mathcal{P}_2(\mathbb{R}^d)$ a finite dimensional projection, whose definition is as follows.

Given a continuous map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and a nonzero integer N, we define the projection U^N of U as the map $U^N : (\mathbb{R}^d)^N \to \mathbb{R}$ defined by

$$U^N(x_1, \ldots, x_N) = U(m_{\boldsymbol{x}}^N), \quad \text{where } m_{\boldsymbol{x}}^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \ \boldsymbol{x} = (x_1, \ldots, x_N) \in (\mathbb{R}^d)^N.$$

The following statement clarifies the meaning of the derivative D_m:

Proposition 2.2. *Assume that U is L-differentiable with a Lipschitz-continuous derivative. Then U^N is of class C^1 and*

$$D_{x_i} U^N(x_1, \ldots, x_N) = \frac{1}{N} D_m U(m_{\boldsymbol{x}}^N, x_i),$$

for $(x_1, \ldots, x_N) \in (\mathbb{R}^d)^N$.

One can, of course, introduce higher-order derivatives of a map $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ in a similar way and extend Proposition 2.2 to higher-order derivatives, see [35, CHAPTER 5].

Itô's formula along a flow of conditional measures. The following Itô's formula, needed in the proofs below and of independent interest, is a generalization of Itô rule for flows of measures and functions defined on the space of measures. Let $(X_t)_{t \geq 0}$ be an Itô process of the form

$$dX_t = b_t dt + \sigma_t dB_t + \sigma_t^0 dW_t, \quad t \geq 0, \tag{2.1}$$

with a given (possibly random) initial condition X_0, where $(B_t)_t$ and $(W_t)_t$ are two d-dimensional Brownian motions, X_0, $(B_t)_t$, and $(W_t)_t$ being independent. Above, $(b_t)_t$, $(\sigma_t)_t$, and $(\sigma_t^0)_t$ are progressively-measurable processes with respect to the filtration generated by X_0, $(B_t)_t$, and $(W_t)_t$, with values in (respectively) $\mathbb{R}^d$, $\mathbb{R}^{d \times d}$, and $\mathbb{R}^{d \times d}$. For simplicity, we assume that that the probability space is given in a product form $(\Omega_0 \times \Omega_1, \mathbb{F}_0 \otimes \mathbb{F}_1, \mathbb{P}_0 \otimes \mathbb{P}_1)$, where $(\Omega_0, \mathbb{F}_0, \mathbb{P}_0)$ supports W, while $(\Omega_1, \mathbb{F}_1, \mathbb{P}_1)$ supports (X_0, B). We denote by $\mathbb{E}^0$ the expectation with respect to $\mathbb{P}^0$ and by $\mathbb{E}^1$ the expectation with respect to $\mathbb{P}^1$. We assume that

$$\mathbb{E}\left[|X_0|^2 + \int_0^T \left(|b_t|^2 + |\sigma_t|^4 + \left|\sigma_t^0\right|^4 \right) dt \right] < +\infty,$$

where $\mathbb{E} = \mathbb{E}^0 \mathbb{E}^1$.

The following result is taken from [36] (see also [20, 45]).

Theorem 2.3. *Let $(X_t)_t$ be as in (2.1) and, for any $t \geq 0$, let m_t be the conditional law of X_t given $(W_s)_s$. Then, for $U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ a sufficiently smooth mapping on $\mathcal{P}_2(\mathbb{R}^d)$,*

$$\begin{aligned} U(m_t) = U(m_0) &+ \int_0^t \mathbb{E}^1\big[D_m U(m_s, X_s) \cdot b_s ds\big] + \int_0^t \mathbb{E}^1\big[(\sigma^0)^* D_m U(m_s, X_s)\big] \cdot dW_s \\ &+ \frac{1}{2} \int_0^t \mathbb{E}^1\big[\mathrm{Tr}\big(D_y D_m U(m_s, X_s)\big(\sigma_s \sigma_s^* + \sigma_s^0 (\sigma_s^0)^*\big)\big]ds \\ &+ \frac{1}{2} \int_0^t \mathbb{E}^1 \tilde{\mathbb{E}}^1\big[\mathrm{Tr}\big(D_m^2 U(m_s, X_s, \tilde{X}_s)\sigma_s^0 (\tilde{\sigma}_s^0)^*\big)\big]ds, \end{aligned}$$

where $\tilde{X}_s$ and $\tilde{\sigma}_s^0$ are independent copies of X_s and σ_s^0 defined on $\Omega_0 \times \tilde{\Omega}_1$, for a copy $\tilde{\Omega}_1$ of Ω_1 equipped with the expectation $\tilde{\mathbb{E}}^1$.

Briefly, $D_y D_m U$ is the y-derivative of the function $y \mapsto D_m U(m, y)$ for a fixed m. Similarly, D_m^2 is the m-derivative of the function $m \mapsto D_m U(m, y)$ for a fixed y, which implies that $D_m^2 U$ can be written in the form $(m, y, y') \mapsto D_m^2 U(m, y, y')$. Under the regularity assumptions mentioned in the statement, all these derivatives exist implicitly and are jointly continuous. They also satisfy appropriate growth conditions that permit giving a meaning to the various expectations appearing in the expansion. The symbol Tr is for the trace.

Potential games. We feel it useful to provide another application of the derivative D_m. There is indeed one special class of mean field games, for which the corresponding MFG system coincides with the first-order condition (or equivalently, with the Pontryagin system) of a control problem. Such games are called potential games, and the control problem lying above a potential game is usually called a mean field control problem. The connection between both can be thus formulated in this way: The minimizers to the mean field control problem are equilibria of the corresponding potential game. This was noted in the earlier articles by Lasry and Lions [85–87], see also [88].

The potential structure turns out to be very useful in practice for the simple reason that it might be easier to work with minimizers than with Nash equilibria. We provide a longer discussion in Section 4 about possible applications to the selection of equilibria when there is no uniqueness.

In the simple framework of (1.2)–(1.4), the potential game typically requires that the cost coefficients F and G derive from a potential, namely

$$\partial_x F(x, m) = D_m \mathcal{F}(m, x), \quad \partial_x G(x, m) = D_m \mathcal{G}(m, x), \tag{2.2}$$

for two smooth functionals $\mathcal{F}$ and $\mathcal{G}$ on $\mathcal{P}_2(\mathbb{R}^d)$. With the trajectory $(X_t^\beta)_t$ as in (1.3), we can associate the cost

$$\mathcal{J}\big((\beta_t)_{0 \leq t \leq T}\big) = \int_0^T \Big(\mathcal{F}\big(\mathcal{L}(X_t)\big) + \frac{1}{2}\mathbb{E}\big[|\beta_t|^2\big]\Big)dt + \mathcal{G}\big(\mathcal{L}(X_T)\big).$$

The following statement may be found under more precise assumptions in [35, CHAPTER 6]:

Proposition 2.4. *Under suitable regularity properties on $\mathcal{F}$ and $\mathcal{G}$, and for given initial distribution $\tilde{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$ for X_0 in (1.3), the optimal trajectories of $\mathcal{J}$ with respect to $(X_0, (B_t)_t)$-progressively measurable controls $(\beta_t)_t$ are solutions of the mean field game* (1.2)–(1.3).

When $(\beta_t)_t$ is identified with a feedback function $\beta : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, equation (1.3) becomes a stochastic differential equation whose marginal law solves the Kolmogorov equation

$$\partial_t m_t - \Delta m_t + \operatorname{div}(\beta_t m_t) = 0,$$

with $m_0 = \tilde{m}_0$. It is then possible to write $\mathcal{J}$ as

$$\mathcal{J}\big((\beta_t)_{0\le t\le T}\big) = \int_0^T \bigg(\mathcal{F}(m_t) + \frac{1}{2}\int_{\mathbb{R}^d} \big|\beta_t(x)\big|^2 m_t(dx)\bigg)dt + \mathcal{G}(m_T),$$

which is more in line with the formulation (1.4) of the mean field game. Although the resulting class of controls is obviously smaller when restricted to feedback controls, the infimum of $\mathcal{J}$ is the same, see Lacker [82].

2.2. The master equation

The master equation was first derived by Lions in [88] as the formal limit of the Nash system (1.1). It is a PDE with unknown $U : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ (with U writing $(t,x,m) \mapsto U(t,x,m)$) and reads

$$\begin{cases} \text{(i)} \quad -\partial_t U_t - (1+\epsilon)\Delta_x U_t \\ \qquad + \dfrac{1}{2}|D_x U_t|^2 + \displaystyle\int_{\mathbb{R}^d} D_m U_t(t,x,m,y) \cdot D_x U_t(t,y,m) m(dy) \\ \qquad - (1+\epsilon) \displaystyle\int_{\mathbb{R}^d} \operatorname{div}_y (D_m U_t(t,x,m,y)) m(dy) \\ \qquad - 2\epsilon \displaystyle\int_{\mathbb{R}^d} \operatorname{div}_x (D_m U_t(t,x,m,y)) m(dy) \\ \qquad - \epsilon \displaystyle\int_{\mathbb{R}^d \times \mathbb{R}^d} \operatorname{Tr}_y (D^2_{mm} U_t(t,x,m,y,y')) m(dy) m(dy') = F(x,m) \\ \qquad\qquad \text{in } (0,T) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d), \\ \text{(ii)} \quad U_T(x,m) = G(x,m) \quad \text{in } \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d). \end{cases} \tag{2.3}$$

This is a kind of hyperbolic equation stated on the infinite-dimensional space $\mathcal{P}_2(\mathbb{R}^d)$. Indeed, when F and G are monotone (recall (1.5)), the solution can be (at least formally) built by the method of characteristics. To ease the presentation, let us explain this when there is no common noise ($\epsilon = 0$). Let $(t_0, m_0) \in [0,T) \times \mathcal{P}_2(\mathbb{R}^d)$ and $(u_t, m_t)_t$ be the unique solution of the MFG system (1.4) stated on $(t_0, T) \times \mathbb{R}^d$ with initial condition $m_{t_0} = \tilde{m}_0$. Let us set $U_{t_0}(x, m_0) = u_{t_0}(x)$. Assuming that U is sufficiently smooth, one can easily check that U solves (2.3) by expanding it along the path $(m_t)_t$ (see [25]). The main issue is to prove that U is indeed smooth. When there is a common noise ($\epsilon > 0$), similar ideas can be implemented, using the stochastic MFG system (1.8) instead of the deterministic one.

Let us loosely summarize the main result concerning (2.3) (see [25, 36]).

Theorem 2.5. *Assume that F and G are smooth enough and are monotone in the sense of* (1.5)*. Then there exists a unique classical solution to* (2.3)*.*

By "a classical solution," we mean a map U for which all the derivatives in (2.3) exist, are bounded and globally Lipschitz continuous. This strong notion of solution is needed below for the convergence results.

It is known that, if one removes the monotonicity condition, then the solution of (2.3) exists on a short time interval but may develop a discontinuity after a while. Most formal properties of the master equation have been introduced and discussed by Lions in [88], who also introduced the so-called Hilbertian approach (lifting the equation to the space of random variables). The actual proof of the existence of a solution of the master equation is a tedious verification that the map U, as defined above from the MFG system, actually gives rise to a classical solution. This required several steps in the literature before the proof was completed: The first paper in this direction is [20], where the classical solutions to the linear Kolmogorov equation associated with a standard Fokker–Planck equation are studied; Gangbo and Swiech [64] address the master equation in short time and without any diffusion term; Chassagneux et al. [45] obtain the existence and uniqueness for the master equation without common noise; Cardaliaguet et al. [25] establish the existence and uniqueness of solutions for the master equation with common noise under the monotonicity condition (see also [36]). Since then, there have been many works on the subject. We provide some references in Section 5.1.

2.3. Convergence of the Nash system

One key feature of the master equation is that it allows building approximate solutions of the Nash system (1.1) whose regularity is independent of the number of players.

Proposition 2.6. *Assume that U is a classical solution of* (2.3) *and let $(u^{N,i})_{i\in\{1,\dots,N\}}$ be its finite-dimensional projections:*

$$u^{N,i}_t(x_1,\dots,x_N) = U_t(x_i, m^{N,i}_{\boldsymbol{x}}), \quad \textit{where } m^{N,i}_{\boldsymbol{x}} := \frac{1}{N-1}\sum_{j=1,\dots,N:j\neq i} \delta_{x_j}.$$

Then u^N almost solves the Nash system (1.1)*:*

$$\begin{cases} -\partial_t u^{N,i}_t - \displaystyle\sum_{j=1}^N \Delta_{x_j} u^{N,i}_t - \epsilon \sum_{j,k=1}^N \mathrm{Tr}(D^2_{x_j x_k} u^{N,i}_t) + \frac{1}{2}\big|D_{x_i} u^{N,i}_t\big|^2 \\ \quad + \displaystyle\sum_{j\neq i} D_{x_i} u^{N,i}_t \cdot D_{x_j} u^{N,j}_t = F(x_i, m^{N,i}_{\boldsymbol{x}}) + r^{N,i}_t(\boldsymbol{x}) \\ \qquad\qquad\qquad\qquad \textit{in } (0,T)\times(\mathbb{R}^d)^N, \quad i\in\{1,\dots,N\}, \\ u^{N,i}_T(\boldsymbol{x}) = G(x_i, m^{N,i}_{\boldsymbol{x}}) \quad \textit{in } (\mathbb{R}^d)^N, \quad i\in\{1,\dots,N\}, \end{cases}$$

where

$$\left|r_t^{N,i}(\boldsymbol{x})\right| \le \frac{1}{N}\left(1 + \frac{1}{N}\sum_{i=1}^{N}|x_i - x_j|\right).$$

The proof relies on Proposition 2.2 and on its extension to higher-order derivatives. Note, however, that the result does not show directly that u^N is close to the solution v^N of (1.1) because each $u^{N,i}$ solves (1.1) only up to an error term of size $1/N$ while the system counts exactly N equations.

The main convergence result of [25] and [36] is the following:

Theorem 2.7. *Let v^N be the solution of the Nash system and assume that U is a classical solution of* (2.3) *with bounded derivatives. Then there exists a constant $C > 0$ such that, for all $i \in \{1, \dots, N\}$ and all $(t, \boldsymbol{x}) \in [0, T] \times (\mathbb{R}^d)^N$,*

$$\left|U_t(x_i, m_{\boldsymbol{x}}^{N,i}) - v_t^{N,i}(\boldsymbol{x})\right| \le \frac{C}{N}\left(1 + |x_i|^2 + \frac{1}{N}\sum_{j=1}^{N}|x_j|^2\right)^{1/2}.$$

Theorem 2.7 provides an obvious comparison between the equilibrium values of the finite game and of the mean field game. Even though it is not obvious at first sight, the statement is in fact reminiscent of earlier results on the convergence of classical mean field particle systems to Fokker–Planck equations. An alternative strategy to the standard coupling argument for proving propagation of chaos (see [101]) consists indeed in studying the action of the semigroup generated by the McKean–Vlasov equation onto the marginal empirical measure of the particle system (see Kolokoltsov [78] and the works of Mouhot, Mischler, and Wennberg [94]). In comparison, the game setting involves an additional optimization step, which makes the analysis really difficult. In order to account for this optimization step, we work instead with forward–backward McKean–Vlasov equations, following the approach developed in [34,36]. We describe the main lines below.

Sketch of the proof of Theorem 2.7. The first step is to provide a probabilistic representation of the solution v^N of the Nash system. This goes through the representation of the equilibrium paths. To this end, we recall that $(\bar{\alpha}^i(t, \boldsymbol{x}) := -D_{x_i} v^{N,i}(t, \boldsymbol{x}))_{i=1,\dots,N}$ is the Nash equilibrium of the N-player game in closed-loop form. For a given starting point $\boldsymbol{x} = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$, the equilibrium trajectories $\boldsymbol{X}^{*N} = (X^{*N,i})_{i\in\{1,\dots,N\}}$ associated to the Nash equilibrium are the solutions to the system

$$\begin{cases} dX_t^{*N,i} = -D_{x_i} v_t^{N,i}(\boldsymbol{X}_t^{*N})dt + \sqrt{2}dB_t^i + \sqrt{2\epsilon}dW_t, \\ X_0^{*N,i} = x_i. \end{cases} \tag{2.4}$$

Adopting a Lagrangian point of view, we may then follow the evolution of the cost and of the control along the system, which prompts us to let

$$Y_t^{*N,i} = v_t^{N,i}(\boldsymbol{X}_t^{*N}), \quad Z_t^{*N,i,j} = D_{x_j} v_t^{N,i}(\boldsymbol{X}_t^{*N}).$$

Classical Itô's formula, combined with the form of the Nash system, leads to the following expansion:

$$dY_t^{*N,i} = -\Big(\frac{1}{2}\big|Z_t^{*N,i,i}\big|^2 + F(X_t^{*N,i}, m^{N,i}_{\boldsymbol{X}_t^{*N}})\Big)dt + \sqrt{2}\sum_{j=1}^N Z_t^{*N,i,j} \cdot (dB_t^j + \sqrt{\epsilon} dW_t). \tag{2.5}$$

In order to test, as suggested before, the action of the solution U to the master equation (which is somehow the analogue of the semigroup generated by a McKean–Vlasov equation but in the nonlinear setting induced by the game structure) on the Nash equilibrium of the N-player game, we need to perform a similar computation, but for the processes

$$\mathcal{Y}_t^{*N,i} = u^{N,i}(t, X_t^{*N}), \quad \mathcal{Z}_t^{*N,i,j} = D_{x_j} u_t^{N,i}(X_t^{*N}), \quad t \in [0,T],$$

with $(u^{N,i})_i$ being as in Proposition 2.6. In fact, Proposition 2.6 now permits expanding $(\mathcal{Y}_t^{*N,i})_t$. We get

$$\begin{aligned} d\mathcal{Y}_t^{*N,i} = &- \Big(\frac{1}{2}\big|\mathcal{Z}_t^{*N,i,i}\big|^2 + F(X_t^{*N,i}, m^{N,i}_{\boldsymbol{X}_t^{*N}}) + r_t^{N,i}(\boldsymbol{X}_t^{*N,i})\Big)dt \\ &+ \sum_{j=1}^N \mathcal{Z}_t^{*N,i,j} \cdot (\mathcal{Z}_t^{*N,j,j} - Z_t^{*N,j,j})dt + \sqrt{2}\sum_{j=1}^N \mathcal{Z}_t^{*N,i,j} \cdot (dB_t^j + \sqrt{\epsilon} dW_t). \end{aligned}$$

Importantly, the two processes $(Y_t^{*N,i})_t$ and $(\mathcal{Y}_t^{*N,i})_t$ satisfy the same boundary conditions at time T, namely $Y_T^{*N,i} = \mathcal{Y}_T^{*N,i} = g(X_T^{*N,i}, m^{N,i}_{\boldsymbol{X}_T^{*N}})$, which prompts us to address the difference process $(Y_t^{*N,i} - \mathcal{Y}_t^{*N,i})_{0 \le t \le T}$. We get

$$\begin{aligned} d(\mathcal{Y}_t^{*N,i} - Y_t^{*N,i}) = &- \Big(\frac{1}{2}\big|\mathcal{Z}_t^{*N,i,i}\big|^2 - \frac{1}{2}\big|Z_t^{*N,i,i}\big|^2 + r_t^{N,i}(\boldsymbol{X}_t^{*N,i})\Big)dt \\ &+ \sum_{j=1}^N \mathcal{Z}_t^{*N,i,j} \cdot (\mathcal{Z}_t^{*N,j,j} - Z_t^{*N,j,j})dt \\ &+ \sqrt{2}\sum_{j=1}^N (\mathcal{Z}_t^{*N,i,j} - Z_t^{*N,i,j}) \cdot (dB_t^j + \sqrt{\epsilon} dW_t). \end{aligned} \tag{2.6}$$

The last term yields a stochastic integral. If there were no dt-term in the right-hand side, then the simple fact that the terminal condition is equal to 0 would say that the stochastic integral is also null. In turn, this would say that $\mathcal{Z}_t^{*N,i,j} - Z_t^{*N,i,j} = 0$ for any t. In other words, the noise provides a strong form of stability in the above equation. This is consistent with the fact that, in the Nash system, the Laplace operator dissipates the energy when time runs backwards. The sum on the second line is also challenging, at least at first sight. However, Proposition 2.2 says that, except when $j = i$, all the terms are of order $1/N$, which guarantees that the whole sum is of order 1. On the first line of the right-hand side, the remainder $r_t^{N,i}$ is also known to be of order $1/N$ on compact sets. In the end, we are thus left with a backward stochastic differential inequation of the form

$$
\begin{aligned}
d(\mathcal{Y}_t^{*N,i} - Y_t^{*N,i}) = & -\Bigg[\frac{1}{2}\big|\mathcal{Z}_t^{*N,i,i}\big|^2 - \frac{1}{2}\big|Z_t^{*N,i,i}\big|^2 \\
& + O\Bigg(\frac{1}{N} + \frac{1}{N^2}\sum_{i,j=1}^{N}\big|X_t^{*i,N} - X_t^{*j,N}\big|\Bigg)\Bigg]dt \\
& + O\Bigg(\frac{1}{N}\sum_{j=1}^{N}\big|\mathcal{Z}_t^{*N,j,j} - Z_t^{*N,j,j}\big| + \big|\mathcal{Z}_t^{*N,i,i} - Z_t^{*N,i,i}\big|\Bigg)dt \\
& + \sqrt{2}\sum_{j=1}^{N}(\mathcal{Z}_t^{*N,i,j} - Z_t^{*N,i,j})\cdot(dB_t^j + \sqrt{\epsilon}dW_t).
\end{aligned}
$$

Above, the symbol $O(\cdot)$ is used for the Landau notation, the underlying constant being (in our setting) deterministic and independent of N and t. Obviously, the goal is to provide a stability analysis of this equation. Needless to say, the main difficulty in this regard is the difference of the two quadratic terms on the first line of the right-hand side. Invoking Proposition 2.2 once again and using the fact that the solution to the master equation is assumed to have bounded derivatives, it is pretty easy to get L^∞-bounds on the process $(\mathcal{Z}_t^{*N,i,i,})_t$, independently of i and N. However, there are no similar inequalities for the process $(Z_t^{*N,i,i,})_t$. This is in fact the main challenge in this proof: Known estimates on the regularity of $v^{N,i}$, and in particular on its gradient, depend on N. Accordingly, most of the analysis relies on the sole properties of the solution U to the master equation. In words, there is no easy way here to linearize the difference of the two quadratic terms in the backward equation. The idea is then to adapt some of the tricks that have been developed in the literature on backward stochastic differential equations with a quadratic dependence on the martingale representation term (here denoted by $(\mathcal{Z}_t^{*N,i,j} - Z_t^{*N,i,j})_t$). In the analysis of the well-posedness of a backward stochastic differential equation, quadratic growth (with respect to the same martingale representation term) is indeed known to be a threshold. This is consistent with the results on nonlinear parabolic PDEs: quadratic growth in the gradient of the solution is also known to be a threshold. Noticeably, the unknown in the backward equation should be in fact regarded as being multidimensional since it comprises all the coordinates $(\mathcal{Y}_t^{*N,i} - Y_t^{*N,i})_{i=1,\dots,N}$. In general, this is known to render the analysis in the quadratic case even more challenging. Anyway, the symmetric structure of the equation here is very helpful and somehow permits thinking as if the equation were set in dimension 1. In the end, a suitable form of exponential transform (very much inspired from the Cole–Hopf transform in the analysis of Hamilton–Jacobi–Bellman equations, see [36, CHAPTER 6] for the details) allows transforming the quadratic equation into a linear one, and then concluding by using standard stability arguments from the theory of backward stochastic differential equations. Essentially, the size of the difference terms $((\mathcal{Y}_t^{*N,i} - Y_t^{*N,i})_t)_{i=1,\dots,N}$ is dictated by the remainder in the equation and is thus of order $1/N$. It then remains to observe that that, at time $t=0$,

$$
\mathcal{Y}_0^{*N,i} - Y_0^{*N,i} = u_0^{N,i}(\boldsymbol{x}) - v_0^{N,i}(\boldsymbol{x}) = U_0(x_i, m_{\boldsymbol{x}}^{N,i}) - v_0^{N,i}(\boldsymbol{x}).
$$

Our sketch of proof hence shows that the left-hand side is of order $1/N$. In fact, a careful inspection would permit tracking the dependence on the initial conditions and recovering the same rate as in the statement. ■

2.4. Propagation of chaos for the N-player game

In fact, the proof of Theorem 2.7 kills two birds with one stone. Indeed, it also permits addressing the large-N behavior of the equilibrium trajectories of the N-player game. Recall indeed from (2.4) that these equilibrium trajectories solve the system of stochastic differential equations

$$dX_t^{*N,i} = -D_{x_i} v_t^{N,i}(X_t^{*N})dt + \sqrt{2}dB_t^i + \sqrt{2\epsilon}dW_t, \tag{2.7}$$

for a given choice of initial conditions. In order to state propagation of chaos in a proper manner, we assume, as in our preliminary description of a mean field game in Section 2, that these initial conditions are given as independent samples $X_0^1, \dots, X_0^N$ from a common distribution $\tilde{m}_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

Noticeably, the drift in (2.7) may be rewritten in terms of the notations introduced in the proof of Theorem 2.7. Indeed, this drift is nothing but $(-Z_t^{*N,i,i})_{0 \le t \le T}$, which is a key quantity in the proof of Theorem 2.7. It is then worth emphasizing that stability arguments for backward stochastic differential equations like those we used in this proof provide more than what is eventually contained in the result. They also provide a similar bound on the quadratic variation (or, equivalently, on the energy) of the martingale representation term in (2.6). Using the fact that $\tilde{m}_0$ is square-integrable, we end up with the fact that

$$\mathbb{E}\int_0^T \left|\mathcal{Z}_t^{*N,i,i} - Z_t^{*N,i,i}\right|^2 dt \le \frac{C}{N^2},$$

for a constant C that is independent of N. Implicitly, the constant C depends on $\tilde{m}_0$ through its second-order moment. Moreover, it is worth recalling that, on the left-hand side, $\mathcal{Z}_t^{*N,i,i} = -D_x U_t(X_t^{*N,i}, m^{N,i}_{\boldsymbol{X}^{*N}})$. In turn, this says that, up to an error of order $1/N$, we can replace the drift in (2.7) by $-D_x U_t(X_t^{*N,i}, m^{N,i}_{\boldsymbol{X}^{*N}})$. Equivalently, by using the regularity properties of $D_x U$, we have

$$\sup_{i=1,\dots,N} \mathbb{E}\Big[\sup_{0\le t\le T} \left|\mathcal{X}_t^{*N,i} - X_t^{*N,i}\right|^2\Big] \le \frac{C}{N^2}, \tag{2.8}$$

where

$$\begin{cases} d\mathcal{X}_t^{*N,i} = -D_x U_t(\mathcal{X}_t^{*N,i}, m^{N,i}_{\boldsymbol{\mathcal{X}}_t^{*N}})dt + \sqrt{2}dB_t^i + \sqrt{2\epsilon}dW_t, \\ \mathcal{X}_0^i = X_0^i. \end{cases} \tag{2.9}$$

Very differently from (2.7), whose structure is made intricate by the presence of v^N, (2.9) is a standard weakly interacting particle system. As such, it is known to converge to the solution of the conditional McKean–Vlasov equation

$$\begin{cases} d\mathcal{X}_t^i = -D_x U_t\big(\mathcal{X}_t^i, \mathcal{L}(\mathcal{X}_t^i|W)\big)dt + \sqrt{2}dB_t^i + \sqrt{2\epsilon}dW_t, \\ \mathcal{X}_0^i = X_0^i. \end{cases}$$

The analysis of the above equation is standard. Under our standing assumptions on U, it follows from a classical contraction argument. In particular, uniqueness for the above equation implies that the conditional law $\mathcal{L}(\mathcal{X}_t^i|W)$ that appears in the dynamics is in fact independent of i. Sznitman's coupling argument [101] then allows estimating the distance between the solution of (2.9) and the solution of the above conditional McKean–Vlasov equation. We get

Theorem 2.8. *For any $\eta > 0$, there exists a constant $C_\eta > 0$ such that, for all $N \geq 1$ and for all $i \in \{1, \dots, N\}$,*

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\big|X^{*N,i} - X_t^i\big|\Big] \leq C_\eta N^{-1/\max\{d,2+\eta\}}.$$

When $d \geq 3$, we can choose $\eta = 0$.

Noticeably, the rate in Theorem 2.8 is much weaker than the rate in (2.8). In fact, the bound in Theorem 2.8 is the same as the bound for the mean 1-Wasserstein distance between a probability distribution in $\mathcal{P}_2(\mathbb{R}^d)$ and the empirical law of an independent sample of it. We refer to Fournier and Guillin's idea [62] for a complete review of the subject.

Some bibliographical comments on the propagation of chaos in Nash equilibria are now in order. The first results concerning this question are due to Fischer [61] and Lacker [81] for open-loop controls (in which players observe only the initial states and the Brownian motions) in problems without common noise: Lacker [81], in particular, identified completely the possible limits, which are always MFG equilibria (in a weak form, with a notion of weak solution similar to [38]). The question of convergence of closed-loop equilibria is more subtle. As shown in a counterexample in [36, I.7.2.5] (inspired from [56]), this convergence does not hold in full generality. At present, the minimal conditions to obtain it are still not clear. Theorem 2.8, proved first in the periodic setting in [25] and then extended to the Euclidean framework in [35], shows that the convergence holds if there exists a classical solution (with bounded derivatives) to the master equation (which implies that equilibria are unique) and if the idiosyncratic noise is nondegenerate (which implies that it is not null). In the same framework (and with $\mathbb{R}^d$ as state space), Delarue et al. [51] and [50] established a central limit theorem and a large deviation principle, using the same idea as in the proof of Theorem 2.8: the main point is to show that the fluctuations and the deviations in the convergence of the N-player game equilibria are mainly due to the fluctuations and the deviations in the convergence of the standard particle system (2.9). In a beautiful work, Lacker [83] extended the result by establishing convergence without assuming the existence of the master equation or any monotonicity property (but keeping the assumption that the idiosyncratic noise is nondegenerate): the limit points are weak MFG equilibria. The main difference with Theorem 2.8 is that [83] is based on a compactness argument (obtained by using the theory of relaxed controls, in which controls are regarded as being measure-valued) and provides no convergence rate. The result relies on the fact that, in some average sense, the deviation of a player barely affects the distribution of the players when N is large. Heuristically, this is due to the presence of the noise, which prevents the players to guess if another has deviated or not. However, in Lacker's approach, there might be a lot of (weak) MFG equilibria, apart from the monotone case where they are unique. This raises subtle questions of selection since only some of these equilibria may be selected when passing to limit: this is what happens in the examples discussed in Bayraktar and Zhang [14], Cecchin, Dai Pra, Fischer, and Pelino, [44], and Delarue and Foguen [52]. We provide more details in Section 4. Let us underline another limitation: The result presented above, as well as Lacker's approach, rely in a crucial way on the presence of a nondegenerate idiosyncratic noise and, to date, nothing is known outside this framework.

Finally, it is important to note that, historically, another approach was first implemented to relate the N-player and mean field games. In short, any solution to the mean field game gives rise to an approximate Nash equilibrium to the N-player game, with an accuracy that gets better and better as N increases. This idea dates back to the earliest papers in the field [73,75]. We refer to [36, CHAPTER 6] for a complete review.

3. THE LONG-TIME BEHAVIOR

In this section we discuss the behavior of MFG equilibria (without common noise) as the time horizon T tends to infinity. This is an interesting question both in terms of theory and applications: for instance, in economics, it is related to the existence of stationary equilibria or business cycles. On the other hand, the answer is not obvious because the MFG system has two boundary conditions, one at the initial time and one at the terminal time. One can therefore expect that convergence holds only far from the initial and terminal times. In order to perform this analysis, it is necessary to require that the solution of the stochastic control problem remains confined in an appropriate sense: the simplest setting in which this is possible is the spatially periodic one. We make this assumption here: we set $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ and denote by $\mathcal{P}(\mathbb{T}^d)$ the set of Borel probability measures on $\mathbb{T}^d$ endowed with the corresponding 2-Wasserstein distance. We consider the solution $(u^T, m^T) = (u_t^T, m_t^T)_{0\leq t\leq T}$ of the MFG system (1.4), now stated on $(0,T)\times\mathbb{T}^d$, in which $F, G : \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d) \to \mathbb{T}$ are "smooth."

3.1. The ergodic MFG system

As explained by Lions in [88], the limit of the MFG system (1.4), as the time horizon T tends to infinity, is expected to be given by the ergodic MFG system

$$\begin{cases} \bar\lambda - \Delta\bar u + \frac{1}{2}|D\bar u|^2 = F(x,\bar m) & \text{in } \mathbb{T}^d, \\ -\Delta\bar m - \operatorname{div}(\bar m\, D\bar u) = 0 & \text{in } \mathbb{T}^d, \\ \int_{\mathbb{T}^d}\bar m = 1, \quad \int_{\mathbb{T}^d}\bar u = 0. \end{cases} \tag{3.1}$$

Here the unknowns are $(\bar\lambda, \bar u, \bar m)$, where $\bar\lambda \in \mathbb{R}$ is the so-called ergodic constant. The interpretation of the system is the following: each player wants to minimize her ergodic cost

$$J(x,\alpha) := \inf_{\alpha}\limsup_{T\to+\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \left\{\frac{1}{2}|\alpha_t|^2 + F(X_t,\bar m)\right\}dt\right]$$

where $(X_t)_{t\geq 0}$ is the solution to

$$\begin{cases} dX_t = \alpha_t dt + \sqrt{2} dB_t, \\ X_0 = x. \end{cases}$$

The measure $\bar{m}$ in (3.1) is then understood as the invariant ergodic measure associated to the optimal trajectory (the existence of which is much easier to prove in the periodic setting). The solution to (3.1) is known to exist under fairly general assumptions on F and to be unique when the coupling function F is monotone (i.e., satisfies (1.5)); see [85, 87].

3.2. The convergence in the monotone setting

In this part we assume that F is smooth and monotone. Under this monotonicity assumption, one can show that the "long-time stability" takes the form of a *turnpike pattern*; namely, the solution (u^T, m^T) of (1.4) becomes nearly stationary for most of the time. The strongest way to formulate this type of behavior is the following exponential estimate:

Theorem 3.1. *There exist* $K, \omega > 0$ *such that for* (u^T, m^T) *and* $(\bar{u}, \bar{m})$ *solving respectively* (1.4) *and* (3.1)*,*

$$\left\| m^T(t) - \bar{m} \right\|_\infty + \left\| Du^T(t) - D\bar{u} \right\|_\infty \leq K(e^{-\omega t} + e^{-\omega(T-t)}), \quad \forall t \in (0, T). \tag{3.2}$$

The reader may notice that the initial condition $\tilde{m}_0$ for m^T and the terminal condition G for u^T are lost at the limit (as $(\bar{\lambda}, \bar{u}, \bar{m})$ does not depend on $\tilde{m}_0$ or G). This result was first stated in Cardaliaguet, Lasry, Lions, and Porretta [29] when the coupling F is monotone and local and in [30] when this coupling is monotone and regularizing. The proof is based in a crucial way on uniform (in t and T) semiconcavity estimates for u^T and on the energy identity established by Lasry and Lions [87]:

$$\begin{aligned} &\int_0^T \int_{\mathbb{T}^d} \frac{1}{2}(m_t^T + \bar{m})(x) \left| Du_t^T(x) - D\bar{u}(x) \right|^2 dt dx \\ &\quad = -\int_0^T \int_{\mathbb{T}^d} \big(F(x, m_t^T) - F(x, \bar{m}) \big)(m_t^T - \bar{m})(x) dt dx \\ &\qquad - \int_{\mathbb{T}^d} \big(G(x, m_T^T) - \bar{u}(x) \big)(m_T^T - \bar{m})(x) dx + \int_{\mathbb{T}^d} (u_0^T - \bar{u})(x)(m_0^T - \bar{m})(x) dx. \end{aligned}$$

This energy identity shows the role of the monotonicity property (1.5) in the analysis.

A consequence of the exponential estimate (3.2) is the existence of a constant C such that

$$\left| u^T(t, x) - \bar{u}(x) - \bar{\lambda}(T - t) \right| \leq C.$$

Following ideas of weak KAM theory (see, for instance, the ICM proceeding by Fathi [60] in the calculus of variation framework), one could expect the existence of a limit for $u^T(t, x) - \bar{\lambda}(T - t)$ as T tends to ∞; moreover, this limit should be given (up to an additive constant) by $\bar{u}$. However, this heuristic is not completely correct and the description of the asymptotic behavior of u^T (eventually established in the paper by Cardaliaguet and Porretta [33]) happens to be more subtle.

To overcome the difficulty that the MFG system is forward–backward, a possible path (towards a long-time expansion of u^T) is to use the master equation (2.3), which is just backward in time. One of the main results of [33] states that the solution of the master equation converges to the solution of the following ergodic master equation:

$$\lambda - \Delta_x \chi(x,m) + \frac{1}{2}\left|D_x \chi(x,m)\right|^2 - \int_{\mathbb{T}^d} \mathrm{div}_y (D_m \chi(x,m,y)) dm(y) \\ + \int_{\mathbb{T}^d} D_m \chi(x,m,y) \cdot D_x \chi(y,m) dm(y) = F(x,m) \quad \text{in } \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d). \tag{3.3}$$

Concerning the existence of (3.3), the following result holds:

Theorem 3.2. *There is a unique constant $\lambda \in \mathbb{R}$ for which the master cell problem* (3.3) *has a (weak) solution. The constant λ coincides with the unique constant $\bar{\lambda}$ for which the ergodic MFG problem* (3.1) *has a solution. Besides, if χ is a solution to* (3.3)*, then $\chi(\cdot, m)$ is of class C^2 (in space) for any $m \in \mathcal{P}(\mathbb{T}^d)$ and*

$$D_x \chi(x, \bar{m}) = D\bar{u}(x) \quad \forall x \in \mathbb{T}^d,$$

where $(\bar{u}, \bar{m})$ is the solution to (3.1).

As in many constructions of a solution to an ergodic problem, the first step consists in building solutions to approximating compact problems and then in proving uniform estimates on these solutions. Here, the compact problems are discounted master equations which can be solved by a method of (infinite-dimensional) characteristics (as for (2.3)). The main issue is to prove estimates on these solutions, independently of the discount rate. In contrast with standard constructions in this area (see Lions–Papanicolau–Varadhan [90] or [60], which analyze the ergodic behavior of (pure) Hamilton–Jacobi equations with a coercive Hamiltonian), the proof of these estimates cannot rely on the coercivity properties of the equation, but must use in a very strong way the bound (3.2), which describes the long-time behavior of the characteristics.

We are now ready to discuss the convergence, as $t \to -\infty$, of the solution U of the master equation (2.3) (now defined in the time interval $(-\infty, 0]$ with terminal condition $U(0,x,m) = G(x,m)$).

Theorem 3.3. *Let χ be a weak solution to the master cell problem* (3.3)*. Then, there exists a constant $c \in \mathbb{R}$ such that*

$$\lim_{t \to -\infty} U(t,x,m) + \bar{\lambda} t = \chi(x,m) + c,$$

uniformly with respect to $(x,m) \in \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d)$.

Moreover, we also have that $D_x U(t,x,m) \to D_x \chi(x,m)$ as $t \to -\infty$, uniformly with respect to (x,m).

This result looks like an extension of the famous Fathi's result on the convergence of the Lax–Oleinik semigroup in weak-KAM theory [60]. This parallel is not completely correct since the master equation is not a Hamilton–Jacobi equation in an infinite-dimensional setting: the comparison principle does not hold, for instance. One has to rely instead on the energy identity described above.

From Theorem 3.3 one can derive the full convergence of the solution (u^T, m^T) of the MFG system:

Corollary 3.4. *Let c be the constant given in Theorem 3.3. For $T > 0$ and $\tilde{m}_0 \in \mathcal{P}(\mathbb{T}^d)$, let (u^T, m^T) be the solution to (1.4). Then, for any $t \geq 0$,*

$$\lim_{T\to+\infty} \big(u^T(t,x) - \bar{\lambda}(T-t)\big) = \chi\big(x, m(t)\big) + c,$$

where the convergence is uniform in x and m solves

$$\partial_t m - \Delta m - \operatorname{div}\big(m D_x \chi(x,m)\big) = 0, \quad m(0) = \tilde{m}_0.$$

In the recent paper [47], Cirant and Porretta managed to show the above corollary without relying on the master equation.

Among many open problems in this area, let us point out the following ones: we have explained in Section 2 that the master equation can be obtained as the limit of the Nash system (1.1). Now that we understand the behavior of the master equation on long time intervals, it would be interesting to see if this convergence holds uniformly in time. Similar results have been obtained, for instance, by Mischler and Mouhot [93] in the framework of kinetic theory. Another very intriguing issue is the long-time behavior of the MFG system in the presence of a common noise: the existence of stationary measures is a completely open problem.

3.3. The long-time behavior without monotonicity

The long-time behavior of the MFG equilibria when the coupling is not monotone is poorly understood and only partial results are known.

The potential case. When the MFG is potential (see (2.2)), then one can extend weak-KAM theory to the infinite-dimensional setup and describe the possible ω-limit sets of the solution of the time-dependent MFG system minimizing a natural energy in terms of a "Mather set." The main point is that this set may not contain an ergodic MFG equilibrium (i.e., an $\bar{m} \in \mathcal{P}(\mathbb{T}^d)$ for which there exists $(\bar{\lambda}, \bar{u})$ such that $(\bar{\lambda}, \bar{u}, \bar{m})$ solves (3.1)): this shows that the ω-limit set of the solutions of the time-dependent MFG system (1.4) that additionally minimize the natural energy may not contain an MFG ergodic equilibrium. In other words, the ergodic MFG system (3.1) may not describe the long-time behavior of these trajectories.

Periodic solutions. The existence of a periodic solution to the MFG system is a fascinating topic on which little is known. The main result in that direction is the analysis by Cirant [46] of a class of examples. It relies on local and global bifurcation methods based on the analysis of eigenfunction expansions of solutions to a suitable linearized problem. Note, however, that the stability of these solutions is not known.

Traveling waves. Intimately related to the notion of equilibria and to periodic solutions, the question of traveling waves has been discussed in the framework of an MFG problem of knowledge growth, first introduced in economics by Lucas and Moll [91]. In this setting the construction of a traveling wave solution is crucial (it is called a balanced-growth path solution in economics) and has been documented by Papanicolaou, Ryzhik and Velcheva [95] and by Porretta and Rossi [98]. The convergence of the solution of the time-dependent problem to this solution remains an open problem.

4. SMOOTHING EFFECT OF THE COMMON NOISE

A natural question is to address the impact of the common noise on the well-posedness of a mean field game. It is indeed useful to observe that, most often, standard mean field games (without common noise) have multiple solutions. In this respect, condition (1.5) is rather restrictive. Just as an additive Brownian motion can restore uniqueness of differential equations driven by nonsmooth vector fields, we can then wonder whether a form of common noise could force equilibria to be unique in a fairly large class of mean field games.

4.1. The linear–quadratic case as a warm-up

It is pretty clear that the form of common noise that is inserted into (1.8) is certainly not sufficient to reach such an aim in full generality. Indeed, the noise is just finite-dimensional whereas the model is infinite-dimensional because of the mean field structure. For sure, we could think of some hypoelliptic structure that could allow the finite-dimensional noise to be transmitted to all the components of the space of probability measures, but this looks a very challenging question. A much easier (but much less ambitious) alternative is to restrict oneself to mean field games whose equilibria are *a priori* known to live in a finite-dimensional subset or, using a standard concept from statistics, to belong to a parametric model of statistical distributions. The typical example in this direction is the class of linear–quadratic mean field games, which has been studied with a lot of attention (see Bardi [10], Bensoussan, Sung, Yam and Yung [16], Carmona, Delarue, and Lachapelle [37], and the works [73, 75] by Caines, Huang, and Malhamé for a tiny example). In short, it corresponds to the case when F and G in (1.8) have the form

$$F(x,m) = \frac{1}{2}\big|Qx + f(\overline{m})\big|^2, \quad G(x,m) = \frac{1}{2}\big|Rx + g(\overline{m})\big|^2, \tag{4.1}$$

where Q and R are matrices of size $d \times e$ (with e being another integer), f, g are Borel functions from $\mathbb{R}^d$ to $\mathbb{R}^e$ and $\overline{m}$ is the mean of m, i.e., $\overline{m} = \int_{\mathbb{R}^d} x dm(x)$ (which implicitly requires m to have a finite first moment). Referring back to Section 1.2, we see that the control problem (1.2)–(1.3) ((1.6)–(1.7) in the presence of common noise) becomes a stochastic control problem with linear–quadratic coefficients depending on the (possibly random) path $(m_t)_t \in C^0([0,T], \mathcal{P}_2(\mathbb{R}^d))$ injected into the coefficients. The key point is that this stochastic control problem has a unique solution (depending on $(m_t)_t$), with the optimal feedback being affine (regardless of the value of the intensity of the noises). In turn, this implies that the equilibrium trajectories must be Gaussian processes (conditional on the initial condition whenever the latter is random). Therefore, for the above choice of F and G, the equilibria are necessarily Gaussian (once again, conditional on the initial condition). Even more, since the volatility coefficient is prescribed in the state dynamics, the variances of the marginal conditional laws of the equilibria given the initial condition are also fixed. In the end, only the means count for determining the equilibria: As expected, the model is parametric. It is then an interesting question to address the impact of the common noise in this specific framework and to see whether the existing well-posedness results can be improved under the action of $(W_t)_t$. A very convenient approach is to use the Pontryagin principle, which provides, under the standing x-convex structure of F and G, a characterization of the equilibria in the form of a forward–backward system of the McKean–Vlasov type. Standard computations (see the aforementioned references together with [35, CHAPTER 3]) then show that, for a given $(W_t)_t$-adapted path $(m_t)_t$ with values in $C^0([0,T],\mathcal{P}_2(\mathbb{R}^d))$, the optimal control in the stochastic control problem described in (1.6)–(1.7) has the following feedback form:

$$\alpha_t^* = -(\eta_t X_t^* + h_t), \tag{4.2}$$

where $(\eta_t)_t$ is the solution of an autonomous deterministic Riccati equation (the form of which is completely independent of the input $(m_t)_t$) and $(h_t)_t$ solves the finite-dimensional backward stochastic differential equation

$$h_t = R^\dagger g(\overline{m}_T) + \int_t^T \big[Q^\dagger f(\overline{m}_s) - \eta_s h_s\big] ds - \int_t^T k_s dW_s, \quad t \in [0,T]. \tag{4.3}$$

Obviously, this equation should be regarded as a finite-dimensional version of the backward equation in (1.8) when the value function therein is sought in a quadratic form.

Forcing uniqueness. Inserting the relationship (4.2) for the optimal feedback into the dynamics (1.7), taking the conditional mean of $(X_t^*)_t$ (with the exponent $*$ being used to denote the optimal trajectory) given the common noise $(W_t)_t$, and then identifying $\mathbb{E}[X_t^*|(W_s)_s]$ with $\overline{m}_t$ (in full consistency with the probabilistic fixed-point formulation of a mean field game), we end up with the following forward–backward system (which is now the finite-dimensional analogue of the whole system (1.8)):

$$\begin{cases} d\overline{m}_t = -(\eta_t \overline{m}_t + h_t)dt + \sqrt{2\epsilon}\, dW_t, & m_0 = \mathbb{E}(X_0), \\ dh_t = -\big(Q^\dagger f(\overline{m}_t) - \eta_t h_t\big)dt + k_t dW_t, & h_T = R^\dagger g(\overline{m}_T). \end{cases} \tag{4.4}$$

Similar to $(v_t)_t$ in (1.8), the $(W_t)_t$-adapted process $(k_t)_t$ is here designed to render the solution $(h_t)_t$ $(W_t)_t$-adapted. Remarkably, system (4.4) just involves the conditional expectation $(\overline{m}_t)_t$. This is in line with the fact that equilibria are known to belong to a parametric model. It then remains to interpret the forward–backward system (4.4) as the system of characteristics of a parabolic PDE. We obtain

$$h_t = \theta_t(\overline{m}_t),$$

where θ solves

$$\partial_t \theta_t(x) + \epsilon \Delta_x^2 \theta_t(x) - \big(\eta_t x + \theta_t(x)\big) \cdot \nabla_x \theta_t(x) + Q^\dagger f(x) - \eta_t \theta_t(x) = 0, \tag{4.5}$$

for $(t,x) \in (0,T) \times \mathbb{R}^d$, with the terminal condition $\theta_T(x) = g(x)$. This PDE is a finite-dimensional version of the master equation (2.3). Obviously, it is much easier to solve.

In particular, when $\epsilon > 0$, the sole presence of the Laplacian forces the existence of a classical solution when f and g are bounded and regular coefficients. In turn, this forces the well-posedness of the system of characteristics (4.4) (see Foguen [102] or [36, CHAPTER 3]):

Proposition 4.1. *Let the cost coefficients F and G be of the same form as in (4.1), with f and g therein being bounded and sufficiently regular coefficients. Then, for any $\epsilon > 0$, the mean field game has a unique solution.*

It must be stressed that the statement becomes false when $\epsilon = 0$ (see the same references for explicit examples). One must then assume more about the coefficients f and g to force uniqueness. For instance, it is easy to reformulate the monotonicity condition (1.5) in terms of f and g: The point is then to require $Q^\dagger f$ and $R^\dagger g$ to satisfy $(Q^\dagger f(x') - Q^\dagger f(x)) \cdot (x' - x) \geq 0$ for any $x, x' \in \mathbb{R}^d$, and similarly for $R^\dagger g$. Regarding the explicit conditions of regularity that f and g must satisfy in Proposition 4.1, a typical instance is to assume that f is bounded and Hölder continuous on the whole space and g, together with its first and second-order derivatives, are bounded and Hölder continuous on the whole space.

4.2. Finite-state mean field games

Another obvious manner to get a parametric model is to force the state space to be finite, in which case the space of probability measures itself becomes finite-dimensional. This requires, however, a modicum of care since the state dynamics can no longer be formulated as in (1.3)–(1.7). In particular, the common noise cannot be chosen in a mere additive fashion.

Games without common noise. When the state space is finite (and is thus chosen as a finite set E), the dynamics of the reference player are usually postulated in the form of a Markov controlled process taking values in E. Typically, the transition rates are explicitly prescribed as functions of the control (see Gomes, Mohr, and Souza [66,67] and Guéant [70]). A simple, but convenient, choice is then to identify the control with the entire transition matrix. In that case, using the same notation $(X_t)_t$ as in (1.3) to denote the trajectory of the reference player, the transition probabilities read (with $\mathbb{P}$ being implicitly identified with $\mathbb{P}^1$ since there is no common noise at this stage of the discussion)

$$\begin{aligned} &\mathbb{P}(X_{t+dt} = j | X_t = i) = \beta_t^{i,j} dt + o(dt), \quad i \neq j, \\ &\mathbb{P}(X_{t+dt} = i | X_t = i) = 1 + \beta_t^{i,i} dt + o(dt), \end{aligned} \tag{4.6}$$

with $((\beta_t^{i,j})_{i,j \in E})_t$ standing for a deterministic path with values in the set of E-indexed matrices satisfying the following two standard prescriptions:

$$\begin{cases} \beta_t^{i,j} \geq 0, \quad i \neq j, \\ \beta_t^{i,i} = -\sum_{j \neq i} \beta_t^{i,j}. \end{cases} \tag{4.7}$$

This formulation is reminiscent of (1.3) in the sense that the transitions do not depend on the choice of the environment $(m_t)_t$ that underpins the cost functional (1.2). In particular, the

Fokker–Planck equation for the marginal law of the $(\beta_t)_t$-controlled process $(X_t)_t$ can be written as

$$\frac{d}{dt} p_t^i = \sum_{j \in E} p_t^j \beta_t^{j,i}, \quad t \in [0,T], \quad i \in E, \tag{4.8}$$

with p_t^i being understood as $\mathbb{P}(X_t = i)$. As for the cost functional, we may choose it as in (1.2) provided that the functions F and G are now defined on $E \times \mathcal{P}(E)$, with $\mathcal{P}(E)$ denoting the space of probability measures (which can be obviously identified with the simplex of dimension $|E| - 1$). To give a clear account that the state space is finite, we will write (in this subsection) $F^x(m)$ and $G^x(m)$ instead of $F(x,m)$ and $G(x,m)$. Of course, there is another slight difference with (1.2), which lies in the interpretation of $(\beta_t)_t$. In (1.2), β_t is implicitly chosen as a control in feedback form: Loosely speaking, we write $\tilde{\beta}_t(X_t)$ for a d-dimensional vector field $\tilde{\beta}_t$; In other words, the quadratic cost in (1.2) is calculated from the pointwise value of the feedback function at X_t. Differently, β_t in (4.7) encodes the entire feedback function: Somehow, it coincides with the entire function $\tilde{\beta}_t$. In this framework, the cost functional (1.2) should read

$$\begin{aligned}
&\mathbb{E}\biggl[\int_0^T \biggl(\frac{1}{2}\sum_{j \neq X_t} |\beta_t^{X_t^\beta, j}|^2 + F^{X_t^\beta}(m_t)\biggr) dt + G^{X_T^\beta}(m_T)\biggr] \\
&= \sum_{i \in E}\biggl[\int_0^T p_t^i \biggl(\frac{1}{2}\sum_{j \neq i} |\beta_t^{i,j}|^2 + F^i(m_t)\biggr) dt + p_T^i G^i(m_T)\biggr] \\
&:= J\bigl((\beta_t)_t; (p_t)_t; (m_t)_t\bigr),
\end{aligned} \tag{4.9}$$

for a given continuous (and here deterministic) path $(m_t)_t$ with values in $\mathcal{P}(E)$.

It it then quite standard to compute the corresponding HJB equation. Since E is finite, it becomes a mere ordinary differential equation. Accordingly, the MFG system (1.4) becomes

$$\begin{cases}
-\partial_t u_t^i + \dfrac{1}{2}\displaystyle\sum_{j \in E} (u_t^i - u_t^j)_+^2 = F^i(m_t), \\
\partial_t m_t^i - \displaystyle\sum_{j \in E} \bigl[m_t^j (u_t^j - u_t^i)_+ - m_t^i (u_t^i - u_t^j)_+\bigr] = 0, \quad i \in E,\ t \in [0,T].
\end{cases} \tag{4.10}$$

Once the system (4.10) is solved, the optimal feedback is given by $\alpha_t^{i,j} = (u_t^i - u_t^j)_+, i \neq j$. Consistently with the notation introduced in (4.8), the probability measure m_t is identified with the collection of nonnegative weights $(m_t^i)_{i \in E}$, with the latter satisfying $\sum_{j \in E} m_t^j = 1$.

Adding a common noise. Differently from (1.4), (4.10) is a finite-dimensional forward–backward system. The question is then how to find a suitable form of finite-dimensional common noise that forces existence and uniqueness. Although it is very similar to the question addressed in Section 4.1 for linear–quadratic quadratic mean field games, the problem is in fact formulated in a different way. Indeed, the analysis carried out in Section 4.1 mostly relies on the probabilistic formulation of the mean field game or, equivalently, on the equation for the dynamics of the reference player. Instead, we want to use here the equation for the dynamics of the population, as it is more adapted to the model in hand. This raises some subtle issues on the structure of the common noise as we want the resulting Fokker–Planck equation to preserve the simplex. In other words, we want to find a form of simplex-valued diffusion process. A very famous instance is the so-called Wright–Fisher process, originally introduced in stochastic models for population genetics (see Kimura [77]). Recast in our framework (the analysis of which is taken from Bayraktar, Cecchin, Cohen, and Delarue [13]), it leads to the following stochastic version of the MFG system (4.10):

$$\begin{cases} d_t u_t^i = \Big(\frac{1}{2}\sum_{j\in E}(u_t^i - u_t^j)_+^2 - F^i(m_t) - \sqrt{\epsilon}\sum_{j\in E}\sqrt{m_t^i m_t^j}\,(v_t^{i,i,j} - v_t^{i,j,i})\Big)dt \\ \qquad\qquad + \sum_{j,k\in E} v_t^{i,j,k}\,dW_t^{j,k}, \\ d_t m_t^i = \sum_{j\in E}\big[m_t^j(u_t^j - u_t^i)_+ - m_t^i(u_t^i - u_t^j)_+\big]dt + \sqrt{\epsilon}\sum_{j\in E}\sqrt{m_t^i m_t^j}\,d\big[W_t^{i,j} - W_t^{j,i}\big], \end{cases} \tag{4.11}$$

for $i \in E$ and $t \in [0,T]$. In the above, $(W_t)_t = ((W_t^{i,j})_{0\le t\le T})_{(i,j)\in E^2}$ is a collection of independent Brownian motions. Following the notations introduced in the statement of Theorem 2.3, it is very useful to distinguish the space carrying $(W_t)_t$ from the space carrying the idiosyncratic noise underpinning the transition rates (4.6): The former will be denoted by $(\Omega_0, \mathbb{F}_0, \mathbb{P}_0)$ and the latter by $(\Omega_1, \mathbb{F}_1, \mathbb{P}_1)$. Accordingly, the expectations are respectively denoted by $\mathbb{E}_0$ and $\mathbb{E}_1$. The product measure on the product space is denoted by $\mathbb{P}$ and the corresponding expectation by $\mathbb{E}$. Intuitively, the process $(v_t)_t$ in the above backward equation plays the same role as the process $(v_t)_t$ in (1.8). In particular, it is worth observing that, in both cases, the process $(v_t)_t$ appears in the dt term of the backward equation.

Before we provide the interpretation of the above system in terms of a mean field game, we write down the resulting form of the master equation (see again [13]):

$$\begin{aligned} &\partial_t U_t^i(m) + \epsilon\sum_{j,k\in E}(m_j\delta_{j,k} - m_j m_k)\partial^2_{m_j m_k} U_t^i(m) \\ &\quad + \sum_{j,k\in E} p_k\big(U_t^k(m) - U_t^j(m)\big)_+\big(\partial_{m_j}U_t^i(m) - \partial_{m_k}U_t^i(m)\big) \\ &\quad + 2\epsilon\sum_{j\in E} p_j\big(\partial_{m_i}U_t^i(m) - \partial_{m_j}U_t^i(m)\big) - \frac{1}{2}\sum_{j\in E}\big(U_t^i(m) - U_t^j(m)\big)_+^2 \\ &\quad + F^i(m) = 0, \end{aligned} \tag{4.12}$$

with the boundary condition $U_T^i(m) = g^i(m)$. The terms induced by the common noise are those featuring the prefactor ϵ. In particular, the master equation without common noise is obtained by letting $\epsilon = 0$. The main impact of the common noise is to generate the second-order differential operator

$$\epsilon\sum_{j,k\in E}(m_j\delta_{j,k} - m_j m_k)\partial^2_{m_j m_k}, \tag{4.13}$$

which is called a (purely second-order) Kimura operator on the simplex of dimension $|E| - 1$. In both (4.12) and (4.13), the derivatives should be formally regarded as intrinsic derivatives on the simplex, with gradients being of dimension $|E| - 1$. However, it is also possible to assume that U has a smooth extension to an $|E|$-dimensional subset of the simplex and then to consider the derivatives as standard $|E|$-dimensional derivatives. It is worth noticing that the resulting derivative used in (4.12) and (4.13) does not coincide with the derivative D_m introduced in Theorem 2.1. The infinite-dimensional analogue of the derivative used in (4.12) and (4.13) is the so-called flat, or linear, functional derivative. In short, it is the restriction, to the space of probability measures, of the derivative on the space of signed measures. It is a potential of the derivative D_m.

Forcing uniqueness. A key feature of the Kimura operator (4.13) lies in the structure of the diffusion matrix: it degenerates near the boundary of the simplex. This is somehow the price to pay to construct a diffusion process that does not leave the simplex. As an issue, it makes much more difficult any attempt to prove smoothing properties (which are precisely what we need in order to force uniqueness to the system (4.11), in full analogy with the result stated in Proposition 4.1). However, a relevant form of Schauder's theory was established in the monograph by Epstein and Mazzeo [59]. In short, it says that linear equations driven by Kimura operators have classical solutions (with a suitable behavior at the boundary) if the first-order and source terms are just Hölder continuous in time and space. This is, however, not sufficient to get similar results for the nonlinear equation (4.12), as the first-order term therein is driven by the solution itself. As U is easily shown to be bounded from a straightforward application of the maximum principle, the next step to fill the gap is thus to prove the following form of *a priori* estimate: For some Hölder exponent, the Hölder norm of a classical solution to a homogeneous parabolic equation driven by a Kimura operator is bounded in terms of the L^∞-norm of the solution and the Hölder norm of the initial condition (if the equation is set forward) or of the terminal condition (if the equation is set backward as (4.12) is). Nevertheless, it is not possible to prove this in full generality. In short, the best results that are known require the presence in (4.13) of a first-order term with strictly positive components along inward normal directions to the boundary. When applying this principle to (4.12), we are led to consider the following modified version of the master equation:

$$\begin{aligned}
&\partial_t U_t^i(m) + \epsilon \sum_{j,k\in E} (m_j \delta_{j,k} - m_j m_k) \partial^2_{m_j m_k} U_t^i(m) \\
&\quad + \sum_{j,k\in E} p_k \big[\varphi(m_j) + \big(U_t^k(m) - U_t^j(m)\big)_+\big]\big(\partial_{m_j} U_t^i(m) - \partial_{m_k} U_t^i(m)\big) \\
&\quad + 2\epsilon \sum_{j\in E} m_j \big(\partial_{m_i} U_t^i(m) - \partial_{m_j} U_t^i(m)\big) - \frac{1}{2} \sum_{j\in E} \big(U_t^i(m) - U_t^j(m)\big)_+^2 \\
&\quad + F^i(m) + \sum_{j\in E} \varphi(m_j)\big(U_t^j(m) - U_t^i(m)\big) = 0, \qquad (4.14)
\end{aligned}$$

with the terminal condition $U_T^i(m) = g^i(m)$, for a smooth function φ from $[0,\infty)$ into itself that is nonzero in the neighborhood of 0. This function φ should be regarded as a penalty: when inserted in the transition rates (4.6), it forces the corresponding solution to the Fokker–Planck equation (4.8) to leave the boundary of the simplex (here and below, the notions of boundary and interior of the simplex are understood when $\mathcal{P}(E)$ is regarded as a subset of

$\mathbb{R}^{|E|-1}$). Notice that this additional penalty φ appears in the first-order term on the second line, which is consistent with our preliminary discussion, but also in the zeroth-order term on the last line, which is necessary to have a relevant interpretation of (4.14) as the master equation of a mean field game (see Definition 4.1 below).

The next statement is also taken from [13]:

Theorem 4.2. *We can find a threshold $\kappa_0 > 0$, only depending on ϵ ($\epsilon > 0$), $\|F\|_\infty$, $\|G\|_\infty$, and T, such that, if $\varphi(0) > \kappa_0$, and if F and G are smooth enough, then equation* (4.14) *has a classical solution, with first-order derivatives in space that are bounded on the whole domain and second-order derivatives in space that are bounded on $[0,T] \times \mathcal{K}$, for any compact subset $\mathcal{K}$ included in the interior of $\mathcal{P}(E)$.*

The existence of a classical solution is then shown to force uniqueness to the corresponding system of characteristics. Due to the presence of the penalty φ, this system does not exactly fit (4.11). The right-version is

$$\begin{cases} d_t u_t^i = \Big(\dfrac{1}{2} \displaystyle\sum_{j \in E} (u_t^i - u_t^j)_+^2 - F^i(m_t) - \sqrt{\epsilon} \sum_{j \in E} \sqrt{m_t^i m_t^j} (v_t^{i,i,j} - v_t^{i,j,i}) \Big) dt \\ \qquad\quad - \displaystyle\sum_{j \in E} \varphi(m_t^j)(u_t^j - u_t^i) + \sum_{j,k \in E} v_t^{i,j,k} d W_t^{j,k}, \\ d_t m_t^i = \displaystyle\sum_{j \in E} \big[m_t^j \big(\varphi(m_t^i) + (u_t^j - u_t^i)_+\big) - m_t^i \big(\varphi(m_t^j) + (u_t^i - u_t^j)_+\big)\big] dt \\ \qquad\quad + \sqrt{\epsilon} \displaystyle\sum_{j \in E} \sqrt{m_t^i m_t^j} d\big[W_t^{i,j} - W_t^{j,i}\big], \end{cases} \tag{4.15}$$

for $i \in E$ and $t \in [0,T]$. In line with Theorem 4.2, we have (see again [13]):

Theorem 4.3. *We can find a threshold $\kappa_0 > 0$, only depending on ϵ ($\epsilon > 0$), $\|F\|_\infty$, $\|G\|_\infty$, and T, such that, if $\varphi(0) > \kappa_0$, and if F and G are smooth enough, then the forward–backward system* (4.15) *has a unique solution when the initial condition $m_0 = (m_0^i)_{i \in E}$ is prescribed in the interior of the simplex.*

To be fair, we should mention that uniqueness holds within a class of solutions with suitable integrability properties. We refer to [13] for the complete version of the statement. As for the constraint on the initial condition, it says that $m_0^i > 0$ for any $i \in E$. The resulting solution $(m_t)_t$ is then shown to stay away from the boundary (which is helpful since the diffusion coefficient in the dynamics of $(m_t)_t$ becomes singular on the boundary). Implicitly, all the statements below are also limited to initial conditions in the interior of the simplex.

It now remains to provide an interpretation of the two systems (4.11) and (4.15) in terms of a mean field game. This goes through the following definition:

Definition 4.1. We say that a $(W_t)_t$-adapted continuous stochastic process $(m_t)_{0 \le t \le T}$ with values in the interior of $\mathcal{P}(E)$ is a solution to the mean field game with common noise of intensity $\sqrt{\epsilon}$ (and without the penalization φ) if $(m_t)_{0 \le t \le T}$ satisfies an equation of the form

$$dm_t^i = \sum_{j \in E} m_t^j \alpha_t^{j,i} dt + \sqrt{\epsilon} \sum_{j \in E} \sqrt{m_t^i m_t^j}\, d(W_t^{i,j} - W_t^{j,i}), \quad t \in [0,T], \tag{4.16}$$

for a bounded $(W_t)_t$-progressively-measurable process $((\alpha_t^{i,j})_{i,j \in E})_t$ satisfying (4.7) and, for any other bounded $(W_t)_t$-progressively-measurable process $((\beta_t^{i,j})_{i,j \in E})_t$ satisfying (4.7), the solution of the equation

$$dp_t^i = \sum_{j \in E} p_t^j \beta_t^{j,i} dt + \sqrt{\epsilon} p_t^i \sum_{j \in E} \sqrt{\frac{m_t^j}{m_t^i}}\, d(W_t^{i,j} - W_t^{j,i}), \quad t \in [0,T], \tag{4.17}$$

satisfies the inequality

$$\mathbb{E}^0\big[J\big((\beta_t)_t; (p_t)_t; (m_t)_t\big)\big] \ge \mathbb{E}^0\big[J\big((\alpha_t)_t; (m_t)_t; (m_t)_t\big)\big],$$

with J being defined as in (4.9).

A similar definition holds for the mean field game with common noise of intensity $\sqrt{\epsilon}$ in the presence of the penalization φ. It suffices to replace $(\alpha_t^{j,i})_t$ by $(\varphi(m_t^i) + \alpha_t^{j,i})_t$ in (4.16) and $(\beta_t^{j,i})_t$ by $(\varphi(p_t^i) + \beta_t^{j,i})_t$ in (4.17).

In fact, Definition 4.1 is rather subtle. Differently from the formulation (1.6)–(1.7) used for continuous state spaces, the current one does not provide an explicit formulation of the (private) dynamics of the reference player within the population. In short, Definition 4.1 is missing an equation similar to (1.7). Instead, equation (4.17) should be regarded as a form of Fokker–Planck equation for some marginal statistics of the reference player given the common noise. Actually, it can be proven that there exists a stochastic process $(X_t, Y_t)_{0 \le t \le T}$ with values in the space $E \times \mathbb{R}_+$ such that

$$p_t^i = \mathbb{E}^1[Y_t \mathbf{1}_{\{X_t = i\}}], \quad t \in [0,T], \quad i \in E,$$

with $(Y_t)_t$ satisfying $\mathbb{E}^0\mathbb{E}^1[Y_t] = 1$. In this formulation, X_t should be regarded as the physical state, at time t, of the reference player, with the latter being also assigned a mass Y_t. The mass of the tagged particle is in fact a density on the entire probability space carrying both types of noise. It is a density accounting for the way the reference player perceives the world. In this respect, it is important to note that the process $(p_t)_t$ does not take values in the simplex, but only in the orthant $(\mathbb{R}_+)^{|E|}$. This follows from the linear structure of equation (4.17) (with $(p_t)_t$ as unknown). The linear structure, here with stochastic coefficients, is consistent with the linear structure of the Fokker–Planck equation (4.8). In order to obtain solutions in a relevant space, integrability conditions on these stochastic coefficients are thus necessary, whence the assumption that $((\alpha_t^{i,j})_{i,j \in E})_t$ and $((\beta_t^{i,j})_{i,j \in E})_t$ are bounded.

4.3. Vanishing viscosity

Following the latter two subsections, a natural question is to address the vanishing viscosity limits of the solutions to the mean field game with common noise and to the corresponding parabolic master equation. Both for linear quadratic mean field games and finite state mean field games, uniqueness of the equilibria may be lost in the framework of Proposition 4.1 and Theorem 4.3 when the common noise is removed. This is the same for the corresponding master equation: Classical solutions may cease to exist and, accordingly, weaker notions of solutions are needed; Uniqueness is then a challenging question.

For sure, we could think of other methods for selecting equilibria. For instance, we could think of returning back to the game with N players and then identifying which equilibria coincide with a limit point of the N-equilibrium as N tends to infinity. This is, however, a very difficult road. As easily seen from the uniformly parabolic structure of the system (1.1), the N-player game (at least in the form studied there) satisfies a form of non-degeneracy that is asymptotically lost when N tends to infinity. The study of the large-N limit thus combines two difficulties at the same time: The whole system becomes more and more degenerate (this is a vanishing viscosity limit) and, meanwhile, some propagation of chaos is expected to occur (this is the mean field limit). In contrast, taking the small noise limit in a mean field game with common noise just raises one of these two issues since the mean field limit has already been taken.

Earlier selection results can be found in Bayraktar and Zhang [14], Cecchin, Dai Pra, Fischer, and Pelino, [44] and Delarue and Foguen [52]. Generally speaking, they are stated for mean field games whose equilibria are known to belong to a one-dimensional parametric model. This covers the following two examples: Linear–quadratic mean field games of the same type as in Section 4.1, but with d therein being equal to 1 (which implies in particular that, conditional on the initial state, the equilibria follow Gaussian distributions with a known variance but an unknown mean); Finite state mean field games on a set E containing two elements only (in which case the simplex is one-dimensional). In all these aforementioned works, selection is directly proved by taking the large-N limit in the finite game. Basically, this is possible thanks to the totally ordered structure of $\mathbb{R}$. Moreover, the master equation then reduces to a scalar conservation law and the selected solution is the entropy solution.

When the effective dimension of the model is greater than or equal to 2, things become much more challenging. A way to make the problem simpler is to address the so-called potential case. As explained in Proposition 2.4, potential games are a special kind of mean field games that coincide with the first-order condition of a mean field control problem. When the state space E is finite, this corresponds to the case where F and G satisfy

$$F^i(x,m) = \partial_{m_i}\mathcal{F}(m), \quad G^i(x,\mu) = \partial_{m_i}\mathcal{G}(m), \quad i \in E, \tag{4.18}$$

for two real-valued functions $\mathcal{F}$ and $\mathcal{G}$ defined on $\mathcal{P}(E)$. In words, F and G are identified with (respectively) the gradient of $\mathcal{F}$ and the gradient of $\mathcal{G}$. The identification is, however, a bit subtle since, formally, these two gradients should be identified with vectors of dimension $|E| - 1$. In turn, this says that the above condition could be slightly relaxed: In short, it would suffice to identify the projections of F and G onto the orthogonal complement of $(1, \dots, 1)$ (which should be regarded as the tangent space to the simplex) with the corresponding intrinsic gradient. Anyway, given $\mathcal{F}$ and $\mathcal{G}$, we can consider the deterministic optimal control problem

$$\inf_{(\beta_t)_t} \mathcal{J}\big((\beta_t)_{0\le t\le T}\big),$$

associated with the cost functional

$$\mathcal{J}\big((\beta_t)_{0\le t\le T}\big) = \int_0^T \bigg(\frac{1}{2}\sum_{i,j\in E:i\neq j} p_t^i\big|\beta_t^{i,j}\big|^2 + \mathcal{F}(p_t)\bigg)dt + \mathcal{G}(p_T), \tag{4.19}$$

and with the dynamics (4.8) (for a given initial condition $(p_0^i)_{i\in E}$), the function $(\beta_t)_t$ satisfying the constraint (4.7) at any time. Then, very similar to Proposition 2.4, we have

Proposition 4.4. *Let $m_0 = (m_0^i)_{i\in E}$ be an initial condition in the interior of the simplex. Under condition* (4.18)*, any bounded minimizer $((\beta_t^{i,j})_{i,j\in E})_{0\le t\le T}$ of the cost functional* (4.19)*, with $(p_0^i)_{i\in E} = (m_0^i)_{i\in E}$ as initial condition in* (4.8)*, yields a solution to the mean field game* (4.9)*.*

The proof follows from a standard application of the Pontryagin principle. The adjoint variable then identifies with $(u_t^i)_{i\in E}$ in the system (4.10). In the statement, the two constraints on $(p_0^i)_{i\in E}$ (which is required to have strictly positive coordinates) and on $((\beta_t^{i,j})_{i,j\in E})_{0\le t\le T}$ (which is required to be bounded) force the corresponding trajectory (4.8) to stay away from the boundary of the simplex (when the latter is viewed as an open subset of dimension of $|E|-1$). This guarantees that, along the trajectory (4.8), the extended Hamiltonian has a unique minimizer, as required in the application of the Pontryagin principle.

Selection of equilibria. Obviously, there is no converse to Proposition 4.4: The set of equilibria of a potential mean field game may be strictly larger than the set of minimizers to the corresponding mean field control. In this respect, a natural selection principle would consist in ruling out the equilibria that are not minimizers of the corresponding mean field control. Very interestingly, this principle is consistent with the results mentioned above when the state space E is of cardinality 2. Indeed, any mean field game on a finite state space with two elements is potential. As such, it derives from a mean field control problem. In particular, a natural question is to ask whether the solutions to the mean field game that are selected by taking the large-N limit in the finite game associated with (4.8)–(4.9) are also minimizers of the corresponding mean field control problem. The answer is yes. The same result remains open when $|E|\ge 3$. However, a simpler (but still interesting) question is to ask whether, under the same property (4.18) as before, the vanishing viscosity limits of the mean field game with common noise, as defined in Section 4.2, are minimizers of the corresponding mean field control problem. Formulated in this way, this question is also open. The main issue is that, in the presence of the common noise (and of the additional penalization φ that is necessary to guarantee the conclusion of Theorem 4.3), the mean field game is no longer potential. In order to get a potential form in (4.15), an additional penalization is necessary. Once the game with common noise is potential, it is pretty easy to take the vanishing viscosity limit in the mean field control problem that lies above. The following result is taken from Cecchin and Delarue [43]:

Theorem 4.5. *Let $m_0 = (m_0^i)_{i \in E}$ be an initial condition in the interior of the simplex. For any $\epsilon > 0$, we can find two functions $\varphi^\epsilon : [0, +\infty) \to [0, +\infty)$ and $(F^{\epsilon,i} : \mathcal{P}(E) \to \mathbb{R})_{i \in E}$, with φ^ϵ converging to 0 uniformly on any compact subset of $(0, +\infty)$, such that:*

(1) *The system (4.11) obtained by replacing (F, φ) by $(F^\epsilon, \varphi^\epsilon)$ is uniquely solvable, the solution of the forward equation being denoted by $((m_t^{\epsilon,i})_{0 \le t \le T})_{i \in E}$;*

(2) *Any weak limits of the sequence of the laws of the processes $((m_t^{\epsilon,i})_t)_{i \in E}$ has a support included in the set of minimizers of $\mathcal{J}$ in (4.19) (with the same initial condition);*

(3) *There exists a family of positive reals $(\delta_\epsilon)_\epsilon$, satisfying $\lim_{\epsilon \to 0} \delta_\epsilon = 0$, such that the trajectories $((m_t^{\epsilon,i})_t)_{i \in E}$ form $(\delta_\epsilon)_\epsilon$-approximate solutions of the mean field game with common noise of intensity $\sqrt{\epsilon}$ and with penalty φ^ϵ, as defined in Definition 4.1. In clear, if $((p_t^i)_{0 \le t \le T})_{i \in E}$ solves (4.17) (with $(\beta_t^{j,i}, m_t)_t$ being replaced by $(\varphi^\epsilon(p_t^i) + \beta_t^{j,i}, m_t^\epsilon)_t$ and with the prescription that $(\beta_t)_t$ is bounded by a fixed constant), then*

$$\left| \mathbb{E}^0 \sum_{i \in E} \left[\int_0^T p_t^i F^i(m_t^\epsilon) dt \right] - \mathbb{E}^0 \sum_{i \in E} \left[\int_0^T p_t^i F^{\epsilon,i}(m_t^\epsilon) dt \right] \right| \le \delta_\epsilon.$$

Obviously, item (3) says that the additional penalization in the definition of F^ϵ has a limited impact: The solution to the mean field game associated with the cost functional driven by F^ϵ is almost a solution of the same mean field game but associated with the cost functional driven by F. For sure, the notion of approximated solution is here consistent with the standard notion of approximated Nash equilibria: when the reference player in the population chooses a feedback function different from that chosen by the others, the best possible improvement (in the cost functional) tends to 0 with ϵ.

Interestingly, uniqueness of the minimizers (and thus of the limit points) in the second item of Theorem 4.5 is in fact the typical situation. Indeed, standard control theory says that the control problem (4.19)–(4.8) has in fact a unique minimizer at any point in time and space where the corresponding value function, which we denote by $\mathcal{V}$, is differentiable (see [23]). However, it is a standard exercise to prove that $\mathcal{V}$ is Lipschitz continuous, hence the fact that uniqueness holds for almost every starting point (in time and space) when the simplex is equipped with the $(|E| - 1)$-dimensional Lebesgue measure. Obviously, in the formulation (4.19)–(4.8), the initial time is 0, but there is no difficulty in adapting the definition to any other time $t \in [0, T]$.

Selection of solutions to the master equation. In fact, $\mathcal{V}$ plays an even more important role in the analysis of the vanishing viscosity limit as it also permits characterizing the limit of the solutions to the second-order master equation (4.14) associated with the common noise of intensity $\sqrt{\epsilon}$, with the penalty φ^ϵ and with the penalization F^ϵ (for the same choices φ^ϵ and F^ϵ as in the statement of Theorem 4.5). The next statement result is also taken from [43]:

Theorem 4.6. *With the same notation as in the statement of Theorem 4.5 and with U^ϵ denoting the solution to equation (4.14) when $\varphi \equiv \varphi^\epsilon$ and $F \equiv F^\epsilon$ therein, the limit*

$$\lim_{\epsilon \to 0} \bigl[U^{\epsilon,i}(t,q) - U^{\epsilon,j}(t,q)\bigr] = \partial_{m_i} \mathcal{V}(t,q) - \partial_{m_j} \mathcal{V}(t,q)$$

holds for almost-every $(t,q) \in [0,T] \times \mathcal{P}(E)$ and for any $i, j \in E$, where $\mathcal{V}$ is the value function of the control problem (4.19)–(4.8).

As we have already explained, the gradient of the value function exists almost-everywhere in time and space. It also important to note that the argument in the limit is not the solution of the master equation itself but the finite differences of it. In short, the limit of the master equation is just identified in dimension $|E| - 1$, which is fully consistent with the fact that the gradient of $\mathcal{V}$ is a vector of dimension $|E| - 1$. Alternatively, the above statement provides the limiting form of the feedback function used in the mean field game with a common noise of intensity $\sqrt{\epsilon}$. The $|E|$-dimensional limit of the function U^ϵ itself can be found by computing the minimal in cost (4.9) when the environment $(m_t)_t$ therein is the solution of the control problem (4.19)–(4.8).

In accordance with the program outlined above, it is a natural question to ask whether the limit established in Theorem 4.6 can be characterized in terms of the original master equation itself (i.e., the master equation (4.12) but with $\epsilon = 0$ therein). The answer is positive. As shown in [43], the master equation can be written in a conservative form. Following earlier results of Kružkov [80] and Lions [89], this conservative form has a unique solution that is bounded and satisfies a weak one-sided Lipschitz condition in space. It coincides the gradient of the value function $\mathcal{V}$. This recovers the existing results when $|E| = 2$.

4.4. Complements and open problems

Even when the state space is finite, the extension of the above results to the nonpotential case is a highly difficult problem.

Another interesting problem is to extend the same results to mean field games on continuous state spaces. The main issue is to define a suitable form of common noise. In short, this requires addressing stochastic processes with values in the infinite-dimensional space $\mathcal{P}_2(\mathbb{R}^d)$ and with sufficiently strong smoothing properties, which is known to be a challenging problem in the literature. There are earlier results in this direction, but they are not sufficient to handle the nonlinearities that make the spice of mean field games: We refer, for instance, to Stannat [100] for smoothing estimates of the Fleming–Viot process, which is an infinite-dimensional version of the Wright–Fisher noise underpinning the forward–backward system. In short, the Dirichlet form of the Fleming–Viot process is driven by the aforementioned linear-functional derivative (which provides a potential of the derivative D_m). In the meantime, the construction of a process with a Dirichlet form associated with the derivative D_m has been addressed in a series of works initiated in von Renesse and Sturm [103], but no canonical definition has yet been given. Another strategy in order to force uniqueness consists in embedding the problem in some L^2 space: following the idea underpinning

Definition 2.1, we can indeed see the unknown in a mean field game as a flow of random variables and not as a flow of probability measures. This makes it possible to use noises in Hilbert spaces. However, this destroys the mean field structure of the problem. We refer to Delarue [49] for results in this direction.

From another perspective, it is important to note that common noises in finite state mean field games can be defined a manner different from (4.11). We refer in particular to Bertucci, Lasry, and Lions [19], the key idea of which is to force the finite-player system to have many simultaneous jumps at some random times prescribed by the common noise. The reader may also have a look at [6], which provides a discrete point of view on the system (1.8). As far as the formulation (4.11) is concerned, a study of the convergence problem, very much in the spirit of Theorem 2.7, is available in [12].

5. FURTHER PROSPECTIVES AND RELATED OPEN PROBLEMS

We will now briefly review some aspects of the theory that we have not covered so far. This is only a summary presentation which demonstrates (if needed) that the field has diversified into many active branches.

5.1. Analysis of the MFG system and of the master equation

The MFG system. In the last two decades there has been a large amount of research on MFG systems of the type (which generalize (1.4)):

$$\begin{cases} \text{(i)} \quad -\partial_t u_t(x) - \Delta u_t(x) + H\big(t,x,Du_t(x),m_t\big) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ \text{(ii)} \quad \partial_t m_t(x) - \Delta m_t(x) - \operatorname{div}\big(m_t(x) D_p H\big(t,x,Du_t(x),m_t\big)\big) = 0 & \text{in } (0,T)\times\mathbb{R}^d, \\ \text{(iii)} \quad m_0(x) = \tilde{m}_0(x), \quad u(T,x) = g(x,m_T) & \text{in } \mathbb{R}^d, \end{cases}$$

and of more general (fully nonlinear) MFG systems (where $D_p H$ is the derivative of the Hamiltonian $H(t,x,p,m)$ with respect to p). It is impossible to give a complete overview of this literature: we refer to the survey [6] and to the references therein for a general presentation of this literature. The question of the existence and regularity of the solutions has been investigated in several frameworks: When the dependence of the Hamiltonian is local (depending on the pointwise value of the density), existence of classical solutions is discussed, for instance, by Cardaliaguet, Lasry, Lions, and Porretta in [29] and by Gomes, Pimentel, and Voskanyan in [68]; Porretta introduced in [97] a notion of a weak solution for these problems and proved uniqueness in this framework. The MFG system can also be set with other boundary conditions: for instance, Neumann boundary condition (Bardi and Cirant [11]), optimal stopping (Bertucci [17]), state constraints (Cannarsa, Capuani, and Cardaliaguet [22]). Mean field games can be also stated in networks (Camilli and Marchi [21] or Achdou, Dao, Ley, and Tchou [3]). Problems with congestion or with density constraints are discussed by Lions [88], Achdou and Porretta [5] and Cardaliaguet, Mészáros, and Santambrogio [32].

Variational aspects. In general, the analysis of the MFG system relies on fixed point techniques. In some frameworks (the local coupling case, for instance) it is possible to use variational methods. This turns out to be very useful for problems in which the diffusion is degenerate and for which this approach allows building weak solutions; see, for instance, the papers by Cardaliaguet and Graber [26] (first-order problems with a local coupling) and by Cardaliaguet, Graber, Porretta, and Tonon [27] (for degenerate second-order problems with a local coupling). Refining earlier result by Lions [88], Santambrogio [99] combined variational techniques with ideas from optimal transport to obtain nice regularity results of first-order MFG systems (system without diffusion). Many other references and results can be found in the survey by Santambrogio in [6].

The master equation. The analysis of the master equation has attracted some attention in the recent years, refining the earlier results [25,36,45,64,88]. Without trying to be exhaustive, one can quote the recent papers: Bertucci [18] for a notion of a weak solution under monotonicity conditions; Cardaliaguet, Porretta, and Cirant [24] for the construction of solutions to general master equations (with common noise or for a major player, see the paragraph below) on short time intervals using a kind of Trotter–Kato scheme; Gangbo, Mészáros, Mou, and Zhang [63] for the existence of a classical solution to the master equation outside the classical monotone framework, obtained by using instead conditions related with displacement convexity. Let us underline that a suitable notion of weak (discontinuous) notion of solution for the master equation is still missing (see, however, Section 4 and [43] by Cecchin and Delarue in the finite-state framework and for potential problems).

MFG problem with a major player. In general, mean field games address problems with a single homogeneous population. It is, of course, not the only interesting configuration. Among the many possible generalizations, one can mention the MFG problems with a major player, in which a controller (the major player) interacts with a population. This problem, first introduced by Huang [72], has been studied (among many other references) by Carmona and Zhu [42] by a probabilistic approach, and in [24] using the master equation. It is related with the principal–agent problems with one principal and infinitely many agents, as explained by Elie, Mastrolia, and Possamaï [57].

5.2. Mean field games of control

Mean field games of controls (sometimes also called extended mean field games) are mean field games in which players interact through the joint distribution of their positions and their controls. Many models in economics are of this type (for instance, agents interact through the price of a good that depends directly on their collective decisions to buy or sell). This kind of problem was first discussed by Gomes and Voskanyan [69]. Weak solutions have been built through a probabilistic approach by Carmona and Lacker [39]. In [35, CHAPTER 4], Carmona and Delarue pointed out the specific structure of the corresponding MFG system, which involves two fixed point problems (the classical one and a static one used to build the distribution of positions and controls from the distribution of positions and the control feedback). This MFG system was also studied in Cardaliaguet and Lehalle [31] (existence of weak solutions for problems with degenerate diffusions) and in Achdou and Kobeissi [4] (classical solution in the diffusive case and with very general interactions). Very recently Djete [54] proved the convergence of open-loop Nash equilibria for the N-player game as N tends to infinity.

5.3. Numerical methods and learning

The fixed-point nature of MFG equilibria makes them difficult to approximate and implement in practice. In the work by Achdou and Capuzzo Dolcetta [2], the authors explain how to reproduce numerically the forward–backward nature of the MFG system in order to obtain convergent numerical schemes, thus starting a series of works of the subject. An up-to-date literature on the numerical methods for mean field games, including effective methods for decoupling the two equations, can be found in Achdou's survey on this topic [6]. Recently, other works have also demonstrated the possible efficiency of tools from machine learning within this complex framework: standard equations for characterizing the equilibria may be approximately solved by means of a neural network; see, for instance, Carmona and Laurière [40,41].

Intimately related to the numerical approximation, the intriguing question of learning ("how do the MFG equilibria actually appear?") has attracted some attention. One of the first results in this direction is the transposition to MFG games of the classical fictitious play by Cardaliaguet and Hadikhanloo [28]: assuming that players know the model and that the MFG problem is potential, the method explains how players could converge to an MFG equilibrium after playing the game many times. Elie, Pérolat, Laurière, Geist, and Pietquin [58] study the effects of diverse reinforcement learning algorithms for agents with no prior information on an MFG equilibrium and learn their policy through repeated experiments. The very recent paper Delarue and Vasileiadis [53] shows that common noise may also serve as an exploration noise for learning the solution of a mean field game.

5.4. Mean field control

Mean field control (MFC) is a distinct theory from mean field games, but both theories are connected in many ways. For instance, potential games are a typical instance of mean field games that are solved by the minimizers of an MFC problem, see Proposition 2.4. The very aim of MFC theory is to address minimization problems set over Kolmogorov equations (when formulated by means of PDEs) or over McKean–Valsov equations (when formulated in a probabilistic fashion). From a particle point of view, MFC problems provide an asymptotic description of large systems of weakly interacting controlled agents who cooperate in order to minimize some common cost. Therefore, in contrast with MFGs, the agents no longer compete, and the solutions of the two problems are different. As such, this asks for a new proof of the corresponding convergence problem. We refer, for instance, to Lacker [82] for a proof based on compactness arguments, and to Djete [55] for a similar result but for models including the law of the control in the mean field interaction. As for the analysis of MFC themselves, the related value function satisfies a form of Hamilton–Jacobi equation. Similar to the master equation, the Hamilton–Jacobi equation is set on the space of probability measures, but Lions' lifting procedure allows lifting it onto an L^2 space (see [88]). This observation can be used in order to adapt the notion of viscosity solutions and thus to handle less regular solutions. We refer to [65] for a recent contribution in this direction in the first-order case, namely when the dynamics of the players are deterministic. In the presence of an idiosyncratic noise in the dynamics (so-called second-order case), the theory is still in progress and a complete theory of existence and uniqueness of viscosity solutions has not yet been achieved. We refer to Cosso and Pham [48] for an overview of the stakes.

FUNDING

F. Delarue acknowledges the financial support of French ANR project ANR-19-P3IA-0002—3IA Côte d'Azur—Nice—Interdisciplinary Institute for Artificial Intelligence.

REFERENCES

[1] Y. Achdou, F. J. Buera, J.-M. Lasry, P.-L. Lions, and B. Moll, Partial differential equation models in macroeconomics. *Philos. Trans. R. Soc. A* **372** (2014), 20130397.

[2] Y. Achdou and I. Capuzzo-Dolcetta, Mean field games: numerical methods. *SIAM J. Numer. Anal.* **48** (2010), 1136–1162.

[3] Y. Achdou, M. K. Dao, O. Ley, and N. Tchou, A class of infinite horizon mean field games on networks. *Netw. Heterog. Media* **14** (2019), 537–566.

[4] Y. Achdou and Z. Kobeissi, Mean field games of controls: finite difference approximations. 2020, arXiv:2003.03968.

[5] Y. Achdou and A. Porretta, Mean field games with congestion. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **35** (2018), 443–480.

[6] Y. Achdou P, F. Cardaliaguet, A. Delarue, Porretta, and F. Santambrogio, *Mean field games*. Lecture Notes in Math., CIME 2281, Springer, 2020.

[7] S. R. Aiyagari, Uninsured idiosyncratic risk and aggregate saving. *Q. J. Econ.* **109** (1994), 659–684.

[8] S. Albeverio, Y. G. Kondratiev, and M. Röckner, Analysis and geometry on configuration spaces. *J. Funct. Anal.* **154** (1998), 444–500.

[9] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2006.

[10] M. Bardi, Explicit solutions of some linear quadratic mean field games. *Netw. Heterog. Media* **7** (2012), 243–261.

[11] M. Bardi and M. Cirant, Uniqueness of solutions in mean field games with several populations and Neumann conditions. In *PDE models for multi-agent phenomena*, pp. 1–20, Springer, 2018.

[12] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue, Finite state mean field games with Wright–Fisher common noise as limits of N-player weighted games. 2020, arXiv:2012.04845.

[13] E. Bayraktar, A. Cecchin, A. Cohen, and F. Delarue, Finite state mean field games with Wright–Fisher common noise. *J. Math. Pures Appl.* **147** (2021), 98–162.

[14] E. Bayraktar and X. Zhang, On non-uniqueness in mean field games. *Proc. Amer. Math. Soc.* **148** (2020), 4091–4106.

[15] A. Bensoussan, J. Frehse, and P. Yam, *Mean field games and mean field type control theory*. Springer Briefs in Mathematics 101, Springer, 2013.

[16] A. Bensoussan, K. C. J. Sung, S. C. P. Yam, and S. P. Yung, Linear quadratic mean field games. *J. Optim. Theory Appl.* **169** (2016), 469–529.

[17] C. Bertucci, Optimal stopping in mean field games, an obstacle problem approach. *J. Math. Pures Appl.* **120** (2018), 165–194.

[18] C. Bertucci, Monotone solutions for mean field games master equations: finite state space and optimal stopping. 2020, arXiv:2007.11854.

[19] C. Bertucci, J.-M. Lasry, and P.-L. Lions, Some remarks on mean field games. *Comm. Partial Differential Equations* **44** (2019), 205–227.

[20] R. Buckdahn, J. Li, S. Peng, and C. Rainer, Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* **45** (2017), 824–878.

[21] F. Camilli and C. Marchi, Stationary mean field games systems defined on networks. *SIAM J. Control Optim.* **54** (2016), no. 2, 1085–1103.

[22] P. Cannarsa, R. Capuani, and P. Cardaliaguet, Mean field games with state constraints: from mild to pointwise solutions of the pde system. 2018, arXiv:1812.11374.

[23] P. Cannarsa and C. Sinestrari, *Semiconcave functions, Hamilton–Jacobi equations, and optimal control*. Progr. Nonlinear Differential Equations Appl. 58, Birkhäuser Boston, Inc., Boston, MA, 2004.

[24] P. Cardaliaguet, M. Cirant, and A. Porretta, Splitting methods and short time existence for the master equations in mean field games. 2020, arXiv:2001.10406.

[25] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions, *The master equation and the convergence problem in mean field games*. Ann. of Math. Stud. 201, Princeton University Press, 2019.

[26] P. Cardaliaguet and P. J. Graber, Mean field games systems of first order. *ESAIM Control Optim. Calc. Var.* **21** (2015), 690–722.

[27] P. Cardaliaguet, P. J. Graber, A. Porretta, and D. Tonon, Second order mean field games with degenerate diffusion and local coupling. *NoDEA Nonlinear Differential Equations Appl.* **22** (2015), 1287–1317.

[28] P. Cardaliaguet and S. Hadikhanloo, Learning in mean field games: the fictitious play. *ESAIM Control Optim. Calc. Var.* **23** (2017), no. 2, 569–591.

[29] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, and A. Porretta, Long time average of mean field games. *Netw. Heterog. Media* **7** (2012).

[30] P. Cardaliaguet, J.-M. Lasry, P.-L. Lions, and A. Porretta, Long time average of mean field games with a nonlocal coupling. *SIAM J. Control Optim.* **51** (2013), 3558–3591.

[31] P. Cardaliaguet and C.-A. Lehalle, Mean field game of controls and an application to trade crowding. *Math. Financ. Econ.* **12** (2018), 335–363.

[32] P. Cardaliaguet, A. R. Mészáros, and F. Santambrogio, First order mean field games with density constraints: pressure equals price. *SIAM J. Control Optim.* **54** (2016), 2672–2709.

[33] P. Cardaliaguet and A. Porretta, Long time behavior of the master equation in mean field game theory. *Anal. PDE* **12** (2019), 1397–1453.

[34] R. Carmona and F. Delarue, Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** (2013), 2705–2734.

[35] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games*. Probab. Theory Stoch. Model. 83, Springer, Cham, 2018.

[36] R. Carmona and F. Delarue, *Probabilistic theory of mean field games with applications. II. Mean field games with common noise and master equations*. Probab. Theory Stoch. Model. 84, Springer, Cham, 2018.

[37] R. Carmona, F. Delarue, and A. Lachapelle, Control of McKean–Vlasov versus mean field games. *Math. Financ. Econ.* **7** (2013), 131–166.

[38] R. Carmona, F. Delarue, and D. Lacker, Mean field games with common noise. *Ann. Probab.* **44** (2016), 3740–3803.

[39] R. Carmona and D. Lacker, A probabilistic weak formulation of mean field games and applications. *Ann. Appl. Probab.* **25** (2015), 1189–1231.

[40] R. Carmona and M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean-field control and games: I – The ergodic case. 2019, arXiv:1907.05980.

[41] R. Carmona and M. Laurière, Convergence analysis of machine learning algorithms for the numerical solution of mean-field control and games: II – The finite horizon case. 2019, arXiv:1908.01613.

[42] R. Carmona and X. Zhu, A probabilistic approach to mean field games with major and minor players. *Ann. Appl. Probab.* **26** (2016), 1535–1580.

[43] A. Cecchin and F. Delarue, Selection by vanishing common noise for potential finite state mean field games. *Comm. Partial Differential Equations* (2021). DOI 10.1080/03605302.2021.1955256.

[44] A. Cecchin, P. D. Pra, M. Fischer, and G. Pelino, On the convergence problem in mean field games: a two state model without uniqueness. *SIAM J. Control Optim.* **57** (2019), 2443–2466.

[45] J.-F. Chassagneux, D. Crisan, and F. Delarue, Classical solutions to the master equation for large population equilibria. 2015, arXiv:1411.3009.

[46] M. Cirant, On the existence of oscillating solutions in non-monotone mean-field games. *J. Differential Equations* **266** (2019), 8067–8093.

[47] M. Cirant and A. Porretta, Long time behavior and turnpike solutions in mildly non-monotone mean field games. *ESAIM Control Optim. Calc. Var.* **27** (2021), 86, 40 pp.

[48] A. Cosso and H. Pham, Zero-sum stochastic differential games of generalized McKean–Vlasov type. *J. Math. Pures Appl.* **129** (2019), 180–212.

[49] F. Delarue, Restoring uniqueness to mean-field games by randomizing the equilibria. *Stoch. Partial Differ. Equ. Anal. Comput.* **7** (2019), 598–678.

[50] F. Delarue, D. Lacker, and K. Ramanan, From the master equation to mean field game limit theory: a central limit theorem. *Electron. J. Probab.* **24** (2019), 1–54.

[51] F. Delarue, D. Lacker, and K. Ramanan, From the master equation to mean field game limit theory: large deviations and concentration of measure. *Ann. Probab.* **48** (2020), 211–263.

[52] F. Delarue and R. F. Tchuendom, Selection of equilibria in a linear quadratic mean-field game. *Stochastic Process. Appl.* **130** (2020), 1000–1040.

[53] F. Delarue and A. Vasileiadis, Exploration noise for learning linear–quadratic mean field games. 2021, arXiv:2107.00839.

[54] M. F. Djete, Mean field games of controls: on the convergence of Nash equilibria. 2020, arXiv:2006.12993.

[55] M. F. Djete, Extended mean field control problem: a propagation of chaos result. 2020, arXiv:2006.12996.

[56] J. Doncel, N. Gast, and B. Gaujal, Are mean-field games the limits of finite stochastic games? *ACM SIGMETRICS Perform. Eval. Rev.* **44** (2016), no. 2, 18–20. Special Issue on the Workshop on MAthematical performance Modeling and Analysis (MAMA 2016).

[57] R. Elie, T. Mastrolia, and D. Possamaï, A tale of a principal and many, many agents. *Math. Oper. Res.* **44** (2019), 440–467.

[58] R. Elie, J. Pérolat, M. Laurière, M. Geist, and O. Pietquin, Approximate fictitious play for mean field games. 2019, arXiv:1907.02633.

[59] C. L. Epstein and R. Mazzeo, *Degenerate diffusion operators arising in population biology*. Ann. of Math. Stud. 185, Princeton University Press, Princeton, NJ, 2013.

[60] A. Fathi, Weak KAM theory: the connection between Aubry–Mather theory and viscosity solutions of the Hamilton–Jacobi equation. In *Proceedings of the International Congress of Mathematicians* 3, pp. 597–621, Kyung Moon Sa, Seoul, 2014.

[61] M. Fischer, On the connection between symmetric N-player games and mean field games. *Ann. Appl. Probab.* **27** (2017), 757–810.

[62] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** (2015), 707–738.

[63] W. Gangbo, A. R. Mészáros, C. Mou, and J. Zhang, Mean field games master equations with non-separable Hamiltonians and displacement monotonicity. 2021, arXiv:2101.12362.

[64] W. Gangbo and A. Swiech, Existence of a solution to an equation arising from the theory of mean field games. *J. Differential Equations* **259** (2015), 6573–6643.

[65] W. Gangbo and A. Tudorascu, On differentiability in the Wasserstein space and well-posedness for Hamilton–Jacobi equations. *J. Math. Pures Appl.* **125** (2019), 119–174.

[66] D. A. Gomes, J. Mohr, and R. R. Souza, Discrete time, finite state space mean field games. *J. Math. Pures Appl.* **93** (2010), 308–328.

[67] D. A. Gomes, J. Mohr, and R. R. Souza, Continuous time finite state mean field games. *Appl. Math. Optim.* **68** (2013), 99–143.

[68] D. A. Gomes, E. A. Pimentel, and V. Voskanyan, *Regularity theory for mean-field game systems*. Springer, Berlin, 2016.

[69] D. A. Gomes and V. K. Voskanyan, Extended deterministic mean-field games. *SIAM J. Control Optim.* **54** (2016), 1030–1055.

[70] O. Guéant, Existence and uniqueness result for mean field games with congestion effect on graphs. *Appl. Math. Optim.* **72** (2015), 291–303.

[71] O. Gueant, J.-M. Lasry, and P.-L. Lions, Mean field games and applications. In *Paris–Princeton lectures on mathematical finance*, pp. 205–266, Springer, Berlin, Heidelberg, 2010.

[72] M. Huang, Large-population LQG games involving a major player: the Nash certainty equivalence principle. *SIAM J. Control Optim.* **48** (2010), 3318–3353.

[73] M. Huang, P. E. Caines, and R. P. Malhamé, Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. In *Proceedings of the 42nd IEEE Conference on Decision and Control* 1, pp. 98–103, IEEE, 2003.

[74] M. Huang, R. P. Malhamé, and P. E. Caines, Large population stochastic dynamic games: closed-loop Mckean–Vlasov systems and the Nash certainty equivalence principle. *Commun. Inf. Syst.* **6** (2006), 221–252.

[75] M. Huang, R. P. Malhamé, and P. E. Caines, Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε-Nash equilibria. *IEEE Trans. Automat. Control* **52** (2007), 1560–1571.

[76] M. Kac, Foundations of kinetic theory. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability, 1954–1955* III, pp. 171–197, University of California Press, Berkeley and Los Angeles, CA, 1956.

[77] M. Kimura, Diffusion models in population genetics. *J. Appl. Probab.* **1** (1964), 177–232.

[78] V. N. Kolokoltsov, *Nonlinear Markov processes and kinetic equations* 182, Cambridge University Press, 2010.

[79] P. Krusell and A. A. Smith Jr., Income and wealth heterogeneity in the macroeconomy. *J. Polit. Econ.* **106** (1998), 867–896.

[80] S. N. Kružkov, Generalized solutions of nonlinear equations of the first order with several independent variables. II. *Mat. Sb.* **72** (1967), 108–134.

[81] D. Lacker, A general characterization of the mean field limit for stochastic differential games. *Probab. Theory Related Fields* **165** (2016), 581–648.

[82] D. Lacker, Limit theory for controlled McKean–Vlasov dynamics. *SIAM J. Control Optim.* **55** (2017), 1641–1672.

[83] D. Lacker, On the convergence of closed-loop Nash equilibria to the mean field game limit. *Ann. Appl. Probab.* **30** (2020), 1693–1761.

[84] O. A. Ladyzhenskaia, V. A. Solonnikov, and N. Uraltseva, *Linear and quasi-linear equations of parabolic type* 23, Amer. Math. Soc., Providence, RI, 1998.

[85] J.-M. Lasry and P.-L. Lions, Jeux à champ moyen. I – le cas stationnaire. *C. R. Math.* **343** (2006), 619–625.

[86] J.-M. Lasry and P.-L. Lions, Jeux à champ moyen. II – horizon fini et contrôle optimal. *C. R. Math.* **343** (2006), 679–684.

[87] J.-M. Lasry and P.-L. Lions, Mean field games. *Jpn. J. Math.* **2** (2007), 229–260.

[88] P.-L. Lions, Cours au College de France, 2007–2012.

[89] P.-L. Lions, *Generalized solutions of Hamilton-Jacobi equations*. Res. Notes Math. 69, Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.

[90] P.-L. Lions, G. Papanicolaou, and S. S. Varadhan, *Homogenization of Hamilton–Jacobi equations*. 1986. Unpublished work.

[91] R. E. Lucas Jr. and B. Moll, Knowledge growth and the allocation of time. *J. Polit. Econ.* **122** (2014), no. 1, 1–51.

[92] H. P. McKean, A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** (1966), 1907–1911.

[93] S. Mischler and C. Mouhot, Kac's program in kinetic theory. *Invent. Math.* **193** (2013), 1–147.

[94] S. Mischler, C. Mouhot, and B. Wennberg, A new approach to quantitative propagation of chaos for drift, diffusion and jump processes. *Probab. Theory Related Fields* **161** (2015), 1–59.

[95] G. Papanicolaou, L. Ryzhik, and K. Velcheva, Travelling waves in a mean field learning model. 2020, arXiv:2002.06287.

[96] S. Peng, Stochastic Hamilton–Jacobi–Bellman equations. *SIAM J. Control Optim.* **30** (1992), 284–304.

[97] A. Porretta, Weak solutions to Fokker–Planck equations and mean field games. *Arch. Ration. Mech. Anal.* **216** (2015), 1–62.

[98] A. Porretta and L. Rossi, Travelling waves in a mean field game model of knowledge diffusion. 2020, arXiv:2010.10828.

[99] F. Santambrogio, Regularity via duality in calculus of variations and degenerate elliptic pdes. *J. Math. Anal. Appl.* **457** (2018), 1649–1674.

[100] W. Stannat, Long-time behaviour and regularity properties of transition semigroups of Fleming–Viot processes. *Probab. Theory Related Fields* **122** (2002), 431–469.

[101] A.-S. Sznitman, Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pp. 165–251, Springer, 1991.

[102] R. F. Tchuendom, Uniqueness for linear–quadratic mean field games with common noise. *Dyn. Games Appl.* **8** (2018), 199–210.

[103] M.-K. von Renesse and K. T. Sturm, Entropic measure and Wasserstein diffusion. *Ann. Probab.* **37** (2009), 1114–1191.

PIERRE CARDALIAGUET

Université Paris Dauphine-PSL, Place du Mal de Lattre de Tassigny, 75775 Paris Cedex 16, France, cardaliaguet@ceremade.dauphine.fr

FRANÇOIS DELARUE

Université Côte d'Azur, Laboratoire J. A. Dieudonné, CNRS, Parc Valrose, 06108 Nice Cedex 02, France, delarue@univ-cotedazur.fr

MACROSCOPIC LIMITS OF CHAOTIC EIGENFUNCTIONS

SEMYON DYATLOV

ABSTRACT

We give an overview of the interplay between the behavior of high energy eigenfunctions of the Laplacian on a compact Riemannian manifold and the dynamical properties of the geodesic flow on that manifold. This includes the Quantum Ergodicity theorem, the Quantum Unique Ergodicity conjecture, entropy bounds, and uniform lower bounds on mass of eigenfunctions. The above results belong to the domain of *quantum chaos* and use *microlocal analysis*, which is a theory behind the classical/quantum, or particle/wave, correspondence in physics. We also discuss the toy model of quantum cat maps and the challenges it poses for Quantum Unique Ergodicity.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 58J51; Secondary 58J40, 81Q50

KEYWORDS

Quantum chaos, quantum ergodicity, eigenfunctions, fractal uncertainty principle

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3704–3723
DOI 10.4171/ICM2022/22

1. INTRODUCTION

This article is an overview of some results on *macroscopic behavior of eigenstates in the high energy limit*. A typical model is given by Laplacian eigenfunctions:

$$-\Delta_g u_\lambda = \lambda^2 u_\lambda, \quad u_\lambda \in C^\infty(M), \quad \|u_\lambda\|_{L^2(M)} = 1.$$

Here we fix a compact connected Riemannian manifold without boundary (M, g) and denote by $\Delta_g \leq 0$ the corresponding Laplace–Beltrami operator. It will be convenient to denote the eigenvalue by λ^2, where $\lambda \geq 0$. The high-energy limit corresponds to taking $\lambda \to \infty$.

One way to study the macroscopic behavior of the eigenfunctions u_λ as $\lambda \to \infty$ is to look at weak limits of the probability measures $|u_\lambda|^2 \, d\operatorname{vol}_g$ where $d\operatorname{vol}_g$ is the volume measure on (M, g):

Definition 1. Let λ_j^2 be a sequence of eigenvalues of $-\Delta_g$ going to ∞. We say that the corresponding eigenfunctions u_{λ_j} converge weakly to some probability measure ν on M, if

$$\int_M a(x) \big| u_{\lambda_j}(x) \big|^2 \, d\operatorname{vol}_g(x) \to \int_M a(x)\, d\nu(x) \quad \text{as} \quad j \to \infty \tag{1.1}$$

for all test functions $a \in C^\infty(M)$.

Definition 1 can be interpreted in the context of quantum mechanics as follows. Consider a free quantum particle on the manifold M. Then the eigenfunctions u_λ are the wave functions of the *pure quantum states* of the particle. The left-hand side of (1.1) is the average value of the observable $a(x)$ for a given pure state; if we let a be the characteristic function of some set $\Omega \subset M$ then this expression is the probability of finding the quantum particle in Ω (this choice is only allowed if $\nu(\partial\Omega) = 0$). Taking $\lambda \to \infty$ gives the high-energy limit.

The statement (1.1) is macroscopic in nature because we first fix the observable a and then let the eigenvalue go to infinity. This is different from *microscopic* properties such as the breakthrough work of Logunov and Malinnikova on the area of the *nodal set* $\{x \in M \mid u_j(x) = 0\}$, see the review [38]. Ironically, the methods used in the macroscopic results described here are *microlocal* in nature (see Section 2 for a review), with the global geometry of M coming in the form of the long time behavior of the geodesic flow.

The results reviewed in this paper address the following fundamental question:

$$\begin{gathered}\text{For a given Riemannian manifold } (M, g)\text{, what can we say}\\ \text{about the set of all weak limits of sequences of eigenfunctions?}\end{gathered} \tag{1.2}$$

It turns out that the answer depends on the dynamical properties of the *geodesic flow* on (M, g). In particular:

- If (M, g) has *completely integrable* geodesic flow then there is a huge variety of possible weak limits. For example, if (M, g) is the round sphere, then there is a sequence of Gaussian beam eigenfunctions converging to the delta measure on any given closed geodesic (see Section 2.2 below).

- If the geodesic flow instead has *chaotic* behavior, more precisely it is ergodic with respect to the Liouville measure, then a density 1 sequence of eigenfunctions converges to the volume measure $d\,\mathrm{vol}_g / \mathrm{vol}_g(M)$. This statement, known as *Quantum Ergodicity*, is reviewed in Section 3.

- If the geodesic flow is *strongly chaotic*, more precisely it satisfies the Anosov property (i.e., it has a stable/unstable/flow decomposition), then the limiting measures have to be somewhat spread out. This comes in two forms: *entropy bounds* and *full support*. See Section 4 for a description of these results. The *Quantum Unique Ergodicity* conjecture states that in this setting any sequence of eigenfunctions converges to the volume measure; it is not known outside of arithmetic cases (see Section 4) and there are counterexamples in the related setting of quantum cat maps (see Section 5).

- Finally, there are several results in cases when the geodesic flow is ergodic but not Anosov, or it exhibits mixed chaotic/completely integrable behavior; see Section 3.

The present article focuses on the last three cases above, which are in the domain of *quantum chaos*. The general principle is that *chaotic behavior of the geodesic flow leads to chaotic/spread out macroscopic behavior of the eigenfunctions of the Laplacian*. See Figure 1 for a numerical illustration.

In particular, we will describe full support statements for weak limits – see Theorems 11 and 16 – proved in [18–20]. The key component is the *fractal uncertainty principle* first introduced by Dyatlov–Zahl [21] and proved by Bourgain–Dyatlov [10]. It originated in *open* quantum chaos, dealing with quantum systems where the underlying classical system allows for escape to infinity and has chaotic behavior. We refer to the reviews of the author [15, 16] for more on fractal uncertainty principle and its applications.

The above developments use *microlocal analysis*, which is a mathematical theory underlying the classical/quantum, or particle/wave, correspondence in physics. In particular, one typically obtains information on the *semiclassical measures*, which are probability measures μ on the cosphere bundle S^*M which are weak limits of sequences of eigenfunctions in a microlocal sense. These measures are sometimes called *microlocal lifts* of the weak limits, because the pushforward of μ to the base M is the weak limit of Definition 1. One of the advantages of these measures compared to the weak limits on M is that they are invariant under the geodesic flow. We give a brief review of microlocal analysis and semiclassical measures in Section 2 below.

2. SEMICLASSICAL MEASURES

Let us write the left-hand side of (1.1) as

$$\int_M a(x)\left|u_{\lambda_j}(x)\right|^2 d\,\mathrm{vol}_g(x) = \langle \mathbf{M}_a u_{\lambda_j}, u_{\lambda_j}\rangle_{L^2(M)}$$

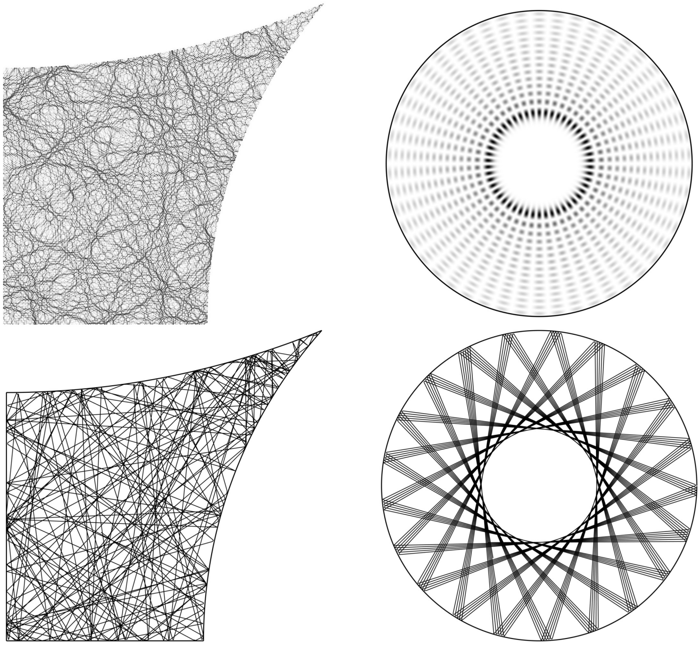


FIGURE 1

(Top) Typical eigenfunctions (with Dirichlet boundary conditions) for two planar domains. The picture on the left (courtesy of Alex Barnett, see [7] and [8] for a description of the method used and for a numerical investigation of Quantum Ergodicity) shows equidistribution, i.e., convergence to the volume measure in the sense of Definition 1. The picture on the right (where the domain is a disk) shows the lack of equidistribution, with the limiting measure supported in an annulus. This difference in quantum behavior is related to the different behavior of the billiard-ball flows on the two domains (which replace geodesic flows in this setting). (Bottom) Two typical billiard-ball trajectories on the domains in question. On the left we see ergodicity (equidistribution of the trajectory for long time), and on the right we see completely integrable behavior.

where $\mathbf{M}_a : L^2(M) \to L^2(M)$ is the multiplication operator by $a \in C^\infty(M)$. To define semiclassical measures, we will allow for more general operators in place of $\mathbf{M}_a$. These operators are obtained by a *quantization procedure*, which maps each smooth compactly supported function a on the cotangent bundle T^*M to an operator on $L^2(M)$ depending on the small number $h > 0$ called the semiclassical parameter:

$$a \in C^\infty_{\mathrm{c}}(T^*M) \quad \mapsto \quad \mathrm{Op}_h(a) : L^2(M) \to L^2(M), \quad 0 < h \ll 1. \tag{2.1}$$

2.1. Semiclassical quantization

We briefly recall several basic principles of semiclassical quantization referring to the books of Zworski [49] and Dyatlov–Zworski [22, APPENDIX E] for the full presentation and pointers to the vast literature on the subject:

- The function a, often called the *symbol* of the operator $\mathrm{Op}_h(a)$, is defined on the cotangent bundle T^*M, whose points we typically denote by (x,ξ) where $x \in M$ and $\xi \in T^*_xM$. The canonical symplectic form on T^*M induces the *Poisson bracket*

$$\{f,g\} := \partial_\xi f \cdot \partial_x g - \partial_x f \cdot \partial_\xi g, \quad f,g \in C^\infty(T^*M).$$

  In physical terms, this corresponds to using Hamiltonian mechanics for the "classical" side of the classical/quantum correspondence, where x is the position variable and ξ is the momentum variable.

- One can work with a broader class of smooth symbols a, where the compact support requirement is changed to growth conditions on the derivatives of a as $\xi \to \infty$. The resulting operators act on (semiclassical) Sobolev spaces, see, e.g.. [22, §E.1.8].

- If $a(x,\xi) = a(x)$ is a function of x only, then

$$\mathrm{Op}_h(a) = \mathbf{M}_a \tag{2.2}$$

  is the corresponding multiplication operator.

- If $a(x,\xi)$ is linear in ξ, that is, $a(x,\xi) = \langle \xi, X_x \rangle$ for some vector field $X \in C^\infty(M;TM)$, then, up to lower-order terms, the operator $\mathrm{Op}_h(a)$ is a rescaled differentiation operator along X,

$$\mathrm{Op}_h(a)u(x) = -ihXu(x) + \mathcal{O}(h). \tag{2.3}$$

  This explains why a should be a function on the cotangent bundle T^*M: linear functions on the fibers of T^*M correspond to vector fields on M. (Quantization procedures do not depend on the choice of a Riemannian metric on M.)

- If $u \in C^\infty(M)$ oscillates at some frequency R, then differentiating u along a vector field X increases its magnitude by about R. One takeaway from (2.3) is that $\mathrm{Op}_h(a)u$ has roughly the same size as u if the function u oscillates at frequencies $\sim h^{-1}$. Thus we treat the semiclassical parameter h as the *effective wavelength* of oscillations of the functions to which we will apply $\mathrm{Op}_h(a)$. We will apply $\mathrm{Op}_h(a)$ to an eigenfunction u_λ, which oscillates at frequency $\sim \lambda$, so we will make the choice

$$h := \lambda^{-1}. \tag{2.4}$$

- If $M = \mathbb{R}^n$ and $a(x,\xi) = a(\xi)$ is a function of ξ only, then $\mathrm{Op}_h(a)$ is a Fourier multiplier,

$$\widehat{\mathrm{Op}_h(a)u}(\xi) = a(h\xi)\hat{u}(\xi), \quad u \in \mathscr{S}(\mathbb{R}^n). \tag{2.5}$$

  Thus in addition to being the momentum variable, we can interpret ξ as a Fourier/frequency variable.

- For general manifolds M, one cannot define a quantization procedure canonically: a typical construction involves piecing together quantizations on copies of $\mathbb{R}^n$ using coordinate charts, see, e.g., [22, §E.1.7]. However, different choices of coordinate charts, etc., will give the same operator modulo lower-order terms $\mathcal{O}(h)$.

Several items above allude to "lower-order terms." We will consider the operators $\operatorname{Op}_h(a)$ in the *semiclassical limit* $h \to 0$ and will often have remainders of the form $\mathcal{O}(h)$, etc., which are operators on $C^\infty(M)$. (More generally, semiclassical analysis gives asymptotic expansions in powers of h with the remainder being $\mathcal{O}(h^N)$ for any N.) This is understood as follows: if the symbols involved are compactly supported in T^*M, then the remainders are bounded in norm as operators on L^2 (with constants in $\mathcal{O}(\bullet)$, of course, independent of h). For more general symbols, one has to take correct semiclassical Sobolev spaces and we skip these details here. We note that in the basic version of semiclassical calculus used in this section, the symbol a does not depend on h, which reflects the macroscopic nature of the results presented below.

Semiclassical quantization has several fundamental algebraic and analytic properties; once these are proved, one can use it as a black box without caring too much for the precise definition of $\operatorname{Op}_h(a)$. Of particular importance are the product, adjoint, and commutator rules:

$$\operatorname{Op}_h(a)\operatorname{Op}_h(b) = \operatorname{Op}_h(ab) + \mathcal{O}(h), \tag{2.6}$$

$$\operatorname{Op}_h(a)^* = \operatorname{Op}_h(\bar{a}) + \mathcal{O}(h), \tag{2.7}$$

$$\big[\operatorname{Op}_h(a), \operatorname{Op}_h(b)\big] = -ih\operatorname{Op}_h\big(\{a,b\}\big) + \mathcal{O}(h^2), \tag{2.8}$$

and the L^2 boundedness statement: if $a \in C_c^\infty(T^*M)$ then $\|\operatorname{Op}_h(a)\|_{L^2\to L^2}$ is bounded uniformly in h.

2.2. Semiclassical measures for eigenfunctions

We can now introduce the main object of study in this article, which are semiclassical measures associated to high frequency sequences of eigenfunctions of the Laplacian. Semiclassical measures were originally introduced independently by Gérard [27] and Lions–Paul [37]. We refer to [49, CHAPTER 5] for a detailed treatment.

Following (2.4), we write the eigenvalue as h^{-2} where h is small. Let (M,g) be a Riemannian manifold and consider a sequence of Laplacian eigenfunctions:

$$-\Delta_g u_j = h_j^{-2} u_j, \quad h_j \to 0, \quad u_j \in C^\infty(M), \quad \|u_j\|_{L^2} = 1.$$

Definition 2. We say that the sequence u_j converges semiclassically to a finite Borel measure μ on the cotangent bundle T^*M, if

$$\big\langle \operatorname{Op}_{h_j}(a)u_j, u_j\big\rangle_{L^2} \to \int_{T^*M} a(x,\xi)\,d\mu(x,\xi) \quad \text{as} \quad j\to\infty \tag{2.9}$$

for all test functions $a \in C_c^\infty(T^*M)$. A measure μ on T^*M is called a *semiclassical measure* if it is the limit of some sequence of Laplacian eigenfunctions.

The statement (2.9) actually applies to a broader class of symbols a with polynomial growth as $\xi \to \infty$. By (2.2), if $a(x,\xi) = a(x)$ depends only on the position variable x, then the left-hand side of (2.9) is the integral $\int_M a|u_j|^2\, d\,\mathrm{vol}_g$. Comparing (2.9) with (1.1), we see that if u_j converges semiclassically to μ, then it converges weakly to the pushforward of μ to the base M. Thus we can think of semiclassical measures as (microlocal) lifts of the weak limits of Definition 1.

A quantum-mechanical interpretation of semiclassical measures is as follows: if $a \in C^\infty(T^*M)$ is a *classical observable* (a function of position and momentum) then $\mathrm{Op}_h(a)$ is the corresponding *quantum observable* and the expression $\langle \mathrm{Op}_h(a)u, u\rangle_{L^2}$ is the average value of the observable a on the quantum particle with wave function u. Thus (2.9) gives macroscopic information on the concentration of the particle in both position and momentum in the high-energy limit. Recalling (2.5), we can also interpret semiclassical measures as capturing the concentration of u_j simultaneously in the position and frequency.

One important property of Definition 2 is the presence of compactness: any sequence of eigenfunctions has a subsequence converging semiclassically to some measure; see [49, THEOREM 5.2] and [22, THEOREM E.42]. Other basic properties of semiclassical measures are summarized in the following

Proposition 3. *Let μ be a semiclassical measure for a Riemannian manifold (M,g). Then:*

- *μ is a probability measure;*
- *μ is supported on the cosphere bundle*

$$S^*M := \{(x,\xi) \in T^*M : |\xi|_g = 1\};$$

- *μ is invariant under the geodesic flow*

$$\varphi^t : S^*M \to S^*M.$$

Here the geodesic flow is naturally a flow on the sphere bundle SM, which is identified with S^*M using the metric g.

We give a sketch of the proof of Proposition 3 to show how the fundamental properties (2.6)–(2.8) can be used. The first claim follows by taking $a = 1$ in (2.9), in which case $\mathrm{Op}_h(a)$ is the identity operator. To see the second claim, we use that the semiclassically rescaled Laplacian $-h^2\Delta_g$ is a quantization of the quadratic function $|\xi|_g^2$ (giving the square of the length of the cotangent vector $\xi \in T_x^*M$ with respect to the metric g), so

$$P(h) := -h^2\Delta_g - 1 = \mathrm{Op}_h\big(|\xi|_g^2 - 1\big) + \mathcal{O}(h), \quad P(h_j)u_j = 0.$$

Now if $a \in C_c^\infty(T^*M)$ vanishes on S^*M, we can write $a = b(|\xi|_g^2 - 1)$ for some $b \in C_c^\infty(T^*M)$. By the product rule (2.6),

$$\mathrm{Op}_{h_j}(a)u_j = \mathrm{Op}_{h_j}(b)P(h_j)u_j + \mathcal{O}(h_j) = \mathcal{O}(h_j),$$

which by (2.9) gives $\int_{T^*M} a\, d\mu = 0$. Since this is true for any a vanishing on S^*M, we see that $\operatorname{supp}\mu \subset S^*M$ as needed.

The last claim is also simple to prove: if $b \in C_c^\infty(T^*M)$ is arbitrary, then

$$0 = \langle [P(h_j), \mathrm{Op}_{h_j}(b)] u_j, u_j \rangle_{L^2} = -ih_j \langle \mathrm{Op}_{h_j}(\{|\xi|_g^2, b\}) u_j, u_j \rangle_{L^2} + \mathcal{O}(h_j^2).$$

Here the first equality follows from the fact that $P(h_j)u_j = 0$ and $P(h_j)$ is self-adjoint; the second one uses the commutator rule (2.8). Now (2.9) shows that the Poisson bracket $\{|\xi|_g^2, b\}$ integrates to 0 with respect to μ. But the Hamiltonian flow of $|\xi|_g^2/2$, restricted to S^*M, is the geodesic flow φ^t, so we get

$$\int_{S^*M} \partial_t|_{t=0}(b \circ \varphi^t)\, d\mu = 0 \quad \text{for all } b \in C_c^\infty(T^*M),$$

from which it follows that $\int_{S^*M} b \circ \varphi^t \, d\mu$ is independent of t and thus μ is invariant under the flow φ^t.

We now give the microlocal formulation of the question (1.2) asked at the beginning of the article:

$$\begin{gathered}\text{For a given Riemannian manifold } (M, g)\text{, what can we say} \\ \text{about the set of all semiclassical measures?}\end{gathered} \tag{2.10}$$

The general expectation is that

- when the geodesic flow on (M, g) is “predictable,” i.e., completely integrable, there are semiclassical measures which can concentrate on small flow-invariant sets;
- on the other hand, when the geodesic flow on (M, g) has chaotic behavior, semiclassical measures have to be more “spread out.”

One of the results supporting the first point above is the following theorem of Jakobson–Zelditch [33]: if M is the round sphere then *any* measure satisfying the conclusions of Proposition 3 is a semiclassical measure. See also the work of Studnia [46] and Arnaiz–Macià [6] in the related case of the quantum harmonic oscillator.

The rest of this article presents various results which support the second point above, in particular giving several ways of defining chaotic behavior of the geodesic flow and the way in which a measure is “spread out.”

3. ERGODIC SYSTEMS

We first describe what happens under a “mildly chaotic” assumption on the geodesic flow $\varphi^t : S^*M \to S^*M$, namely that it is *ergodic* with respect to the Liouville measure. Here the Liouville measure $\mu_L = c\, d\,\mathrm{vol}_g(x)\, dS(\xi)$ is a natural flow-invariant probability measure on S^*M, with dS denoting the volume measure on the sphere S_x^*M corresponding to g and c some constant. By definition, the flow φ^t is ergodic with respect to μ_L if every φ^t-invariant Borel subset $\Omega \subset S^*M$ has $\mu_L(\Omega) = 0$ or $\mu_L(\Omega) = 1$.

We say that a sequence of eigenfunctions u_j *equidistributes* if it converges to μ_L in the sense of Definition 2, that is, in the high-energy limit the probability of finding the corresponding quantum particle in a set becomes proportional to the volume of this set. A central

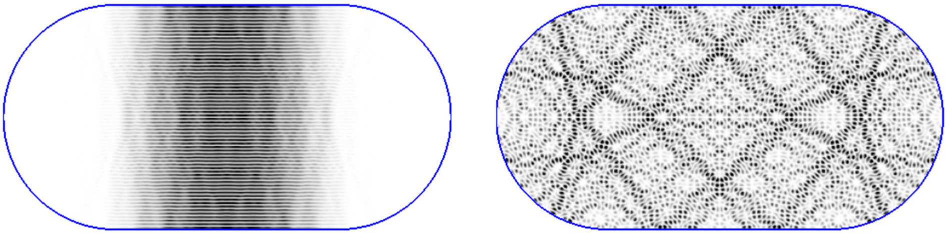


FIGURE 2
Two Dirichlet eigenfunctions for a Bunimovich stadium, courtesy of Alex Barnett (see the caption to Figure 1): the right one shows equidistribution, but the left one does not. Quantum Ergodicity implies that most eigenfunctions look from afar like that on the right.

result in quantum chaos is the following Quantum Ergodicity theorem of Shnirelman [44], Zelditch [47], and Colin de Verdière [14], which states that when the geodesic flow is ergodic, most eigenfunctions equidistribute:

Theorem 4. *Assume that the geodesic flow is ergodic with respect to the Liouville measure. Then for any choice of orthonormal basis of eigenfunctions* $\{u_k\}$ *there exists a density* 1 *subsequence* u_{k_j} *which converges semiclassically to* μ_L *in the sense of Definition* 2.

See [49, CHAPTER 15] and the review of Dyatlov [17] for more recent expositions of the proof. The version of Theorem 4 for compact manifolds with boundary was proved by Gérard–Leichtnam [28] for convex domains in $\mathbb{R}^n$ with $W^{2,\infty}$ boundaries and Zelditch–Zworski [48] for compact Riemannian manifolds with piecewise C^∞ boundaries. In this setting one imposes (Dirichlet or Neumann) boundary conditions on the eigenfunctions, and the geodesic flow is naturally replaced by the billiard-ball flow (reflecting off the boundary). See Figures 1 and 2 for numerical illustrations.

A natural question is whether the entire sequence of eigenfunctions equidistributes, i.e., whether μ_L is the *only* semiclassical measure. For general manifolds with ergodic classical flows this is not always true, as proved by Hassell [32]. In particular, for the case of the Bunimovich stadium shown on Figure 2, the paper [32] shows that for almost every choice of the parameter of the stadium (i.e., the aspect ratio of its central rectangle) there exist semiclassical measures which are not the Liouville measure.

Another natural question is what happens when the classical flow has *mixed* behavior, e.g., S^*M is the union of two flow-invariant sets of positive Lebesgue measure such that the flow is ergodic on one of them and completely integrable on the other. *Percival's Conjecture* claims that this mixed behavior translates to macroscopic behavior of eigenfunctions, namely one can split any orthonormal basis of eigenfunctions into three parts: one of them equidistributes in the ergodic region, another has semiclassical measures supported in the completely integrable region, and the remaining part has density 0. A version of this conjecture for mushroom billiards was proved by Gomes in his thesis [29, 30]; see also the earlier work of Galkowski [26] and Rivière [41].

4. STRONGLY CHAOTIC SYSTEMS

We now describe what is known when the geodesic flow on M is assumed to be strongly chaotic. The latter assumption is understood in the sense of the following *Anosov property*:

Definition 5. Let (M, g) be a compact Riemannian manifold without boundary. We say that the geodesic flow $\varphi^t : S^*M \to S^*M$ has the Anosov property if there exists a flow/unstable/stable decomposition of the tangent spaces

$$T_\rho(S^*M) = E_0(\rho) \oplus E_u(\rho) \oplus E_s(\rho), \quad \rho \in S^*M,$$

where E_0 is the one-dimensional space spanned by the generator of the flow, while E_u, E_s depend continuously on ρ, are invariant under the flow φ^t, and satisfy the exponential decay condition for some $\theta > 0$:

$$\left|d\varphi^t(\rho)v\right| \leq Ce^{-\theta|t|}|v|, \qquad \begin{cases} v \in E_u(\rho), & t \leq 0, \\ v \in E_s(\rho), & t \geq 0. \end{cases}$$

A large family of manifolds with Anosov geodesic flows is given by compact Riemannian manifolds of negative sectional curvature, see the book of Anosov [5]. An important special case is given by *hyperbolic surfaces*, which are compact, oriented Riemannian manifolds of dimension 2 with Gauss curvature identically equal to -1. See Figure 3 for a numerical illustration.

The Anosov property implies that the geodesic flow is ergodic with respect to the Liouville measure, so Quantum Ergodicity applies to give that most eigenfunctions equidis-

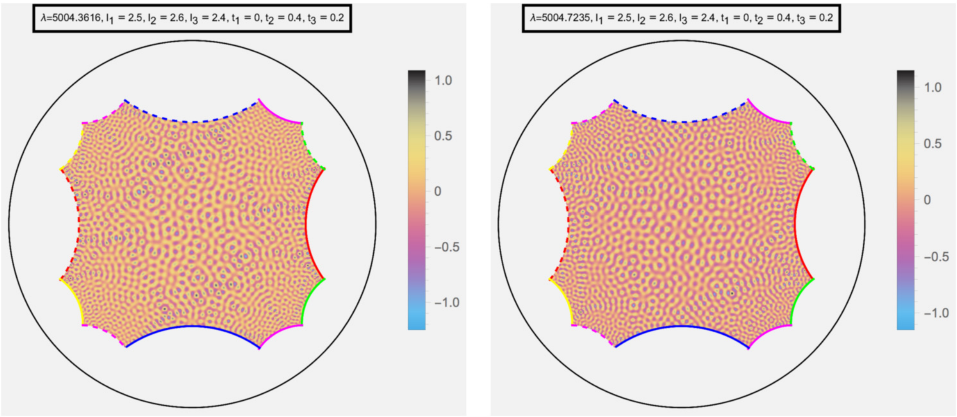


FIGURE 3

Two Laplacian eigenfunctions on a hyperbolic surface, courtesy of Alex Strohmaier (see Strohmaier–Uski [45]). Here we view the surface as a quotient of the hyperbolic plane by a group of isometries, or equivalently as the result of gluing together appropriate sides of the pictured fundamental domain. On a microscopic level the two eigenfunctions look different, but the macroscopic features are the same – both show equidistribution.

tribute. The major open question is the following *Quantum Unique Ergodicity* conjecture which claims equidistribution for the entire sequence of eigenfunctions:

Conjecture 6. *Assume that (M, g) is a compact Riemannian manifold with Anosov geodesic flow. Then μ_L is the only semiclassical measure.*

Conjecture 6 was originally stated by Rudnick–Sarnak [42] in the context of hyperbolic surfaces. It is known in the special case of *arithmetic* hyperbolic surfaces, which have additional symmetries commuting with the Laplacian, called Hecke operators, and we consider a joint basis of eigenfunctions of the Laplacian and a Hecke operator; see Lindenstrauss [36] and Brooks–Lindenstrauss [13]. In general, in spite of significant partial progress described below, the conjecture is open. One of the issues with a potential proof is that Quantum Unique Ergodicity fails in the related setting of quantum cat maps; see Theorem 14 below.

4.1. Entropy bounds

A major step towards Quantum Unique Ergodicity (Conjecture 6) are *entropy bounds*, originating in the work of Anantharaman [1]:

Theorem 7. *Assume that the geodesic flow on (M, g) has the Anosov property. Then any semiclassical measure μ has positive Kolmogorov–Sinai entropy, $\mathbf{h}_{\mathrm{KS}}(\mu) > 0$.*

Here the Kolmogorov–Sinai entropy $\mathbf{h}_{\mathrm{KS}}(\mu)$ is a nonnegative number associated to each flow-invariant measure μ; roughly speaking, it expresses the complexity of the flow from the point of view of that measure, and is one way to measure how "spread out" the measure is—measures which are more concentrated have lower entropy, and measures which are more spread out have higher entropy. Theorem 7 in particular implies the following conjecture of Colin de Verdière [14]:

$$\text{On a hyperbolic surface, no semiclassical measure can be supported on a closed geodesic} \tag{4.1}$$

since the entropy of a measure supported on a closed geodesic is zero.

The lower bound on entropy in Theorem 7 is in general complicated. However, in the case of hyperbolic (i.e., constant negative curvature) manifolds Anantharaman–Nonnenmacher [3] gave the following easy to state bound:

Theorem 8. *Assume that (M, g) is an n-dimensional hyperbolic manifold. Then any semiclassical measure μ satisfies*

$$\mathbf{h}_{\mathrm{KS}}(\mu) \geq \tfrac{n-1}{2}. \tag{4.2}$$

We remark that the Liouville measure in this setting has entropy $n - 1$, so (4.2) in some sense excludes "half" of all invariant measures as possible semiclassical measures. For other entropy(-type) bounds, see the works of Anantharaman–Koch–Nonnenmacher [2], Rivière [39, 40], and Anantharaman–Silberman [4].

The constant in the bound (4.2) matches (in the case of surfaces) the counterexamples for quantum cat maps given in Theorem 14 below. Thus an important milestone on the way to Quantum Unique Ergodicity would be to prove the following:

Conjecture 9. *Let μ be a semiclassical measure on an n-dimensional hyperbolic manifold (M, g). Then $\mathbf{h}_{\mathrm{KS}}(\mu) > \frac{n-1}{2}$.*

We conclude this subsection with another conjecture which would go a long way towards Quantum Unique Ergodicity but does not exclude the counterexample of Theorem 14:

Conjecture 10. *Let μ be a semiclassical measure on a compact manifold (M, g) with Anosov geodesic flow. Then we have $\mu = \alpha\mu_L + (1-\alpha)\mu'$ for some $\alpha \in (0, 1]$, where μ_L is the Liouville measure and μ' is some probability measure on S^*M.*

4.2. Full support property

Another way to characterize how much a measure μ is "spread out" is by looking at its support, $\operatorname{supp}\mu \subset S^*M$. For surfaces with Anosov geodesic flows, Dyatlov–Jin [19] (in the hyperbolic case) and Dyatlov–Jin–Nonnenmacher [20] (in the general case) showed that the support of every semiclassical measure is the entire S^*M:

Theorem 11. *Let μ be a semiclassical measure on a compact surface (M, g) with Anosov geodesic flow. Then $\operatorname{supp}\mu = S^*M$, that is, $\mu(U) > 0$ for every nonempty open set $U \subset S^*M$.*

Theorem 11 and entropy bounds give different restrictions on the set of possible semiclassical measures. On the one hand (assuming (M, g) is a hyperbolic surface for simplicity), the entropy bound (4.2) implies that the Hausdorff dimension of $\operatorname{supp}\mu$ is at least 2, but there exist flow-invariant measures supported on proper subsets of S^*M of dimension arbitrarily close to 3. On the other hand, there exist measures which have full support and small entropy: one can, for example, take a convex combination of the Liouville measure and a measure supported on a closed geodesic.

The key new ingredient in the proof of Theorem 11 is the *fractal uncertainty principle* of Bourgain–Dyatlov [10]. We state the following version appearing in [20]:

Theorem 12. *Let $\nu, h \in (0, 1)$ and assume that $X, Y \subset \mathbb{R}$ are ν-porous up to scale h, namely for any interval $I \subset \mathbb{R}$ of length $|I| \in [h, 1]$, there exists a subinterval $J \subset I$ of length $|J| = \nu|I|$ such that $X \cap J = \emptyset$ (and similarly for Y). Then there exist constants $C, \beta > 0$ depending only on ν such that for all $f \in L^2(\mathbb{R})$,*

$$\operatorname{supp}\hat{f} \subset h^{-1}Y \quad \Rightarrow \quad \|\mathbf{1}_X f\|_{L^2(\mathbb{R})} \le Ch^{\beta}\|f\|_{L^2(\mathbb{R})}. \tag{4.3}$$

One should think of the parameter ν in Theorem 12 as fixed and h as going to 0. The sets X, Y can depend on h as long as they are ν-porous; a basic example is given by $\frac{h}{10}$-neighborhoods of some sets which are porous up to scale 0 (e.g., Cantor sets). The estimate (4.3) can be interpreted as follows: if a function f lives in the (semiclassically rescaled) frequency space in a porous set Y, then only a small part of the L^2-mass of f can concentrate on the porous set X. We refer the reader to the review [15] for more details.

The proof of Theorem 11 can be roughly summarized as follows (restricting to the case of hyperbolic surfaces for simplicity): assume that a sequence of eigenfunctions $\{u_j\}$ converges semiclassically to a measure μ such that $\mu(\mathcal{U}) = 0$ for some nonempty open set $\mathcal{U} \subset S^*M$. Using microlocal methods, one can show that u_j is in a certain sense concentrated on both of the sets

$$\Omega_\pm(h_j) := \{\rho \in S^*M \mid \varphi^{\mp t}(\rho) \notin \mathcal{U} \text{ for all } t \in [0, \log(1/h_j)]\}$$

of geodesics which do not cross the set $\mathcal{U}$ in the future or in the past for time $\log(1/h_j)$. Here one can barely make sense of localization in the position–frequency space on each of the sets $\Omega_\pm(h_j)$, i.e., construct operators $A_\pm$ which localize to these sets and write $u_j = A_+u_j + o(1) = A_-u_j + o(1)$. However, the sets $\Omega_\pm(h)$ have porous structure (see Figure 5 below for the related case of quantum cat maps), and one can use the Fractal Uncertainty Principle to show that $\|A_+A_-\|_{L^2\to L^2} = o(1)$, giving a contradiction. We refer to [15] for a detailed exposition of the proof.

Theorem 11 only applies to surfaces because the Fractal Uncertainty Principle is only known for subsets of $\mathbb{R}$. A naïve generalization of Theorem 12 to higher dimensions is false: for example, the sets

$$X = [0, h/10] \times [0, 1], \quad Y = [0, 1] \times [0, h/10] \subset \mathbb{R}^2$$

are both $\frac{1}{10}$-porous up to scale h (where we replace intervals by balls in the definition of porosity), but they do not satisfy an estimate of type (4.3): the Fourier transform of the indicator function of $h^{-1}Y$ has large L^2 mass on X. (See [16, §6] for a more detailed discussion.) However, this does not translate to a counterexample for semiclassical measures, leaving the door open for the following:

Conjecture 13. *Let μ be a semiclassical measure on a compact manifold (M, g) with Anosov geodesic flow. Then* $\operatorname{supp} \mu = S^*M$.

An analog of Conjecture 13 is known for certain quantum cat maps, see Theorem 16 below.

5. QUANTUM CAT MAPS

We finally discuss *quantum cat maps*, which are toy models in quantum chaos with microlocal properties similar to Laplacians on hyperbolic manifolds (though the extensive research on them demonstrates that they are a "tough toy to crack"). They were originally introduced by Hannay and Berry in [31]. We start with two-dimensional quantum cat maps which are analogous to hyperbolic surfaces. These maps quantize toral automorphisms (a.k.a. "Arnold cat maps")

$$x \mapsto Ax \bmod \mathbb{Z}^2, \quad x \in \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2 \tag{5.1}$$

where $A \in \mathrm{SL}(2,\mathbb{Z})$ is a 2×2 integer matrix with determinant 1. We make the assumption that A is *hyperbolic*, i.e., it has no eigenvalues on the unit circle. A basic example of such a matrix is

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}. \tag{5.2}$$

Quantizations of the map (5.1) are not operators on L^2 of a manifold, instead they are unitary $N \times N$ matrices, where the integer N is related to the semiclassical parameter h as follows:

$$2\pi N h = 1.$$

The semiclassical limit $h \to 0$ studied above now turns into the limit $N \to \infty$.

Before introducing quantizations of cat maps, we briefly discuss the adaptation of the quantization procedure (2.1) to this setting, which has the form

$$a \in C^\infty(\mathbb{T}^2) \quad \mapsto \quad \mathrm{Op}_N(a) : \mathbb{C}^N \to \mathbb{C}^N. \tag{5.3}$$

That is, functions on the 2-torus are quantized to $N \times N$ matrices. The quantization procedure also depends on a twist parameter $\theta \in \mathbb{T}^2$, but we suppress this in the notation. (If N is even, then we can always just take $\theta = 0$ in what follows.) See, for example, [18, §2.2] for more details.

Now, for $A \in \mathrm{SL}(2,\mathbb{Z})$, its quantization is a family of unitary $N \times N$ matrices $B_N : \mathbb{C}^N \to \mathbb{C}^N$ which satisfies the following *exact Egorov's theorem*:

$$B_N^{-1} \mathrm{Op}_N(a) B_N = \mathrm{Op}_N(a \circ A) \quad \text{for all } a \in C^\infty(\mathbb{T}^2). \tag{5.4}$$

Such B_N exists and is unique modulo multiplication by a unit length scalar. The statement (5.4) intertwines conjugation by B_N (corresponding to quantum evolution) with pullback by the map (5.1) (corresponding to classical evolution). It is analogous to Egorov's Theorem for Riemannian manifolds (see, e.g., [49, THEOREM 15.2]), which states that

$$e^{-ith\Delta_g/2} \mathrm{Op}_h(a) e^{ith\Delta_g/2} = \mathrm{Op}_h(a \circ \varphi^t) + \mathcal{O}(h)$$

where the geodesic flow $\varphi^t : S^*M \to S^*M$ is extended to T^*M as the Hamiltonian flow of $|\xi|_g^2/2$. Thus the quantum cat map B_N should be thought of as an analog of the Schrödinger propagator $e^{ith\Delta_g/2}$, eigenfunctions of B_N are analogous to Laplacian eigenfunctions, and the dynamics of the geodesic flow in this setting is replaced by the dynamics of the map (5.1).

Using the quantization (5.3), we can define similarly to (2.9) semiclassical measures associated to sequences of eigenfunctions

$$B_{N_j} u_j = \lambda_j u_j, \quad u_j \in \mathbb{C}^{N_j}, \quad \|u_j\|_{\ell^2} = 1, \quad N_j \to \infty.$$

These are probability measures on $\mathbb{T}^2$ which are invariant under the map (5.1) (as can be seen directly from Egorov's theorem (5.4)).

When the matrix A is hyperbolic, the map (5.2) is ergodic with respect to the Lebesgue measure on $\mathbb{T}^2$. Using this fact, Bouzouina–de Bièvre [11] showed Quantum Ergodicity in this setting: if we put together orthonormal bases of eigenfunctions of B_N

for all N, then there exists a density 1 subsequence of this sequence which converges to the Lebesgue measure.

On the other hand, Faure–Nonnenmacher–De Bièvre [25] showed that Quantum Unique Ergodicity fails for quantum cat maps:

Theorem 14. *Let $A \in \mathrm{SL}(2,\mathbb{Z})$ be a hyperbolic matrix. Fix any periodic trajectory $\gamma \subset \mathbb{T}^2$ of the map (5.1). Then there exists a sequence of eigenfunctions u_j of the quantum cat map B_{N_j}, for some $N_j \to \infty$, which converge semiclassically to the measure*

$$\tfrac{1}{2}\delta_\gamma + \tfrac{1}{2}\mu_L \tag{5.5}$$

where δ_γ is the delta probability measure on the trajectory γ and μ_L is the Lebesgue measure on $\mathbb{T}^2$.

We remark that the choice of N_j in Theorem 14 is highly special: one takes them so that the matrix A^{k_j} is the identity modulo $2N_j$ where k_j is very small, namely $k_j \sim \log N_j$. This implies that the quantum cat map B_{N_j} also has a short period, namely $B_{N_j}^{k_j}$ is a scalar. See the papers of Dyson–Falk [23] and Bonechi–De Bièvre [9] for more information on the periods of the cat map. A numerical illustration of Theorem 14 is given on Figure 4.

The entropy of the measure (5.5) is equal to half the entropy of the Lebesgue measure. This matches the constant in the entropy bound of Theorem 8. Since from the point of view of microlocal analysis quantum cat maps have similar properties to hyperbolic surfaces,

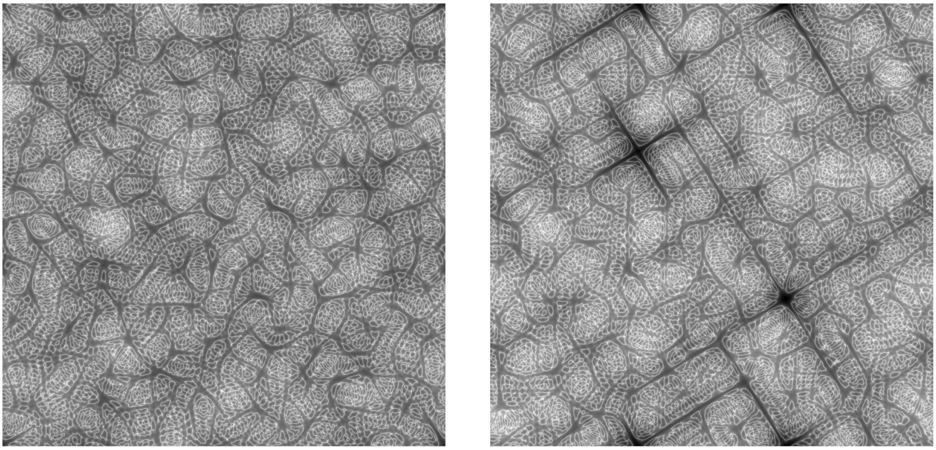


FIGURE 4

Phase space concentration for two eigenfunctions of the quantum cat map with A given by (5.2) and $N = 1292$. More specifically, we plot the absolute value of a smoothened out Wigner transform of the eigenfunction on the logarithmic scale (see, e.g., [18, §2.2.5]). On the left is a typical eigenfunction, showing equidistribution. On the right is a particular eigenfunction of the type constructed in [25], corresponding to a measure of the type (5.5) featuring the closed trajectory $\{(\frac{1}{3},0),(\frac{2}{3},\frac{1}{3}),(\frac{2}{3},0),(\frac{1}{3},\frac{2}{3})\}$. The existence of such an eigenfunction relies on the careful choice of N: A^{18} is the identity matrix modulo $2N$.

significant new insights would be needed to show that a counterexample of the kind (5.5) cannot occur for hyperbolic surfaces.

Faure–Nonnenmacher [24] showed that the constant $\frac{1}{2}$ in (5.5) is sharp: the mass of the pure point part of any semiclassical measure for a quantum cat map is less than or equal to the mass of its Lebesgue part. Brooks [12] generalized this to a statement that the mass of lower entropy components of any semiclassical measure is less than or equal to the mass of higher entropy components; this in particular implies an entropy bound analogous to (4.2).

There is also an analogue of arithmetic Quantum Unique Ergodicity in the setting of cat maps: Kurlberg–Rudnick [35] introduced Hecke operators which commute with B_N and showed that any sequence of joint eigenfunctions of B_N and these operators converges to the Lebesgue measure. This does not contradict the counterexample of Theorem 14 since for the values of N_j chosen there, the map B_{N_j} has eigenvalues of high multiplicity.

We now discuss the recent results on support of semiclassical measures for cat maps, proved using the fractal uncertainty principle. For two-dimensional cat maps, Schwartz [43] showed the following:

Theorem 15. *Let μ be a semiclassical measure for a quantum cat map associated to some hyperbolic matrix $A \in \mathrm{SL}(2,\mathbb{Z})$. Then* $\operatorname{supp}\mu = \mathbb{T}^2$.

Similarly to Section 4.2, the proof uses that no function can be localized simultaneously on the two sets

$$\Omega_{\pm}(N) := \left\{\rho \in \mathbb{T}^2 \;\middle|\; A^{\mp j}(\rho) \notin \mathcal{U} \text{ for all } j = 0, \dots, \frac{\log N}{\log|\lambda_+|}\right\}$$

where λ_+ is the eigenvalue of A such that $|\lambda_+| > 1$. Here $\mathcal{U} \subset \mathbb{T}^2$ is some nonempty open set. See Figure 5.

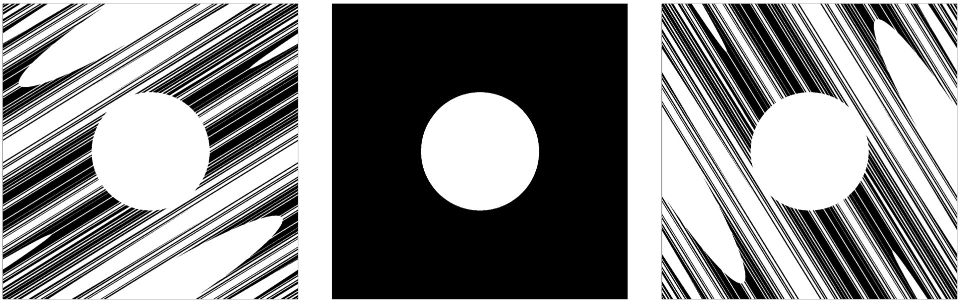


FIGURE 5

A set $\mathcal{U} \subset \mathbb{T}^2$ (center picture, in white) and the corresponding sets $\Omega_+(N)$, $\Omega_-(N)$ (left/right picture). The set $\Omega_+(N)$ is "smooth" in the unstable direction of the matrix A and porous in the stable direction, with the porosity constant depending only on $\mathcal{U}$. Same is true for $\Omega_-(N)$ but switching the roles of the stable/unstable directions. The fractal uncertainty principle of Theorem 12 can be used to show that no function can be localized on both $\Omega_+(N)$ and $\Omega_-(N)$.

We finally discuss the quantum cat map analog of the higher-dimensional Conjecture 13, by considering quantum cat maps associated to symplectic integer matrices $A \in \mathrm{Sp}(2n, \mathbb{Z})$. In this setting Dyatlov–Jézéquel [18] proved

Theorem 16. *Let μ be a semiclassical measure for a quantum cat map associated to a matrix $A \in \mathrm{Sp}(2n, \mathbb{Z})$ such that:*

- *A has a simple eigenvalue λ_+ such that all other eigenvalues satisfy $|\lambda| < \lambda_+$; and*
- *the characteristic polynomial of A is irreducible over the rationals.*

Then $\operatorname{supp} \mu = \mathbb{T}^{2n}$.

Here the first condition makes it possible to still use the one-dimensional Fractal Uncertainty Principle in the proof.

We remark that there are examples of semiclassical measures which do not have full support for some matrices A satisfying the first condition of Theorem 16 but not the second condition. In particular, there exist semiclassical measures supported on tori associated to any A-invariant rational Lagrangian subspace of $\mathbb{R}^{2n}$. See the work of Kelmer [34] and the discussion in [18, APPENDIX A].

FUNDING

This work was partially supported by NSF CAREER grant DMS-1749858 and a Sloan Research Fellowship.

REFERENCES

[1] N. Anantharaman, Entropy and the localization of eigenfunctions. *Ann. of Math. (2)* **168** (2008), no. 2, 435–475.

[2] N. Anantharaman, H. Koch, and S. Nonnenmacher, Entropy of eigenfunctions. In *New trends in mathematical physics*, edited by V. Sidoravičius, pp. 1–22, Springer, Netherlands, Dordrecht, 2009.

[3] N. Anantharaman and S. Nonnenmacher, Half-delocalization of eigenfunctions for the Laplacian on an Anosov manifold. *Ann. Inst. Fourier (Grenoble)* **57** (2007), no. 7, 2465–2523.

[4] N. Anantharaman and L. Silberman, A Haar component for quantum limits on locally symmetric spaces. *Israel J. Math.* **195** (2013), no. 1, 393–447.

[5] D. V. Anosov, *Geodesic flows on closed Riemann manifolds with negative curvature*. Proc. Steklov Inst. Math. 90 (1967), American Mathematical Society, Providence, RI, 1969.

[6] V. Arnaiz and F. Macià, Localization and delocalization of eigenmodes of harmonic oscillators. 2020, arXiv:2010.13436.

[7] A. Barnett, Asymptotic rate of quantum ergodicity in chaotic Euclidean billiards. *Comm. Pure Appl. Math.* **59** (2006), no. 10, 1457–1488.

[8] A. Barnett and A. Hassell, Fast computation of high-frequency Dirichlet eigenmodes via spectral flow of the interior Neumann-to-Dirichlet map. *Comm. Pure Appl. Math.* **67** (2014), no. 3, 351–407.

[9] F. Bonechi and S. De Bièvre, Exponential mixing and $|\ln \hbar|$ time scales in quantized hyperbolic maps on the torus. *Comm. Math. Phys.* **211** (2000), no. 3, 659–686.

[10] J. Bourgain and S. Dyatlov, Spectral gaps without the pressure condition. *Ann. of Math. (2)* **187** (2018), no. 3, 825–867.

[11] A. Bouzouina and S. De Bièvre, Equipartition of the eigenfunctions of quantized ergodic maps on the torus. *Comm. Math. Phys.* **178** (1996), no. 1, 83–105.

[12] S. Brooks, On the entropy of quantum limits for 2-dimensional cat maps. *Comm. Math. Phys.* **293** (2010), no. 1, 231–255.

[13] S. Brooks and E. Lindenstrauss, Joint quasimodes, positive entropy, and quantum unique ergodicity. *Invent. Math.* **198** (2014), no. 1, 219–259.

[14] Y. Colin de Verdière, Ergodicité et fonctions propres du laplacien. In *Bony–Sjöstrand–Meyer seminar, 1984–1985*, pp. Exp. No. 13, 8, École Polytech, Palaiseau, 1985.

[15] S. Dyatlov, Control of eigenfunctions on hyperbolic surfaces: an application of fractal uncertainty principle. *Journ. Equ. Dériv. Partielles*, exposé 4, 14 pages (2017).

[16] S. Dyatlov, An introduction to fractal uncertainty principle. *J. Math. Phys.* **60** (2019), no. 8, 081505, 31.

[17] S. Dyatlov, Around quantum ergodicity. *Ann. Math. Québec* (2021), https://link.springer.com/article/10.1007%2Fs40316-021-00165-7.

[18] S. Dyatlov and M. Jézéquel, Semiclassical measures for higher dimensional quantum cat maps. 2021, arXiv:2108.10463.

[19] S. Dyatlov and L. Jin, Semiclassical measures on hyperbolic surfaces have full support. *Acta Math.* **220** (2018), no. 2, 297–339.

[20] S. Dyatlov, L. Jin, and S. Nonnenmacher, Control of eigenfunctions on surfaces of variable curvature. *J. Amer. Math. Soc.* (2021). https://www.ams.org/journals/jams/0000-000-00/S0894-0347-2021-00979-7/.

[21] S. Dyatlov and J. Zahl, Spectral gaps, additive energy, and a fractal uncertainty principle. *Geom. Funct. Anal.* **26** (2016), no. 4, 1011–1094.

[22] S. Dyatlov and M. Zworski, *Mathematical theory of scattering resonances*. Grad. Stud. Math. 200, American Mathematical Society, Providence, RI, 2019.

[23] F. J. Dyson and H. Falk, Period of a discrete cat mapping. *Amer. Math. Monthly* **99** (1992), no. 7, 603–614.

[24] F. Faure and S. Nonnenmacher, On the maximal scarring for quantum cat map eigenstates. *Comm. Math. Phys.* **245** (2004), no. 1, 201–214.

[25] F. Faure, S. Nonnenmacher, and S. De Bièvre, Scarred eigenstates for quantum cat maps of minimal periods. *Comm. Math. Phys.* **239** (2003), no. 3, 449–492.

[26] J. Galkowski, Quantum ergodicity for a class of mixed systems. *J. Spectr. Theory* **4** (2014), no. 1, 65–85.

[27] P. Gérard, Mesures semi-classiques et ondes de Bloch. In *Séminaire sur les Équations aux Dérivées Partielles, 1990–1991*, pp. Exp. No. XVI, 19, École Polytech, Palaiseau, 1991.

[28] P. Gérard and E. Leichtnam, Ergodic properties of eigenfunctions for the Dirichlet problem. *Duke Math. J.* **71** (1993), no. 2, 559–607.

[29] S. Gomes, Quantum ergodicity in mixed and KAM hamiltonian systems. 2017, arXiv:1709.09919.

[30] S. P. Gomes, Percival's conjecture for the Bunimovich mushroom billiard. *Nonlinearity* **31** (2018), no. 9, 4108–4136.

[31] J. Hannay and M. Berry, Quantization of linear maps on a torus-Fresnel diffraction by a periodic grating. *Phys. D* **1** (1980), no. 3, 267–290.

[32] A. Hassell, Ergodic billiards that are not quantum unique ergodic. *Ann. of Math. (2)* **171** (2010), no. 1, 605–619.

[33] D. Jakobson and S. Zelditch, Classical limits of eigenfunctions for some completely integrable systems. In *Emerging applications of number theory (Minneapolis, MN, 1996)*, pp. 329–354, IMA Vol. Math. Appl. 109, Springer, New York, 1999.

[34] D. Kelmer, Arithmetic quantum unique ergodicity for symplectic linear maps of the multidimensional torus. *Ann. of Math. (2)* **171** (2010), no. 2, 815–879.

[35] P. Kurlberg and Z. Rudnick, Hecke theory and equidistribution for the quantization of linear maps of the torus. *Duke Math. J.* **103** (2000), no. 1, 47–77.

[36] E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)* **163** (2006), no. 1, 165–219.

[37] P.-L. Lions and T. Paul, Sur les mesures de Wigner. *Rev. Mat. Iberoam.* **9** (1993), no. 3, 553–618.

[38] A. Logunov and E. Malinnikova, Review of Yau's conjecture on zero sets of laplace eigenfunctions. 2019, arXiv:1908.01639.

[39] G. Rivière, Entropy of semiclassical measures for nonpositively curved surfaces. *Ann. Henri Poincaré* **11** (2010), no. 6, 1085–1116.

[40] G. Rivière, Entropy of semiclassical measures in dimension 2. *Duke Math. J.* **155** (2010), no. 2, 271–336.

[41] G. Rivière, Remarks on quantum ergodicity. *J. Mod. Dyn.* **7** (2013), no. 1, 119–133.

[42] Z. Rudnick and P. Sarnak, The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Comm. Math. Phys.* **161** (1994), no. 1, 195–213.

[43] N. Schwartz, The full delocalization of eigenstates for the quantized cat map. 2021, arXiv:2103.06633.

[44] A. Shnirelman, Ergodic properties of eigenfunctions. *Uspekhi Mat. Nauk* **29** (1974), no. 6(180), 181–182.

[45] A. Strohmaier and V. Uski, An algorithm for the computation of eigenvalues, spectral zeta functions and zeta-determinants on hyperbolic surfaces. *Comm. Math. Phys.* **317** (2013), no. 3, 827–869.

[46] E. Studnia, Quantum limits for harmonic oscillator. 2020, arXiv:1905.07763.

[47] S. Zelditch, Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Math. J.* **55** (1987), no. 4, 919–941.

[48] S. Zelditch and M. Zworski, Ergodicity of eigenfunctions for ergodic billiards. *Comm. Math. Phys.* **175** (1996), no. 3, 673–682.

[49] M. Zworski, *Semiclassical analysis*. Grad. Stud. Math. 138, American Mathematical Society, Providence, RI, 2012.

SEMYON DYATLOV

Massachusetts Institute of Technology, Cambridge, MA 02139, USA,
dyatlov@math.mit.edu

VARIATIONAL HOMOGENIZATION: OLD AND NEW

RITA FERREIRA, IRENE FONSECA, AND RAGHAVENDRA VENKATRAMAN

ABSTRACT

This note is a summary of the contributions of the authors to the variational viewpoint on homogenization. After providing a broad context, two recent projects are discussed in detail: one concerning the large-scale behavior of quasicrystals, and the other involving phase transitions in periodically heterogeneous media.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 35B27; Secondary 35B25, 49S05, 49-06

KEYWORDS

Calculus of variations, periodic and quasiperiodic homogenization, phase transitions, PDE constrained optimization

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3724–3755
DOI 10.4171/ICM2022/129

1. INTRODUCTION

Homogenization, a subject with a long and rich history, deals with the macrobehavior of a medium as a large-scale average of its microscopic properties. The earliest investigations seeking such effective models, appear to go back to Maxwell [76], Lord Rayleigh [84], and others, around the start of the 20th century. For instance, in [84] Lord Rayleigh considers an arrangement of cylindrical rods of constant thermal conductivity in a rectangular array within an otherwise uniform medium. Assuming that the conductivity of the rods is significantly different from that of the background medium, the subject of homogenization addresses questions such as: on length-scales much larger than the period of the arrangement of the rods, can one approximate the heat distribution in the composite material, by instead studying an effective, homogeneous material? Remarkably, in [84] Lord Rayleigh discovers an explicit formula for the effective conductivity in the case of the above planar arrangement.

The study of homogenization has witnessed immense growth in the last half century, and continues to flourish. As it supplies tools for analysis of situations that involve multiple spatio-temporal scales, it is not surprising that homogenization plays an important role in such diverse fields as materials science [2, 88], fluid mechanics and mixing [54], climate modeling [35], biology [15, 16, 68], machine learning and data science [91]. The ubiquity of homogenization, on the one hand, and the intractability of direct computational approaches for large, multiscale problems, on the other, renders the analytical study of homogenization vitally important. The goal of this survey is to report progress, and the state-of-the-art, in one segment of this vast subject, focusing on the contributions of the authors to variational methods in homogenization. In particular, we do not discuss the recent burst of activity in stochastic homogenization [8, 69], applications of homogenization to study discrete and possibly random structures such as point clouds [89], optimal control theory and numerical analysis associated with homogenization [91].

The main thrust of this article is on variational methods. As a concrete example, we consider the benchmark problem in homogenization

$$\begin{cases} -\nabla\cdot\left(a\left(\dfrac{x}{\varepsilon}\right)\nabla u_\varepsilon\right)=0 & \text{in } \Omega\,, \\ u_\varepsilon = g & \text{on } \partial\Omega\,. \end{cases} \tag{1.1}$$

Here, $\Omega \subset \mathbb{R}^N$ is a bounded Lipschitz domain, $a : \mathbb{R}^N \to (0,\infty)$ is a given periodic, measurable, bounded, uniformly elliptic, symmetric matrix field, $0 < \varepsilon \ll 1$ represents the length-scale of the heterogeneities, and $g \in L^2(\partial\Omega)$ is a given Dirichlet datum. Homogenization seeks to find an "effective" constant matrix $\overline{a}$ that is independent of the domain Ω and of the boundary condition g, such that the limit of solutions $\{u_\varepsilon\}_\varepsilon$ to (1.1) exists (call it u_0), and solves the "homogenized" partial differential equation (PDE)

$$\begin{cases} -\nabla\cdot\overline{a}\nabla u = 0 & \text{in } \Omega\,, \\ u = g & \text{on } \partial\Omega\,. \end{cases} \tag{1.2}$$

It is, of course, of interest to also study quasiperiodic, or random choices of a. Early works that addressed the question of justification of the formal two-scale asymptotic expansion led to the development of important functional analytic tools that rely on the structure of the PDE. These include [14], methods of compensation compactness [87], G- and H-convergence [48], Bloch decomposition [34], among others. When the matrix field a is symmetric, the problem (1.1) has a variational formulation. Indeed, solutions to (1.1) are the unique minimizers to the sequence of variational problems

$$\min_{u|_{\partial\Omega}=g} E_\varepsilon(u) := \frac{1}{2}\int_\Omega \left\langle a\left(\frac{x}{\varepsilon}\right)\nabla u \cdot \nabla u\right\rangle dx. \tag{1.3}$$

The notion of Γ-convergence is well suited for the study of the $\varepsilon \to 0^+$ asymptotics of the energies E_ε in (1.3). This notion of convergence of a family of functionals defined on a Banach space was introduced by De Giorgi in 1975 (see [49]). As such, along with appropriate compactness, this scheme of convergence of functionals is the weakest notion that ensures that global minimizers of the approximating functional converge to a global minimizer of the limiting functional. In the example in (1.3) above, the limiting energy takes the form

$$E_0(u) := \frac{1}{2}\int_\Omega \langle \overline{a}\nabla u, \nabla u\rangle\, dx.$$

Γ-convergence is stable under continuous perturbations, and is therefore well adapted to the multiscale analysis of nonlinear problems that have variational structure. More crucially, it is sufficiently robust to allow for the limiting problem to be defined on a different space than the approximating problems (see Section 4 for an example). Being based on soft compactness and lower-semicontinuity arguments, approaches based on Γ-convergence are particularly well-suited when fine information that is uniform in the small parameter (such as a spectral gap) is difficult or even impossible to obtain. There is, however, a price to pay using Γ-convergence techniques in that the underlying arguments do not often yield rates of convergence.

2. AN OVERVIEW OF CONTRIBUTIONS TO HOMOGENIZATION

In [60], Fonseca and Francfort consider a quasistatic model aiming at understanding the interaction between damage and fracture. To prove that a certain incremental problem at a fixed time step is well posed, they state and use a homogenization conjecture (see [60, CONJECTURE 3.15]), known at the time to be true in some important convex examples (cf. [60, REMARK 3.16]). This conjecture was then proved to be true in the general convex case in [10], while the nonconvex case, including the quasiconvex one, remains open.

In [20], Fonseca, Bouchitté, and Mascarenhas introduced the so-called global method for relaxation, which is central to the study of minimization problems via the direct method in the calculus of variations. This method provides a unified pathway to identify the integral representation of the lower-semicontinuous envelope of certain functionals that naturally arise in several applications, such as in phase transitions, fracture mechanics, plasticity, and image segmentation. Moreover, as an application of their methodology, they address in [20, SECTION 4.3] a homogenization problem associated with integral energies coupling bulk and surface terms, which generalized several results in the literature, including [23].

In [24], Fonseca, Braides, and Francfort study dimension reduction problems for heterogeneous thin domains in the context of nonlinear elasticity. The domains considered are of the type

$$\Omega_\varepsilon := \big\{(x', x_3) \in \mathbb{R}^2 \times \mathbb{R} : x' \in \omega,\ |x_3| < \varepsilon h_\varepsilon(x')\big\},$$

where $\omega \subset \mathbb{R}^2$ is a bounded domain and h_ε is a smooth ε-dependent profile, while the elastic integral energy of the system involves a p-growth Carathéodory function $f_\varepsilon \equiv f_\varepsilon(x', x_3; \xi)$. As one of the main applications of their general asymptotic analysis, they consider the homogenization problem corresponding to the case where the profile h_ε is assumed to be periodic, and the elastic density f_ε is assumed to be independent of x_3 and periodic with respect to x', with the same period as h_ε. They obtain an integral representation for the effective energy on the middle section ω.

Another contribution to the study of minimization problems via the direct method in the calculus of variations is that of Fonseca and Müller in [63], where they address the study of lower semicontinuity and relaxation of functionals of the type

$$(u, v) \mapsto \int_\Omega f\big(x, u(x), v(x)\big)\, \mathrm{d}x,$$

where, for N, m, $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^N$ is an open and bounded domain, $u : \Omega \to \mathbb{R}^m$, and $v : \Omega \to \mathbb{R}^d$ satisfies a partial differential constraint of the type $\mathcal{A}v = 0$. Here, $\mathcal{A}$ is a constant-coefficient linear partial differential operator of the form

$$\mathcal{A}v := \sum_{i=1}^{N} A^{(i)} \frac{\partial v}{\partial x_i} \quad \text{with } A^{(i)} \in \mathbb{R}^{l\times d} \text{ for all } i \in \{1, \dots, N\} \text{ and some } l \in \mathbb{N} \tag{2.1}$$

(see Section 3 for a more detailed description of these operators). In the literature, this context is nowadays referred as the $\mathcal{A}$-free setting. A typical example of such operators is $\mathcal{A} = \operatorname{curl}$, in which case $v = \nabla w$ for some potential w. In particular, for $w = u$ we are led to the so-called gradient case, where the integral energies take the form

$$u \mapsto \int_\Omega f\big(x, u(x), \nabla u(x)\big)\, \mathrm{d}x.$$

Though relevant in many applications, the curl case does not cover some important ones in which v must satisfy other linear partial differential constraints, such as Maxwell's equations in the case of electromagnetism, or, in the case of linear elasticity, v is the symmetric part of a gradient. Therefore, the $\mathcal{A}$-free fields setting offers a unified abstract approach to several of these PDE constraints.

In [25], besides further developing the analysis in [63], Fonseca, Braides, and Leoni address an homogenization problem in the $\mathcal{A}$-free setting. More precisely, they characterize the effective behavior of integrals energies of the form

$$v \mapsto \int_\Omega f\left(\frac{x}{\varepsilon}, v(x)\right) \mathrm{d}x \quad \text{subjected to } \mathcal{A}v = 0,$$

where $\varepsilon > 0$ is the usual homogenization small parameter, and the integrand f is periodic in the first variable and satisfies certain continuity, p-growth, and coercivity conditions.

The periodic homogenization result in [25] was generalized by Fonseca and Krömer [61] by working under weaker continuity assumptions and, most importantly, without assuming coercivity on f. Moreover, they extended the widely used two-scale convergence method (see [1, 82]) to the $\mathcal{A}$-free setting.

Also in the context of periodic homogenization in the general $\mathcal{A}$-free framework, Fonseca and Davoli consider in [42, 43] operators with variable coefficients, which is not a straightforward extension of the constant coefficient case. More precisely, these two papers are devoted to the study of the effective behavior, as $\varepsilon \to 0$, of integral energies of the form

$$v \mapsto \int_{\Omega} f\left(x, \frac{x}{\varepsilon^{\alpha}}, v(x)\right) \mathrm{d}x, \tag{2.2}$$

subject to periodically oscillating differential constraints of the type

$$\mathcal{A}_{\varepsilon} v := \sum_{i=1}^{N} A^{i}\left(\frac{\cdot}{\varepsilon^{\beta}}\right) \frac{\partial v}{\partial x_i} \to 0 \quad \text{strongly in } W^{-1,p}(\Omega; \mathbb{R}^{l}) \tag{2.3}$$

or, in divergence form,

$$\mathcal{A}_{\varepsilon} v := \sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left(A^{i}\left(\frac{\cdot}{\varepsilon^{\beta}}\right) v\right) \to 0 \quad \text{strongly in } W^{-1,p}(\Omega; \mathbb{R}^{l}), \tag{2.4}$$

where $p \in (1, +\infty)$, $A^{i}(x) \in \mathbb{R}^{l \times d}$ for all $x \in \mathbb{R}^{N}$ and $i \in \{1, \dots, N\}$, $\alpha, \beta > 0$ are parameters, and f is assumed periodic in the second variable. Different asymptotic regimes are expected according to the ratio between α and β. The case in which $\beta > 0$ and $\alpha = 0$ with f independent of the first two variables ($f(x, y, \xi) \equiv f(\xi)$) is addressed in [43] under the $\mathcal{A}$-constraint (2.4). Also, Fonseca and Davoli consider in [43] the $\alpha > 0$ and $\beta > 0$ case under the $\mathcal{A}$-constraint (2.3). The remaining cases are announced in [42, 43] to be treated in forthcoming works.

In [59], Fonseca, Ferreira, and Venkatraman initiated a similar research project to that of Fonseca and Davoli [42, 43] but, in contrast with the works mentioned above, outside of the periodic setting. In a nutshell, [59] addresses the effective behavior, as $\varepsilon \to 0$, of integral energies as in (2.2), with $\mathcal{A}$ as in (2.1), assuming a quasicrystalline assumption on the second variable of f in place of periodicity, which poses new challenges. We refer to Section 3 for a more detailed motivation and description of this work.

Next, we mention some authors' contributions concerning the gradient case, $\mathcal{A} = \operatorname{curl}$, or related cases. In [66], Fonseca and Zappale consider first and second order-derivatives in the multiscale case aimed at composites that may feature periodic properties at more than one microscale. The integral energies are of the form

$$u \mapsto \int_{\Omega} f\left(\frac{x}{\varepsilon}, \frac{x}{\varepsilon^{2}}, D^{s} u(x)\right) \mathrm{d}x,$$

where $s \in \{1, 2\}$, and f is assumed to be convex in the last variable and continuous. Besides considering general convex energies in the multiscale setting, one of the main novelties of [66] is the characterization of multiscale limits of second-order derivatives. Prior to [66], this characterization was only known for first-order derivatives.

Later, Fonseca and Baía [11] address the effective behavior, as $\varepsilon \to 0$, of integral energies of the form

$$u \mapsto \int_\Omega f\left(x, \frac{x}{\varepsilon}, \nabla u(x)\right) \mathrm{d}x$$

without assuming any convexity-type condition on f. This work extends those in the literature by not requiring uniform continuity in space.

In [57,58], Fonseca and Ferreira revisit the multiscale framework in the case where f grows at most linearly. These studies fall within the realm of the space of functions of bounded variation, BV, and are aimed at identifying effective energies for composite materials in the presence of fracture or cracks. Precisely, they generalize in [58] the notion of two-scale convergence for sequences of Radon measures with finite total variation in [3] to the case of multiple periodic length scales of oscillations. The main result concerns the characterization of the multiscale limit of $\{(u_\varepsilon \mathcal{L}^N{}_{\lfloor\Omega}, Du_{\varepsilon\lfloor\Omega})\}_\varepsilon \subset \mathcal{M}(\Omega;\mathbb{R}^d) \times \mathcal{M}(\Omega;\mathbb{R}^{d\times N})$ whenever $\{u_\varepsilon\}_\varepsilon$ is a bounded sequence in $\mathrm{BV}(\Omega;\mathbb{R}^d)$, where $\mathcal{M}(\Omega;\mathbb{R}^m)$ with $m \in \mathbb{N}$ is the Banach space of bounded Radon $\mathbb{R}^m$-valued measures, endowed with the total variation norm $|\cdot|$. This result requires considerable modifications of the single microscale case treated in [3], and is based on fine analytical and measure-theoretic arguments. Using this characterization, Fonseca and Ferreira treat in [57] multiscale homogenized problems in the space BV of functions of bounded variation of the form

$$\begin{aligned} u \mapsto &\int_\Omega f\left(\frac{x}{\varrho_1(\varepsilon)}, \dots, \frac{x}{\varrho_n(\varepsilon)}, \nabla u(x)\right) \mathrm{d}x \\ &+ \int_\Omega f^\infty\left(\frac{x}{\varrho_1(\varepsilon)}, \dots, \frac{x}{\varrho_n(\varepsilon)}, \frac{\mathrm{d}D^s u}{\mathrm{d}|D^s u|}(x)\right) \mathrm{d}\left|D^s u\right|(x) \end{aligned}$$

for $u \in \mathrm{BV}(\Omega;\mathbb{R}^d)$. Here, the distributional derivative of u, Du, is decomposed into its absolutely continuous part with respect to the N-dimensional Lebesgue measure, $\nabla u \mathcal{L}^N{}_{\lfloor\Omega}$, and its singular part, $D^s u$. Moreover, $f^\infty(y_1, \dots, y_n, \xi) := \limsup_{t\to\infty} f(y_1, \dots, y_n, t\xi)/t$ is the recession function of a function $f : \mathbb{R}^{nN} \times \mathbb{R}^{d\times N} \to \mathbb{R}$, separately periodic in the first n variables, and $\varrho_1, \dots, \varrho_n$ are positive functions on $(0,\infty)$, representing the length-scales, such that for all $i \in \{1, \dots, n\}$ and $j \in \{2, \dots, n\}$, $\lim_{\varepsilon\to 0} \varrho_i(\varepsilon) = 0$, $\lim_{\varepsilon\to 0} \varrho_j(\varepsilon)/\varrho_{j-1}(\varepsilon) = 0$. In the case of one microscale, Fonseca and Ferreira recover the result in [3] under more general conditions, as well as the results in [19,45]. For two or more microscales, they obtain new results in the literature.

In [28], Fonseca and Bufford extend to $L^1(\Omega)$ the paramount two-scale compactness property, which asserts that from every bounded sequence, one can extract a subsequence that two-scale converges, with the average over the periodic cell coinciding with the usual weak two-scale limit. This L^1-extension is obtained under an equiintegrability condition on the sequence, and is proved in [28] using three different approaches: an adaptation of the L^p case with $p > 1$, a measure-theoretic argument, and the periodic-unfolding method.

In [27], Fonseca, Bufford, and Davoli address a multiscale homogenization problem in the context of dimension reduction in nonlinear elasticity, aiming at characterizing effective energies for thin, elastic plate-type composites. The energies considered are of the form

$$u \mapsto \frac{1}{h} \int_{\Omega_h} f\left(\frac{x'}{\varepsilon(h)}, \frac{x'}{\varepsilon^2(h)}, \nabla u(x)\right) \mathrm{d}x =: F_h(u)$$

for $u \in W^{1,2}(\Omega_h; \mathbb{R}^3)$, where $\Omega_h := \omega \times (-\frac{h}{2}, \frac{h}{2}) \subset \mathbb{R}^2 \times \mathbb{R}$, $x = (x', x_3) \in \omega \times (-\frac{h}{2}, \frac{h}{2})$, $h > 0$, f is periodic in its first two arguments, and satisfies both common assumptions in nonlinear elasticity and a nondegeneracy condition in a neighborhood of the set of proper rotations. The main result in [27] concerns the characterization of the effective energy associated with the rescaled energies $\frac{1}{h^2} F_h(\cdot)$ depending on the values of

$$\gamma_1 := \lim_{h \to 0} \frac{h}{\varepsilon(h)} \quad \text{and} \quad \gamma_2 := \lim_{h \to 0} \frac{h}{\varepsilon^2(h)},$$

where $\lim_{h\to 0} \varepsilon(h) = \lim_{h \to 0} \varepsilon^2(h) = 0$. This rescaling of the energies corresponds to Kirchhoff's nonlinear bending theory for plates, and the values of γ_1 and γ_2 represent the relative ratios between the thickness parameter h and the two homogenization length-scales, ε and ε^2. These authors obtain different limit models depending on these ratios. Their results extend those in [72, 90] to the multiscale case, and a key and nontrivial step in [27] is the characterization of the three-scale limit of the sequence of linearized elastic stresses. Indeed, the presence of three scales increases the technicality of the problem in all scaling regimes.

Very recently, in [36, 37], Fonseca, Cristoferi, Hagerty, and Popovici study a variational model for fluid–fluid phase transitions with small scale heterogeneities in the case where the small heterogeneities are of the same order of the scale governing the phase transition, and characterized by a small parameter $\varepsilon > 0$. The main result is the limit behavior, as $\varepsilon \to 0$, of integral energies of the form

$$u \mapsto \int_{\Omega} \left[\frac{1}{\varepsilon} W\left(\frac{x}{\varepsilon}, u(x)\right) + \varepsilon \left|\nabla u(x)\right|^2 \right] \mathrm{d}x,$$

where $W : \mathbb{R}^N \times \mathbb{R}^d \to [0, +\infty)$ is a double-well potential that is periodic in the first variable and has two zeros. This limit behavior is given not by an isotropic interfacial energy as one might expect given the isotropy of the surface energy penalization, $\varepsilon|\nabla u|^2$, but instead it has an anisotropic interfacial energy. This anisotropy results from the intricate interaction between homogenization and the phase transitions, and is encoded in the limit cell problem. In [32], the authors study fine properties on the minimizers of the family of problems defining the asymptotic cell formula obtained in [36, 37]. They also obtain bounds for the limiting anisotropic surface tension in terms of the large-scale behavior of the distance function to hyperplanes in certain periodic Riemannian metrics. This work, along with a discussion of [36], is the content of Section 5.2.

3. HOMOGENIZATION OF QUASICRYSTALLINE FUNCTIONALS VIA TWO-SCALE-CUT-AND-PROJECT CONVERGENCE

The work [59] addresses a homogenization problem aimed at understanding composites with a quasicrystalline microstructure. Such composites have been playing a central role in materials science and other areas of engineering [9,18,53,70,73,74,81,93]; for example, Al–Cu–Fe quasicrystalline materials in polymer-based composites have significantly shown to improve wear-resistance to volume loss, and a two-fold increase in the elastic moduli. The 2011 Nobel Prize in Chemistry was awarded to Dan Shechtman for the striking discovery of quasicrystals, which was announced in the early 1980s.

A key feature of a quasicrystalline structure is that its properties are ordered but are neither periodic nor random. In particular, the mathematical study of quasicrystalline composites does not fit within the *classical* periodic homogenization theory, while almost-periodic and stochastic homogenization approaches do not take full advantage of the quasi-crystalline feature of the problem, often leading to asymptotic formulas that pose computational difficulties and are not stable under perturbations. Instead, in [59], a homogenization procedure based on the two-scale-cut-and-project convergence, introduced in [21] and recently revisited in [92], is adopted and further developed. This two-scale-cut-and-project homogenization procedure leads to a more tractable (even if higher-dimensional) cell problem.

To describe the problem and the results in [59], we first recall the cut-and-project method to model quasicrystals. This method was introduced by de Bruijn [44] and further developed by Duneau and Katz [52], and extends Penrose's ideas of aperiodic tilings of the plane [83] to higher dimensions (also see [21]). Roughly speaking, we can model an N-dimensional quasicrystalline patterns by cutting periodic tilings in an m-dimensional space, with $m > N$, through an N-dimensional subspace with irrational slope. To be precise, given an N-dimensional quasicrystal R and representing by $\sigma_R : \mathbb{R}^n \to \mathbb{R}$ a constitutive property of R, we can find $m \in \mathbb{N}$, with $m > N$, a Y^m-periodic function $\sigma : \mathbb{R}^m \to \mathbb{R}$ with $Y^m \subset \mathbb{R}^m$ a parallelotope, and a linear map $\boldsymbol{R} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\sigma_R(x) = \sigma(\boldsymbol{R}x). \tag{3.1}$$

In the homogenization literature, the structural condition (3.1) is referred to as quasi-periodicity [30,75]. We refer to [21,59] for relevant examples of such linear maps $\boldsymbol{R}$. Here, and in the sequel, we do not distinguish the linear map from its associated matrix in $\mathbb{R}^{m\times N}$, and denote both by $\boldsymbol{R}$. Also, we do not distinguish between the transpose matrix and the adjoint of $\boldsymbol{R}$, and denote both by $\boldsymbol{R}^*$.

In general, there are multiple choices for m, σ, and $\boldsymbol{R}$ (see [21]). However, the homogenization analysis in this cut-and-project setting does not depend on $\boldsymbol{R}$ provided it satisfies the following diophantine condition

$$\boldsymbol{R}^* k \neq 0 \quad \text{for all } k \in \mathbb{Z}^m \backslash \{0\}, \tag{3.2}$$

where $\boldsymbol{R}^*$ denotes the transpose of $\boldsymbol{R}$. This condition implies that some entries of $\boldsymbol{R}$ must be irrational, justifying the expression irrational slope used above.

In [59], we address the homogenization problem of characterizing the asymptotic behavior, as $\varepsilon \to 0^+$, of integral energies of the form

$$F_\varepsilon(u) := \int_\Omega f_R\left(x, \frac{x}{\varepsilon}, u(x)\right) \mathrm{d}x \tag{3.3}$$

for $u \in L^p(\Omega;\mathbb{R}^d)$ satisfying $\mathcal{A}u = 0$, where $p \in (1,\infty)$ and

$$\mathcal{A}u := \sum_{i=1}^{N} A^{(i)} \frac{\partial u}{\partial x_i} \quad \text{with } A^{(i)} \in \mathbb{R}^{l\times d} \text{ for all } i \in \{1,\dots,N\}.$$

The precise meaning of the preceding condition $\mathcal{A}u = 0$, in which case we say that u is $\mathcal{A}$-free, is by duality, i.e.,

$$\int_\Omega u \cdot \mathcal{A}^*\phi \,\mathrm{d}x = 0$$

for all $\phi \in C_c^1(\Omega;\mathbb{R}^l)$, where $\mathcal{A}^*$ is the formal adjoint of $\mathcal{A}$ with $\mathcal{A}^*\phi := -\sum_{i=1}^{N}(A^{(i)})^T \frac{\partial\phi}{\partial x_i}$.

As usual within studies involving $\mathcal{A}$-free vector fields, we assume that $\mathcal{A}$ satisfies the constant-rank property [63,80,87]; that is, there exists $r \in \mathbb{N}$ such that for all $w \in \mathbb{R}^n \setminus \{0\}$, we have

$$\operatorname{rank}\mathbb{A}(w) = r, \tag{3.4}$$

where $\mathbb{A} : \mathbb{R}^n \to \mathbb{R}^{l\times d}$ denotes the symbol of $\mathcal{A}$, and is defined by $\mathbb{A}(w) := \sum_{i=1}^{N} A^{(i)} w_i$ for $w \in \mathbb{R}^n$.

A key step to study the asymptotic behavior of the integral energies in (3.3) via the two-scale-cut-and-project convergence is the characterization of the two-scale-cut-and-project limits (or, for brevity, $\boldsymbol{R}$-two-scale limits) associated with L^p-bounded sequences of $\mathcal{A}$-free vector fields. As we mentioned before, this method has the benefit of taking full advantage of the quasicrystalline feature of the problem and, in contrast with the random homogenization case, leads to a simple and more tractable cell formula (see (3.13) below). Before stating our main homogenization result associated with the integral energies in (3.3), Theorem 3.9 below, we revise the main definitions and results regarding the cut-and-project-two-scale convergence obtained in [59], which are of interest on their own.

The notion of $\boldsymbol{R}$-two-scale convergence was introduced in [21] (also see [92]) as an extension of the usual notion of two-scale convergence [1,82] to enable the study of composites whose underlying microstructure has a quasicrystalline feature. In [21,92], the authors consider sequences in L^2 and their arguments are based on Fourier analysis, relying heavily on Parseval's and Plancherel's identities. Also, in [21] the authors characterize the $\boldsymbol{R}$-two-scale limit of bounded sequences in $W^{1,2}$, while in [92] the authors characterize the limit associated with bounded sequences in L^2 that are divergence-free or curl-free. In [59], besides generalizing these results to the more general setting of L^p with $p \in (1,\infty)$, we provide a unified approach to all the previous cases by considering bounded sequences in L^p that are $\mathcal{A}$-free, in the spirit of [61] concerning the periodic case.

We start by introduction the definition of $\boldsymbol{R}$-two-scale convergence. In what follows, we assume that ε takes values on an arbitrary sequence of positive numbers that converges to zero. Moreover, we use the subscript # within function spaces to highlight an underlying periodicity, in which case the domain indicates the periodicity cell. For instance, $C_\#(Y^m) = \{u \in C(\mathbb{R}^m) : u \text{ is } Y^m\text{-periodic}\}$ and, for a parallelotope in $\mathbb{R}^n$, $\Pi \subset \mathbb{R}^n$, $L^p_\#(\Pi) = \{u \in L^p_{\mathrm{loc}}(\mathbb{R}^n) : u \text{ is } \Pi\text{-periodic}\}$. Also, given a Lebesgue measurable set $B \subset \mathbb{R}^k$, with $k \in \mathbb{N}$, we use the average notation $\fint_B \cdot$ in place of $\frac{1}{\mathcal{L}^k(B)}\int_B \cdot$, where $\mathcal{L}^k(B)$ denotes the k-dimensional Lebesgue measure of B.

Definition 3.1 (***R***-two-scale convergence). A sequence $\{u_\varepsilon\}_\varepsilon \subset L^p(\Omega;\mathbb{R}^k)$ is said to ***R***-two-scale converge to a function $u \in L^p(\Omega \times Y^m;\mathbb{R}^k)$, and we write $u_\varepsilon \xrightarrow{\boldsymbol{R}\text{-}2sc} u$ if for all $\varphi \in L^{p'}(\Omega; C_\#(Y^m;\mathbb{R}^k))$ we have

$$\lim_{\varepsilon\to 0^+} \int_\Omega u_\varepsilon(x)\cdot\varphi\left(x, \frac{\boldsymbol{R}x}{\varepsilon}\right) \mathrm{d}x = \int_\Omega \fint_{Y^m} u(x,y)\cdot\varphi(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{3.5}$$

The next proposition states some basic properties of ***R***-two-scale convergence in $L^p(\Omega;\mathbb{R}^k)$, and we refer to [59, REMARKS 3.2 AND 3.3 AND PROPOSITIONS 3.4 AND 3.5] for its proof.

Proposition 3.2. *Let $\{u_\varepsilon\}_\varepsilon \subset L^p(\Omega;\mathbb{R}^k)$, $u \in L^p(\Omega\times Y^m;\mathbb{R}^k)$, and $\bar{u} \in L^p(\Omega;\mathbb{R}^k)$. Then,*

(i) *(uniqueness of **R**-two-scale limits) There exists at most a function $\tilde{u} \in L^p(\Omega\times Y^m;\mathbb{R}^k)$ such that $u_\varepsilon \xrightarrow{\boldsymbol{R}\text{-}2sc} \tilde{u}$.*

(ii) *(on the test functions) If $\{u_\varepsilon\}_\varepsilon$ is bounded in $L^p(\Omega;\mathbb{R}^k)$, then $u_\varepsilon \xrightarrow{\boldsymbol{R}\text{-}2sc} u$ if and only if (3.5) holds for all $\varphi \in C_c^\infty(\Omega; C_\#^\infty(Y^m;\mathbb{R}^k))$.*

(iii) *(**R**-two-scale and weak limits) If $u_\varepsilon \xrightarrow{\boldsymbol{R}\text{-}2sc} u$, then $u_\varepsilon \rightharpoonup \bar{u}_0$ weakly in $L^p(\Omega;\mathbb{R}^k)$, where $\bar{u}_0(\cdot) := \fint_{Y^m} u(\cdot,y)\,\mathrm{d}y$. In particular, $\{u_\varepsilon\}_\varepsilon$ is bounded in $L^p(\Omega;\mathbb{R}^k)$.*

(iv) *(**R**-two-scale and strong limits) If $u_\varepsilon \to \bar{u}$ in $L^p(\Omega;\mathbb{R}^k)$, then $u_\varepsilon \xrightarrow{\boldsymbol{R}\text{-}2sc} \bar{u}$.*

The next proposition provides an important example of sequences that ***R***-two-scale converge, which is at the core of several homogenization results using the ***R***-two-scale convergence. In particular, it is used to prove the compactness property with respect to the ***R***-two-scale convergence stated in Proposition 3.4.

Proposition 3.3. *Let $\psi \in L^1(\Omega; C_\#(Y^m;\mathbb{R}^k))$, and assume that **R** satisfies (3.2). Then $\{\psi(\cdot, \frac{\boldsymbol{R}\cdot}{\varepsilon})\}_\varepsilon$ is an equiintegrable sequence in $L^1(\Omega;\mathbb{R}^k)$ such that*

$$\left\| \psi\left(\cdot, \frac{\boldsymbol{R}\cdot}{\varepsilon}\right)\right\|_{L^1(\Omega;\mathbb{R}^k)} \leqslant \|\psi\|_{L^1(\Omega;C_\#(Y^m;\mathbb{R}^k))} = \int_\Omega \sup_{y\in Y^m} |\psi(x,y)|\,\mathrm{d}x \tag{3.6}$$

and

$$\lim_{\varepsilon\to 0^+} \int_\Omega \psi\left(x, \frac{\boldsymbol{R}x}{\varepsilon}\right)\mathrm{d}x = \int_\Omega \fint_{Y^m} \psi(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{3.7}$$

*In particular, if $\psi \in L^p(\Omega; C_\#(Y^m;\mathbb{R}^k))$, then $\{\psi(\cdot, \frac{\boldsymbol{R}\cdot}{\varepsilon})\}_\varepsilon$ is a p-equiintegrable sequence in $L^p(\Omega;\mathbb{R}^k)$ that **R**-two-scale converges to ψ.*

The proof of Proposition 3.3 can be found in [59, PROPOSITION 3.7 AND COROLLARY 3.8], while the proof of the following compactness result can be found in [59, PROPOSITION 3.9].

Proposition 3.4. *Let $\{u_\varepsilon\}_\varepsilon \subset L^p(\Omega;\mathbb{R}^k)$ be a bounded sequence, and assume that **R** satisfies (3.2). Then, there exist a subsequence $\varepsilon' \preceq \varepsilon$ and a function $u \in L^p(\Omega\times Y^m;\mathbb{R}^k)$ such that $u_{\varepsilon'} \xrightarrow{\boldsymbol{R}\text{-}2sc} u$.*

As shown in [21, Remark 2.8], this compactness property may fail in the case in which $\boldsymbol{R}$ does not satisfy (3.2).

To characterize the $\boldsymbol{R}$-two-scale limits associated with L^p-bounded sequences of $\mathcal{A}$-free vector fields, we recall below the notion of $(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*})$-free vector fields introduced in [59] (see [59, Definition 3.7 and Remark 3.6]).

Definition 3.5 ($(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*})$-free fields). Let $w \in L^p(\Omega; L^p_\#(Y^m; \mathbb{R}^d))$, and define $\bar{w}_0 \in L^p(\Omega; \mathbb{R}^d)$ and $\bar{w}_1 \in L^p(\Omega; L^p_\#(Y^m; \mathbb{R}^d))$ by setting $\bar{w}_0 := ⨍_{Y^m} w(\cdot, y)\, \mathrm{d}y$ and $\bar{w}_1 := w - \bar{w}_0$. We say that w is $(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*})$-free if the two following conditions hold:

$$\text{(i)} \quad \text{for all } \phi \in C^1_c(\Omega; \mathbb{R}^l), \text{ we have } \int_\Omega w_0 \cdot \mathcal{A}^*\phi\, \mathrm{d}x = 0, \tag{3.8}$$

$$\text{(ii)} \quad \text{for a.e. } x \in \Omega \text{ and for all } \psi \in C^1_\#(Y^m; \mathbb{R}^l),$$

$$\text{we have } \int_{Y^m} \bar{w}_1(x, y) \cdot \mathcal{A}^*_{\boldsymbol{R}} \psi(y)\, \mathrm{d}y = 0, \tag{3.9}$$

where

$$\mathcal{A}^* := -\sum_{i=1}^N (A^{(i)})^T \frac{\partial}{\partial x_i} \quad \text{and} \quad \mathcal{A}^*_{\boldsymbol{R}} := -\sum_{i=1}^N \sum_{m=1}^m (A^{(i)})^T \boldsymbol{R}_{mi} \frac{\partial}{\partial y_m}.$$

For brevity, we write $\mathcal{A}\bar{w}_0 = 0$ and $\mathcal{A}^y_{\boldsymbol{R}^*}\bar{w}_1 = 0$ to mean (i) and (ii), respectively.

Next, we state our main result regarding the characterization of the limits of bounded sequences in L^p that are $\mathcal{A}$-free.

Theorem 3.6. *Let $\boldsymbol{R} \in \mathbb{R}^{m\times n}$ satisfy (3.2). A function $u \in L^p(\Omega \times Y^m; \mathbb{R}^d)$ is the $\boldsymbol{R}$-two-scale limit of an $\mathcal{A}$-free sequence $\{u_\varepsilon\}_\varepsilon \subset L^p(\Omega; \mathbb{R}^d)$ if and only if u is $(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*})$-free in the sense of Definition 3.5, that is,*

$$\mathcal{A}\bar{u}_0 = 0 \quad \textit{and} \quad \mathcal{A}^y_{\boldsymbol{R}^*}\bar{u}_1 = 0 \tag{3.10}$$

in the sense of (3.8) and (3.9), respectively, where $\bar{u}_0 := ⨍_{Y^m} u(\cdot, y)\, \mathrm{d}y$ and $\bar{u}_1 := u - \bar{u}_0$.

The proof of Theorem 3.6 in [59] uses similar arguments to those in [61] concerning the periodic case (see [61, Theorem 2.12]). The sufficient part in Theorem 3.6, which guarantees that (3.10) fully characterizes the $\boldsymbol{R}$-two-scale limits, is new in the literature even for $p = 2$ and $\mathcal{A} :=$ curl or $\mathcal{A} :=$ div treated in [21,92]. Furthermore, in [59, Section 5], we give an alternative proof of Theorem 3.6 for the $\mathcal{A} :=$ curl case using arguments based on Fourier analysis that differ from those in [21,92] because Parseval's and Plancherel's identities do not hold for $p \neq 2$. This alternative proof provides the equivalent alterative characterization for the $\boldsymbol{R}$-two-scale limit of bounded sequences in $W^{1,p}$ in Theorem 3.7 below, and we believe it provides useful arguments to study homogenization problems involving quasicrystalline functionals in the $\mathcal{A} :=$ curl case.

Theorem 3.7. *Let $\boldsymbol{R} \in \mathbb{R}^{m\times n}$ satisfy (3.2) and let $Y^m \subset \mathbb{R}^m$ be a parallelotope. Then, a function $v \in L^p(\Omega \times Y^m; \mathbb{R}^n)$ is the $\boldsymbol{R}$-two-scale limit of a sequence $\{\nabla v_\varepsilon\}_\varepsilon$ with $\{v_\varepsilon\}_\varepsilon$ bounded in $W^{1,p}(\Omega)$ if and only if there exist $v_0 \in W^{1,p}(\Omega)$ and $v_1 \in L^p(\Omega; \mathcal{G}^p_{\boldsymbol{R}})$ such that*

$$v = \nabla v_0 + v_1,$$

where

$$\mathcal{G}^p_{\boldsymbol{R}} := \big\{ w \in L^p_\#(Y^m;\mathbb{R}^n) : \hat{w}_k = \lambda_k \boldsymbol{R}^* k \text{ for some } \{\lambda_k\}_{k\in\mathbb{Z}^m} \subset \mathbb{C} \text{ with } \lambda_0 = 0 \big\} \quad (3.11)$$

with $\hat{w}_k := \fint_{Y^m} w(y) e^{-2\pi i k\cdot y}\,\mathrm{d}y$, $k \in \mathbb{Z}^m$, *denoting the Fourier coefficients of* w.

Remark 3.8. We recall that if $u_\varepsilon \in L^p(\Omega;\mathbb{R}^n)$ is curl-free in $\mathbb{R}^n$ with Ω simply connected, then there exists $v_\varepsilon \in W^{1,p}(\Omega)$ such that $u_\varepsilon = \nabla v_\varepsilon$. Thus, in terms of the notations in the two previous results with $d = N$, we have $\bar{u}_0 = \nabla v_0$ and $\bar{u}_1 = v_1$. In particular, (3.11) provides an alternative characterization of $\mathcal{A}_{\boldsymbol{R}^*}$- and $\mathcal{A}^y_{\boldsymbol{R}^*}$-free vector fields (see Definition 3.5) in the $\mathcal{A} :=$ curl case (also see [59, REMARK 5.7] for a more detailed analysis).

Finally, we state the main homogenization result in [59] associated with the integral energies in (3.3), proved under the following assumptions on the Lagrangian, $f_R : \Omega \times \mathbb{R}^n \times \mathbb{R}^d \to [0,\infty)$:

(H1) (quasicrystallinity) there exist $m \in \mathbb{N}$, with $m > N$, a matrix $\boldsymbol{R} \in \mathbb{R}^{m\times n}$ satisfying (3.2), and a continuous function $f : \Omega \times \mathbb{R}^m \times \mathbb{R}^d \to [0,\infty)$ such that the function $f(x,\cdot,\xi)$ is Y^m-periodic for each $(x,\xi) \in \Omega \times \mathbb{R}^d$, with Y^m denoting a parallelotope in $\mathbb{R}^m$, and

$$f_R(x,z,\xi) = f(x,\boldsymbol{R}z,\xi)$$

for all $(x,z,\xi) \in \Omega \times \mathbb{R}^n \times \mathbb{R}^d$.

(H2) (growth) there exist $p \in (1,\infty)$ and $C > 0$ such that

$$0 \leqslant f_R(x,z,\xi) \leqslant C\big(1 + |\xi|^p\big)$$

for all $(x,z,\xi) \in \Omega \times \mathbb{R}^n \times \mathbb{R}^d$.

For the proof in [59] of the Γ-liminf inequality in Theorem 3.9 below, we require, in addition,

(H3) (convexity) for all $(x,y) \in \Omega \times \mathbb{R}^m$, the function $\xi \mapsto f(x,y,\xi)$ is convex and C^1.

Theorem 3.9. *Let $\Omega \subset \mathbb{R}^n$ be an open and bounded set, let $f_R : \Omega \times \mathbb{R}^n \times \mathbb{R}^d \to [0,\infty)$ be a function satisfying (H1)–(H3), let F_ε be the functional introduced in* (3.3)*, and assume that* (3.4) *holds. Then, the sequence $\{F_\varepsilon\}_\varepsilon$ Γ-converges on $\mathcal{U}_{\mathcal{A}} := \{u \in L^p(\Omega;\mathbb{R}^d) : \mathcal{A}u = 0\}$ as $\varepsilon \to 0^+$, with respect to the weak topology in $L^p(\Omega;\mathbb{R}^d)$, to the functional $\mathcal{F}_{\mathrm{hom}}$ defined, for $u \in \mathcal{U}_{\mathcal{A}}$, by*

$$\mathcal{F}_{\mathrm{hom}}(u) := \inf_{w\in\mathcal{W}_{\mathcal{A}}} \int_\Omega \fint_{Y^m} f\big(x,y,u(x)+w(x,y)\big)\,\mathrm{d}x\,\mathrm{d}y,$$

where

$$\mathcal{W}_{\mathcal{A}} := \Big\{ w \in L^p\big(\Omega; L^p_\#\big(Y^m;\mathbb{R}^d\big)\big) : w \text{ is } \big(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*}\big)\text{-free in the sense of Definition 3.5}, \\ \text{with} \int_{Y^m} w(\cdot,y)\,\mathrm{d}y = 0 \Big\}. \quad (3.12)$$

Precisely, given an arbitrary sequence $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^+$ *converging to* 0*, the following pair of statements holds:*

(1) *(*Γ*-liminf inequality) Let* $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}_{\mathcal{A}}$ *be a sequence such that* $u_n \rightharpoonup u$ *in* $L^p(\Omega;\mathbb{R}^d)$ *for some* $u \in L^p(\Omega;\mathbb{R}^d)$*. Then,* $u \in \mathcal{U}_{\mathcal{A}}$ *and*

$$\liminf_{n\to\infty} F_{\varepsilon_n}(u_n) \geqslant \mathcal{F}_{\hom}(u).$$

(2) *(recovery sequence) For every* $u \in \mathcal{U}_{\mathcal{A}}$*, there exists sequence* $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}_{\mathcal{A}}$ *such that* $u_n \rightharpoonup u$ *in* $L^p(\Omega;\mathbb{R}^d)$ *and*

$$\limsup_{n\to\infty} F_{\varepsilon_n}(u_n) \leqslant \mathcal{F}_{\hom}(u).$$

Moreover, for all $u \in \mathcal{U}_{\mathcal{A}}$*, we have*

$$\mathcal{F}_{\hom}(u) = \int_\Omega f_{\hom}\big(x, u(x)\big)\,\mathrm{d}x,$$

where

$$f_{\hom}(x,\xi) := \inf_{v\in\mathcal{V}_{\mathcal{A}}} \fint_{Y^m} f\big(x, y, \xi + v(y)\big)\,\mathrm{d}y \tag{3.13}$$

with

$$\mathcal{V}_{\mathcal{A}} := \left\{ v \in L^p_\#(Y^m;\mathbb{R}^d) : v \text{ is } \mathcal{A}_{\boldsymbol{R}^*}\text{-free in the sense of (3.9) and } \int_{Y^m} v(y)\,\mathrm{d}y = 0 \right\}. \tag{3.14}$$

Remark 3.10 (On the hypotheses of Theorem 3.9, cf. [59, REMARK 1.2]). (i) In the homogenization literature, measurability of f with respect to the fast-variable is often preferred over continuity. As we discuss in [59, SECTION 2], measurability of f_R requires, in general, Borel-measurability of f. A common approach to deal with lack of continuity is to combine periodicity with Scorza–Dragoni's-type results that, up to a set of small measure, allow reducing the problem to the continuity setting. Here, however, we cannot use such an argument because a set of small m-dimensional Lebesgue measure, the ambient space for the fast variable in terms of (the periodic function) f, may not have small N-dimensional Lebesgue, the ambient space for the fast variable in terms of (the quasicrystalline function) $f_{\boldsymbol{R}}$. (ii) The nonconvex case raises nontrivial difficulties in the quasicrystalline setting, and will be the subject of a forthcoming work. (iii) In the Sobolev setting, homogenization of integral energies of the form (3.3) under *nonperiodic* assumptions was undertaken in [22, 41, 71] in the $\mathcal{A} :=$ curl case, assuming coercivity. Within the quasicrystalline framework, Theorem 3.9 extends these results to the general $\mathcal{A}$-free setting and without coercivity.

The proof in [59] of Theorem 3.9, which we sketch next, is based on Γ-convergence and on two-scale convergence adapted to the quasicrystalline setting, also called two-scale-cut-and-project convergence.

Proof of Theorem 3.9. We refer to [59] for a detailed proof of the assertions in Theorem 3.9. Here, we only present a sketch of the proof. Let $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^+$ be an arbitrary sequence converging to 0.

Step 1. Fix $u \in \mathcal{U}_{\mathcal{A}}$ and assume that $w \in \mathcal{W}_{\mathcal{A}} \cap C^1(\overline{\Omega}; C^1_\#(Y^m;\mathbb{R}^d))$. For $(x,y) \in \Omega \times Y^m$, define

$$\psi(x,y) := f\big(x, y, u(x) + w(x,y)\big).$$

Using (H1), (H2), the continuity of f, and the regularity of w, we conclude that $\psi \in L^1(\Omega; C_\#(Y^m))$. Then, by Proposition 3.3, we have

$$\lim_{n\to\infty} \int_\Omega f_R\bigg(x, \frac{x}{\varepsilon_n}, w_n(x)\bigg)\,\mathrm{d}x = \int_\Omega \fint_{Y^m} f\big(x, y, u(x) + w(x,y)\big)\,\mathrm{d}y\,\mathrm{d}x, \tag{3.15}$$

where, for $x \in \Omega$,

$$w_n(x) := u(x) + w\bigg(x, \frac{\boldsymbol{R}x}{\varepsilon_n}\bigg).$$

It can be checked that

$$\begin{aligned}&\{w_n\}_{n\in\mathbb{N}} \quad \text{is a } p\text{-equiintegrable sequence in } L^p(\Omega;\mathbb{R}^d),\\ &w_n \rightharpoonup u \text{ weakly in } L^p(\Omega;\mathbb{R}^d), \quad \mathcal{A}w_n \to 0 \text{ in } W^{-1,p}(\Omega;\mathbb{R}^l).\end{aligned}$$

Then, using an $\mathcal{A}$-free periodic extension lemma established in [63, LEMMA 2.15] (also see [61, LEMMA 2.8] and [59, LEMMA 2.3]), we can find a sequence $\{u_n\}_{n\in\mathbb{N}} \subset L^p(\Omega;\mathbb{R}^d)$ such that

$$\{u_n\}_{n\in\mathbb{N}} \text{ is } p\text{-equiintegrable}, \quad \mathcal{A}u_n = 0 \text{ in } L^p(\Omega;\mathbb{R}^l), \quad u_n - w_n \to 0 \text{ in } L^p(\Omega;\mathbb{R}^d). \tag{3.16}$$

In particular, $u_n \rightharpoonup u$ weakly in $L^p(\Omega;\mathbb{R}^d)$. Moreover, from (3.15) and a continuity-type result for f_R under (3.16) proved in [59, LEMMA 4.2], we have

$$\begin{aligned}\lim_{n\to\infty} \int_\Omega f_R\bigg(x, \frac{x}{\varepsilon_n}, u_n(x)\bigg)\,\mathrm{d}x &= \lim_{n\to\infty} \int_\Omega f_R\bigg(x, \frac{x}{\varepsilon_n}, w_n(x)\bigg)\,\mathrm{d}x\\ &= \int_\Omega \fint_{Y^m} f\big(x, y, u(x) + w(x,y)\big)\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

Using the preceding arguments and a density argument, we can show that for each $\delta > 0$, $u \in \mathcal{U}_{\mathcal{A}}$, and $w \in \mathcal{W}_{\mathcal{A}}$, there exists a sequence $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}_{\mathcal{A}}$ such that $u_n \rightharpoonup u$ weakly in $L^p(\Omega;\mathbb{R}^d)$ as $n \to \infty$, and

$$\lim_{n\to\infty} \int_\Omega f_R\bigg(x, \frac{x}{\varepsilon_n}, u_n(x)\bigg)\,\mathrm{d}x \leqslant \int_\Omega \fint_{Y^m} f\big(x, y, u(x) + w(x,y)\big)\,\mathrm{d}y\,\mathrm{d}x + \delta. \tag{3.17}$$

Hence, taking the infimum over $w \in \mathcal{W}_{\mathcal{A}}$ first, and then letting $\delta \to 0$ in (3.17), we get

$$\Gamma\text{-}\limsup_{n\to\infty} F_{\varepsilon_n}(u) \leqslant \mathcal{F}_{\mathrm{hom}}(u),$$

where

$$\begin{aligned}\Gamma\text{-}\limsup_{n\to\infty} F_{\varepsilon_n}(u) := \inf\Big\{ \limsup_{n\to\infty} F_{\varepsilon_n}(u_\varepsilon) : u_n \rightharpoonup u \text{ in } L^p(\Omega;\mathbb{R}^d) \text{ as } n \to \infty,\\ \mathcal{A}u_n = 0 \text{ for all } n \in \mathbb{N}\Big\}.\end{aligned}$$

Step 2. Here, we prove the Γ-liminf inequality. Let $\{u_n\}_{n\in\mathbb{N}} \subset \mathcal{U}_{\mathcal{A}}$ be a sequence such that $u_n \rightharpoonup u$ in $L^p(\Omega;\mathbb{R}^d)$ for some $u \in L^p(\Omega;\mathbb{R}^d)$.

Because $u_n \in \mathcal{U}_{\mathcal{A}}$ for all $n \in \mathbb{N}$ and convergence $u_n \rightharpoonup u$ in $L^p(\Omega;\mathbb{R}^d)$, we have $u \in \mathcal{U}_{\mathcal{A}}$. Moreover, by the sufficient part in Theorem 3.6 and by Proposition 3.2, we have

$u_n \xrightarrow{\boldsymbol{R}\text{-}2sc} v$ for a vector-field v that is $(\mathcal{A}, \mathcal{A}^y_{\boldsymbol{R}^*})$-free in the sense of Definition 3.5, with $\int_{Y^m} v(\cdot, y)\,\mathrm{d}y = u(\cdot)$. In particular, we have the decomposition

$$v = u + v_1, \quad v_1 \in L^p\big(\Omega; L^p_\#(Y^m;\mathbb{R}^d)\big), \quad \mathcal{A}^y_{\boldsymbol{R}^*} v_1 = 0, \quad \int_{Y^m} v_1(\cdot, y)\,\mathrm{d}y = 0.$$

Let $\{\psi_j\}_{j\in\mathbb{N}} \subset C_c(\Omega; C_\#(Y^m; \mathbb{R}^l))$ be a sequence converging to v in $L^p(\Omega \times Y^m; \mathbb{R}^d)$ and pointwise in $\Omega \times Y^m$. By (H3), we have, for all $n,\ j \in \mathbb{N}$,

$$f\left(x, \frac{\boldsymbol{R}x}{\varepsilon_n}, u_n(x)\right) \geqslant f\left(x, \frac{\boldsymbol{R}x}{\varepsilon_n}, \psi_j\left(x, \frac{\boldsymbol{R}x}{\varepsilon_n}\right)\right) + \frac{\partial f}{\partial \xi}\left(x, \frac{\boldsymbol{R}x}{\varepsilon}, \psi_j\left(x, \frac{\boldsymbol{R}x}{\varepsilon}\right)\right) \cdot \left(u_n(x) - \psi_j\left(x, \frac{\boldsymbol{R}x}{\varepsilon}\right)\right).$$

Integrating this estimate over Ω and passing to the limit as $n \to \infty$, Proposition 3.3 and (H2)–(H3) yield

$$\begin{aligned}\liminf_{n\to\infty} F_{\varepsilon_n}(u_n) &= \liminf_{n\to\infty} \int_\Omega f\left(x, \frac{\boldsymbol{R}x}{\varepsilon_n}, u_n(x)\right)\mathrm{d}x \\ &\geqslant \int_\Omega \fint_{Y^m} f\big(x, y, \psi_j(x, y)\big)\,\mathrm{d}x\,\mathrm{d}y \\ &\quad + \int_\Omega \fint_{Y^m} \frac{\partial f}{\partial \xi}\big(x, y, \psi_j(x, y)\big) \cdot \big(v(x, y) - \psi_j(x, y)\big)\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{3.18}$$

for all $j \in \mathbb{N}$. Letting $j \to \infty$ in this inequality, we obtain from Fatou's lemma and (H1) that

$$\begin{aligned}\liminf_{n\to\infty} F_{\varepsilon_n}(u_n) &\geqslant \int_\Omega \fint_{Y^m} f\big(x, y, v(x, y)\big)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_\Omega \fint_{Y^m} f\big(x, y, u(x) + v_1(x, y)\big)\,\mathrm{d}y\,\mathrm{d}x \\ &\geqslant \inf_{w\in\mathcal{W}_\mathcal{A}} \int_\Omega \fint_{Y^m} f\big(x, y, u(x) + w(x, y)\big)\,\mathrm{d}x\,\mathrm{d}y = \mathcal{F}_{\mathrm{hom}}(u).\end{aligned}$$

Step 3. From Steps 1 and 2, we conclude that for all $u \in \mathcal{U}_\mathcal{A}$, we have

$$\mathcal{F}_{\mathrm{hom}}(u) = \Gamma\text{-}\liminf_{n\to\infty} F_{\varepsilon_n}(u) = \Gamma\text{-}\limsup_{n\to\infty} F_{\varepsilon_n}(u), \tag{3.19}$$

where

$$\Gamma\text{-}\liminf_{n\to\infty} F_{\varepsilon_n}(u) := \inf\Big\{\liminf_{n\to\infty} F_{\varepsilon_n}(u_n) : u_n \rightharpoonup u \text{ in } L^p(\Omega;\mathbb{R}^d) \text{ as } n \to \infty,\ \mathcal{A}u_n = 0 \text{ for all } n \in \mathbb{N}\Big\}.$$

Formula (3.19) asserts that $\{F_\varepsilon\}_\varepsilon$ Γ-converges as $\varepsilon \to 0^+$, with respect to the weak topology in $L^p(\Omega;\mathbb{R}^d)$, to $\mathcal{F}_{\mathrm{hom}}$ on $\mathcal{U}_\mathcal{A}$, and is equivalent to proving that both the Γ-liminf inequality and the recovery sequence properties in Theorem 3.9 hold (see [39]).

Step 4. Fix $u \in \mathcal{U}_\mathcal{A}$, and let $w \in \mathcal{W}_\mathcal{A}$. It can be checked that

$$x \in \Omega \mapsto f_{\mathrm{hom}}\big(x, u(x)\big) \tag{3.20}$$

is a measurable map. Moreover, for a.e. $x \in \Omega$, we have $w(x, \cdot) \in \mathcal{V}_\mathcal{A}$. Thus, for a.e. $x \in \Omega$,

$$\inf_{v\in\mathcal{V}_\mathcal{A}} \fint_{Y^m} f\big(x, y, u(x) + v(y)\big)\,\mathrm{d}y \leqslant \fint_{Y^m} f\big(x, y, u(x) + w(x, y)\big)\,\mathrm{d}y.$$

Integrating this estimate over Ω, and then taking the infimum over $w \in \mathcal{W}_{\mathcal{A}}$, we conclude that

$$\int_\Omega f_{\mathrm{hom}}\big(x, u(x)\big)\,\mathrm{d}x \leqslant \mathcal{F}_{\mathrm{hom}}(u).$$

The proof of the converse inequality makes use of a measurable selection criterion proved in [61, LEMMA 3.10] (also see [31]), and we refer to [59, PROPOSITION 4.6] for the details. ■

4. PHASE TRANSITIONS IN HETEROGENEOUS MEDIA

Heterogeneous media abound in nature, ranging from biological tissues [68] to geological formations [4]. An essential thermodynamic feature of such systems is phase transitions. The presence of heterogeneities during phase transformations is, in general, expected to lead to complex interactions such as pinning and depinning phenomena of interfacial structures, and stick–slip behaviors for possibly anisotropic interface motion [17]. In [36], Fonseca, Cristoferi, Hagerty, and Popovici initiate a project to understand the interaction of the dynamics of phase transitions with heterogeneities. Further progress is made in [32], and the goal of this section is to outline these developments.

The study of pattern formation in equilibrium configurations phase separation is an extremely complex phenomenon, which has attracted the interest of many mathematicians. In the case of homogeneous substances, variational models such as the Modica–Mortola functional (see [78,79,86]) and its vectorial (see [12,65]), anisotropic (see [13,64]), and nonisothermal variants (see [38]), have been proven capable of describing the stable configurations observed in experiments. For composite materials, it has been realized experimentally (see [17]) that the microscopic scale heterogeneities can affect the macroscopic equilibrium configurations, as well as the dynamics of interfaces. Therefore, physics requires the mathematical models to include these microscopic effects.

In this paper, we consider a variational approach to the study of phase transitions in heterogeneous media in the case where the scale of the heterogeneities is the same as those at which the phase transitions phenomenon takes place. In particular, we study a Modica–Mortola like phase field model where the heterogeneities are modeled by oscillations in the potential. To be precise, let $d, N \geqslant 1$, fix an open bounded set $\Omega \subset \mathbb{R}^N$ with Lipschitz boundary and, for $\varepsilon > 0$, define the energy $\mathcal{F}_\varepsilon : H^1(\Omega; \mathbb{R}^d) \to [0, \infty]$ as

$$\mathcal{F}_\varepsilon(u) := \int_\Omega \left[\frac{1}{\varepsilon} W\left(\frac{x}{\varepsilon}, u(x)\right) + \varepsilon \left|\nabla u(x)\right|^2 \right] \mathrm{d}x. \tag{4.1}$$

Here $u \in H^1(\Omega; \mathbb{R}^d)$ represents the phase field variable. The assumptions that the double-well potential $W : \mathbb{R}^N \times \mathbb{R}^d \to [0, \infty)$ has to satisfy differ according to the questions addressed, and therefore we will present them in each section.

We are interested in understanding what is the sharp interface limit as the parameter $\varepsilon \to 0$. Local minimizers of this limit under a mass constraint will describe equilibrium configurations.

Previous investigations on models related to the one considered in this paper have been undertaken by several authors. In particular, in [6] (see also [5]) Ansini, Braides, and Chiadò Piat considered the case where oscillations are in the forcing term $f(\nabla u)$ (which generalizes $|\nabla u|^2$), while in [50] and [51] by Dirr, Lucia, and Novaga investigated the interaction of the fluid with a periodic mean zero external field. Moreover, in [26], Braides and Zeppieri studied the Γ expansion of the scalar one-dimensional case, allowing the zeros of the potential to jump in a specific way. Finally, the case of higher-order derivatives is examined in [67] by Francfort and Müller.

5. PHASE FIELD MODEL

In this section, we present the results obtained in [32,33,36,37].

5.1. Sharp interface limit

In order to study the sharp interface limit of the energy (4.1), we assume that the double-well potential $W : \mathbb{R}^N \times \mathbb{R}^d \to [0, \infty)$ satisfies the following properties:

(A1) For all $p \in \mathbb{R}^d$, $x \mapsto W(x, p)$ is Q-periodic, where $Q := (-1/2, 1/2)^N$;

(A2) W is a Carathéodory function, i.e.,

(i) for all $p \in \mathbb{R}^d$, the function $x \mapsto W(x, p)$ is measurable,

(ii) for a.e. $x \in Q$, the function $p \mapsto W(x, p)$ is continuous;

(A3) There exist $z_1, z_2 \in \mathbb{R}^d$ such that, for a.e. $x \in Q$, $W(x, p) = 0$ if and only if $p \in \{z_1, z_2\}$,

(A4) There exists a continuous function $\widetilde{W} : \mathbb{R}^d \to [0, \infty)$, vanishing only at $p = z_1$ and at $p = z_2$, such that $\widetilde{W}(p) \leqslant W(x, p)$ for a.e. $x \in Q$;

(A5) There exist $C > 0$ and $q \geqslant 2$ such that

$$\frac{1}{C}|p|^q - C \leqslant W(x, p) \leqslant C\big(1 + |p|^q\big)$$

for a.e. $x \in Q$ and all $p \in \mathbb{R}^d$.

Remark 5.1. Assumption (A2)(i) above is the strongest we can ask when modeling periodic inclusions of different materials. Indeed, when each cell Q is composed of k different inclusions of materials each in a region $E_1, \dots, E_k \subset Q$, the potential W takes the form

$$W(x, p) := \sum_{i=1}^{k} W_i(p)\chi_{E_i}(x),$$

where $W_i : \mathbb{R}^d \to [0, \infty)$ are continuous functions with quadratic growth at infinity and such that $W_i(p) = 0$ if and only if $p \in \{z_1, z_2\}$. Therefore the function W in the first variable is, in general, only measurable. Moreover, the continuity of W in the second variable, the nondegeneracy of the potential (A4), and the growth at infinity in the second variable (A5) are compatible with what is usually assumed in the physical literature.

The limiting functional will be an interfacial energy whose energy density is defined via a cell formula as follows.

Definition 5.2. For $\nu \in \mathbb{S}^{N-1}$, let $u_{0,\nu} : \mathbb{R}^N \to \mathbb{R}^d$ be the function

$$u_{0,\nu}(x) := \begin{cases} z_1 & \text{if } x \cdot \nu \leqslant 0, \\ z_2 & \text{if } x \cdot \nu > 0, \end{cases}$$

and denote by $\mathcal{Q}_\nu$ the family of cubes centered at the origin with unit length sides and having two faces orthogonal to ν. For $T > 0$, $Q_\nu \in \mathcal{Q}_\nu$, and $\rho \in C_c^\infty(B(0,1))$ with $\int_{\mathbb{R}^N} \rho(x)\mathrm{d}x = 1$, where $B(0,1)$ is the unit ball in $\mathbb{R}^N$, consider the class of functions

$$\mathcal{C}(\rho, Q_\nu, T) := \{u \in H^1(TQ_\nu; \mathbb{R}^d) : u = u_{0,\nu} * \rho \text{ on } \partial(TQ_\nu)\}.$$

We define the function $\sigma : \mathbb{S}^{N-1} \to [0,\infty)$ as

$$\sigma(\nu) := \lim_{T\to\infty} g(\nu, T),$$

where, for each $\nu \in \mathbb{S}^{N-1}$ and $T > 0$,

$$g(\nu, T) := \frac{1}{T^{N-1}} \inf\left\{\int_{TQ_\nu} \left[W\big(y, u(y)\big) + |\nabla u|^2\right]\mathrm{d}y : Q_\nu \in \mathcal{Q}_\nu,\ u \in \mathcal{C}(\rho, Q_\nu, T)\right\}.$$

The main properties of the function $\sigma : \mathbb{S}^{N-1} \to [0,\infty)$ that are relevant for our study are collected in the following result. For the proof, see [36, LEMMA 4.1, REMARK 4.2, LEMMA 4.3, PROPOSITION 4.4].

Lemma 5.3. *The following hold:*

(i) *For every $\nu \in \mathbb{S}^{N-1}$, the quantity $\sigma(\nu)$ is well defined and finite;*

(ii) *The value of $\sigma(\nu)$ does not depend on the choice of the mollifier ρ;*

(iii) *The map $\nu \mapsto \sigma(\nu)$ is upper semicontinuous on $\mathbb{S}^{N-1}$;*

(iv) *The infimum in the definition of $g(\nu, T)$ may be taken with respect to one fixed cube $Q_\nu \in \mathcal{Q}_\nu$, i.e., given $\nu \in \mathbb{S}^{N-1}$, for any $Q_\nu \in \mathcal{Q}_\nu$ it holds*

$$\sigma(\nu) = \lim_{T\to\infty} \frac{1}{T^{N-1}} \inf\left\{\int_{TQ_\nu} \left[W\big(y, u(y)\big) + |\nabla u|^2\right]\mathrm{d}y : u \in \mathcal{C}(\rho, Q_\nu, T)\right\}.$$

We are now in position to introduce the limiting functional.

Definition 5.4. Define the functional $\mathcal{F}_0 : L^1(\Omega; \mathbb{R}^d) \to [0,\infty]$ as

$$\mathcal{F}_0(u) := \begin{cases} \int_{\partial^* A} \sigma(\nu_A(x))\,\mathrm{d}\mathcal{H}^{N-1}(x) & \text{if } u \in \mathrm{BV}(\Omega; \{z_1, z_2\}), \\ +\infty & \text{else,} \end{cases} \tag{5.1}$$

where $A := \{u = z_1\}$, and $\nu_A(x)$ denotes the measure theoretic external unit normal to the reduced boundary $\partial^* A$ of A at the point x.

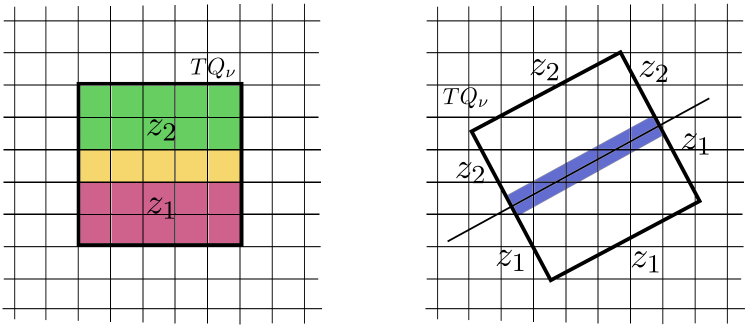


FIGURE 1

The source of anistropy for the limiting functional. If $\nu_A(x)$ is oriented with a direction of periodicity of W, the (local) recovery sequence would simply be obtained by using a rescaled version of the recovery sequence for $\sigma(\nu_A(x))$ in each yellow cube and by setting z_1 in the green region, and z_2 in the pink one. If, instead, $\nu_A(x)$ is not oriented with a direction of periodicity of W, the above procedure does not guarantee that we recover the desired energy, since the energy of such functions is *not* the sum of the energy of each cube.

Remark 5.5. Note that by Lemma 5.3 (i), it holds that $\mathcal{F}_0(u) < \infty$ for all $u \in \mathrm{BV}(\Omega;\{z_1,z_2\})$, and, by Lemma 5.3 (ii), the definition does not depend on the choice of the mollifier ρ.

Theorem 5.6. *Let* $\{\varepsilon_n\}_{n\in\mathbb{N}} \subset (0,1)$ *be a sequence such that* $\varepsilon_n \to 0^+$ *as* $n \to \infty$. *Assume that (A1), (A2), (A3), (A4), and (A5) hold.*

(i) *If* $\{u_n\}_{n\in\mathbb{N}} \subset H^1(\Omega;\mathbb{R}^d)$ *is such that*

$$\sup_{n\in\mathbb{N}} \mathcal{F}_{\varepsilon_n}(u_n) < +\infty$$

then, up to a subsequence (not relabeled), $u_n \to u$ *in* $L^1(\Omega;\mathbb{R}^d)$, *for some function* $u \in \mathrm{BV}(\Omega;\{z_1,z_2\})$.

(ii) *The functional* $\mathcal{F}_0$ *is the* Γ*-limit in the* L^1 *topology of the family of functionals* $\{\mathcal{F}_{\varepsilon_n}\}_{n\in\mathbb{N}}$.

Remark 5.7. The most interesting aspect of the above result is the anisotropic character of the limiting functional. This might come as a surprise since the initial functional $\mathcal{F}_\varepsilon$ is isotropic in its penalization of gradients, but there is a hidden anisotropy: the possible mismatch between the directions of periodicity of W and the local orientation of the limiting interface $\partial^* A$ (see Figure 1).

We would like to comment on the main ideas behind the proof of Theorem 5.6. Compactness follows by using classical arguments (see [65]), since the nondegeneracy assumption (A4) allows reducing to the case of a nonoscillating potential

$$\mathcal{F}_{\varepsilon_n}(u_n) \geq \int_\Omega \left[\frac{1}{\varepsilon_n}\widetilde{W}\big(u_n(x)\big) + \varepsilon_n\left|\nabla_n u(x)\right|^2\right] \mathrm{d}x.$$

The liminf inequality (see [36, PROPOSITION 6.1]) is based on a standard blow-up argument (see [62]) at a point $x_0 \in \partial^* A$ to reduce to the case where the limiting function is $u_{0,\nu}$ and the domain is $Q_\nu \in \mathcal{Q}_\nu$, where $\nu = \nu_A(x_0)$. Then, a technical lemma (see [36, LEMMA 3.1]) in the spirit of De Giorgi's slicing method (see [46]) allows modifying the given sequence $\{u_n\}_{n\in\mathbb{N}} \subset H^1(Q_\nu;\mathbb{R}^d)$ into a new sequence $\{v_n\}_{n\in\mathbb{N}} \subset H^1(Q_\nu;\mathbb{R}^d)$ with $v_n \to u_{0,\nu}$ in L^1 such that

$$\liminf_{n\to\infty} \mathcal{F}_{\varepsilon_n}(u_n) \geqslant \limsup_{n\to\infty} \mathcal{F}_{\varepsilon_n}(v_n),$$

and $v_n = \rho_n * u_{0,\nu}$ on ∂Q_ν, where $\rho_n(x) := \varepsilon_n^{-N} \rho(x/\varepsilon_n)$. The required inequality then follows by using a change of variable, and the definition of $\sigma(\nu)$ together with Lemma 5.3 (iv).

The main challenges are related to the proof of the limsup inequality (see [36, PROPOSITION 7.1]) for a function $u \in \mathrm{BV}(\Omega, \{a,b\})$, which requires new geometric arguments. The idea is first to prove the result for functions $u \in \mathrm{BV}(\Omega;\{a,b\})$, whose outer normals to the reduce boundary have rational coordinates, and then use the density of this class of functions in $\mathrm{BV}(\Omega;\{a,b\})$ together with Reshetnyak's upper semicontinuity theorem (by Lemma 5.3 (iii) the function $\nu \mapsto \sigma(\nu)$ is upper semicontinuous on $\mathbb{S}^{N-1}$) to conclude in the general case. In order to tackle the first step, we use a general strategy developed by De Giorgi, which can be seen as a sort of *reverse* blow-up argument: we consider the localized Γ-limsup as a map on Borel sets and we prove that it is indeed a Radon measure λ. This is done by using a simplification of the De Giorgi–Letta coincidence criterion for Borel measures (see [47]) by Dal Maso, Fonseca, and Leoni (see [40, COROLLARY 5.2]). Next, we show that λ is absolutely continuous with respect to the measure $\mu := \mathcal{H}^{N-1}\lfloor_{\partial^* A}$. The result follows by proving that the density of λ with respect to μ at a point $x_0 \in \partial^* A$ is bounded above by $\sigma(\nu_A(x_0))$. It is in this step that we exploit the fact that $\nu_A(x_0) \in \mathbb{S}^{N-1} \cap \mathbb{Q}^{N-1}$. Indeed, by using the fact that W is periodic (with a different period) also as a function on any cube Q whose faces are normal to directions in $\mathbb{S}^{N-1} \cap \mathbb{Q}^{N-1}$, we can estimate the energy of a configuration similar to that in Figure 1 on the left.

Remark 5.8. The strategy used to prove the above result is robust enough to be easily adapted to prove the analogous result when a mass constraint is enforced. Moreover, as a consequence of the Γ-limit result, we get that the function $\sigma : \mathbb{S}^{N-1} \to [0,\infty)$ is continuous, and its 1-homogeneous extension is convex.

The upshot of the foregoing result is that microscopic heterogeneities during phase transitions result in anisotropic surface tensions at the macroscopic level. Natural follow-up questions are:

(1) Beyond convexity, what can one say about the effective surface tension σ? What functions σ are attainable as effective surface tensions of phase transitions in periodic media?

(2) Considering the gradient flow dynamics of an energy as in (4.1), what are the $\varepsilon \to 0$ asymptotics? Does one indeed obtain a suitable weak formulation of anisotropic mean curvature flow, by analogy with the isotropic setting? Further-

more, what happens to the asymptotics of the gradient flow when the length-scales of homogenization and phase transitions differ?

In [32], we provide partial answers to the first question above, by relating it to a geometry problem. In the sequel, we assume the product form of the potential W:

$$W(y,\xi) := a(y)(1-u^2)^2, \quad y \in \mathbb{R}^N, u \in \mathbb{R}. \tag{5.2}$$

Here $a : \mathbb{R}^N \to \mathbb{R}$ is Q-periodic, and nondegenerate in the sense that

$$\theta \leqslant a(y) \leqslant \Theta, \quad y \in \mathbb{R}^N, \tag{5.3}$$

for some $0 < \theta < \Theta < \infty$. Note that assumptions (A1)–(A5) of Section 5.1 are satisfied with $z_1 = -1$, $z_2 = 1$, and $\widetilde{W}(p) := (1-p^2)^2$. The fact that u is scalar-valued is crucial for a number of the results proven in [32, 33] since we use arguments based on the maximum principle. However, this is not the case of all the results, and we will indicate this as appropriate.

5.2. Bounds on the anisotropic surface tension σ

5.2.1. A geometric framework

Consider the periodic Riemannian metric on $\mathbb{R}^N$ that is conformal to the Euclidean one, defined as follows: given points $x, y \in \mathbb{R}^N$, we set

$$d_{\sqrt{a}}(x,y) := \inf_{\gamma} \int_0^1 \sqrt{a(\gamma(t))}\big|\dot{\gamma}(t)\big|\,dt,$$

where the infimum is taken over Lipschitz continuous curves $\gamma : [0,1] \to \mathbb{R}^N$ such that $\gamma(0) = x$, $\gamma(1) = y$. It is easily seen that the formula defining $d_{\sqrt{a}}$ is independent of the parameterization of the competitor curves γ. Furthermore, standard arguments via the Hopf–Rinow theorem imply that $\mathbb{R}^N$ with the metric $d_{\sqrt{a}}$ is a complete metric space. Equivalently, geodesically complete: given any pair of points $x, y \in \mathbb{R}^N$ there exists a distance-minimizing geodesic joining them, whose length is equal to $d_{\sqrt{a}}(x,y)$ (see [86] for details). Now fix a direction $\nu \in \mathbb{S}^{N-1}$, and consider the plane Σ_ν through the origin with normal ν,

$$\Sigma_\nu := \big\{y \in \mathbb{R}^N : y \cdot \nu = 0\big\}.$$

Next, define the signed distance function in the $d_{\sqrt{a}}$-metric to the plane Σ_ν, via

$$h_\nu(y) := \operatorname{sgn}(y \cdot \nu) \inf_{z \in \Sigma_\nu} d_{\sqrt{a}}(y,z),$$

where the signum function is defined as

$$\operatorname{sgn}(t) := \begin{cases} 1, & t \geqslant 0, \\ -1, & t < 0. \end{cases}$$

It can be shown (see [32, LEMMA 2.2]) that h_ν is Lipschitz continuous, with

$$\big|\nabla h_\nu(y)\big| = \sqrt{a(y)} \quad \text{at a.e. } y \in \mathbb{R}^N. \tag{5.4}$$

These observations, together with (5.3), yield

$$\begin{aligned} \sqrt{\theta}(y\cdot\nu) \leqslant h_\nu(y) \leqslant \sqrt{\Theta}(y\cdot\nu), &\quad y\cdot\nu \geqslant 0,\\ \sqrt{\Theta}(y\cdot\nu) \leqslant h_\nu(y) \leqslant \sqrt{\theta}(y\cdot\nu), &\quad y\cdot\nu < 0. \end{aligned} \tag{5.5}$$

In order to explain the relationship that the $d_{\sqrt{a}}$-metric bears with the anisotropic surface tension σ, it is useful to revisit the case $a \equiv 1$, and the celebrated Modica–Mortola example. Then,

$$\sigma(\nu) = \lim_{T\to\infty} \frac{1}{T^{N-1}} \inf\left\{ \int_{TQ_\nu} \left[W\big(u(y)\big) + |\nabla u|^2\right] : u \in \mathcal{C}(\rho, Q_\nu, T)\right\}.$$

Elementary algebraic manipulations that effectively boil down to completing the square, yield that the infimum above is asymptotically reached by the one-dimensional profile satisfying equipartition of energy. In the model case of (5.2), this entails that the optimal cost is achieved by the choice $u(y) = q \circ (y\cdot\nu)$, where $q := \tanh$. The associated cost is given by

$$\sigma(\nu) \equiv \sigma_0 := \int_{-\infty}^{\infty} \left[W\big(q\circ(y\cdot\nu)\big) + \left|\nabla\big(q\circ(y\cdot\nu)\big)\right|^2\right] d(y\cdot\nu) = 2\int_{-1}^{1} \sqrt{W(s)}\, ds. \tag{5.6}$$

To make the connection to the $\sqrt{a}$-metric, we begin by noting that when $a \equiv 1$ we have $h_\nu(y) \equiv y\cdot\nu$. Our main motivation, then, is to obtain a similar formula that is exact when a is nonconstant, or at least supplies reasonable bounds for the nonconstant $\nu \mapsto \sigma(\nu)$. We do so by encoding the heterogeneous effects of a into the geometry of the underlying space, i.e., by working in the $\sqrt{a}$-metric. We turn to making these comments precise.

Fix $\nu \in \mathbb{S}^{N-1}$. Then, the cell formula defining $\sigma(\nu)$, proven in [36,37] and specialized to our setting, reads (see Lemma 5.3 (iv))

$$\begin{aligned} \sigma(\nu) = \lim_{T\to\infty} \frac{1}{T^{N-1}} \inf\bigg\{ \int_{TQ_\nu} &\left[a(y)W(u) + |\nabla u|^2\right] dy : u \in H^1(TQ_\nu),\\ &u = \rho * u_{0,\nu} \text{ on } \partial(TQ_\nu)\bigg\}. \end{aligned}$$

Here, we recall that $u_{0,\nu}(y) := \operatorname{sgn}(y\cdot\nu)$ and ρ is any standard smooth normalized mollifier (it is shown in Lemma 5.3 (ii) that $\sigma(\nu)$ is independent of this choice). A preliminary step is to observe, by De Giorgi's slicing method (see [32, LEMMA A.1]), that, equivalently,

$$\begin{aligned} \sigma(\nu) = \lim_{T\to\infty} \frac{1}{T^{N-1}} \inf\bigg\{ \int_{TQ_\nu} &\left[a(y)W(u) + |\nabla u|^2\right] dy : u \in H^1(TQ_\nu),\\ &u = q \circ h_\nu \text{ along } \partial(TQ_\nu)\bigg\}. \end{aligned} \tag{5.7}$$

For each fixed $T \gg 1$, by the direct method of the calculus of variations, the variational problem inside the limit has a minimizer. Such a minimizer is, perhaps, not unique, but for each T we select one, and call it u_T. We discuss various properties of u_T below in Section 5.2.2. In light of (5.7), it is clear by energy comparison that

$$\sigma(\nu) \leqslant \liminf_{T\to\infty} \frac{1}{T^{N-1}} \int_{TQ_\nu} \left[a(y)W(q\circ h_\nu) + \left|\nabla(q\circ h_\nu)\right|^2\right] dy.$$

Towards proving the opposite bound, we introduce the function $\phi : \mathbb{R} \to \mathbb{R}$, by

$$\phi(z) := 2\int_0^z \sqrt{W(s)}\, ds.$$

This function plays a fundamental role in the Modica–Mortola analysis corresponding to $a \equiv 1$. For any $T \gg 1$, using (5.4) and completing squares, we find

$$\begin{aligned}
&\frac{1}{T^{N-1}}\int_{TQ_\nu} \big[a(y)W(u_T) + |\nabla u_T|^2\big]\, dy \\
&\quad= \frac{2}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \sqrt{W(u_T)}\nabla u_T\, dy + \frac{1}{T^{N-1}}\int_{TQ_\nu} \big|\nabla u_T - \sqrt{W(u_T)}\nabla h_\nu\big|^2\, dy \\
&\quad\geqslant \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(u_T)\big)\, dy \\
&\quad= \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(q\circ h_\nu)\big)\, dy + \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(u_T) - \phi(q\circ h_\nu)\big)\, dy \\
&\quad= \frac{1}{T^{N-1}}\int_{TQ_\nu} |\nabla h_\nu|^2 \phi'(q\circ h_\nu) q'(h_\nu)\, dy \\
&\quad+ \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(u_T) - \phi(q\circ h_\nu)\big)\, dy \\
&\quad= \frac{1}{T^{N-1}}\int_{TQ_\nu} 2a(y)W(q\circ h_\nu)\, dy + \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(u_T) - \phi(q\circ h_\nu)\big)\, dy \\
&\quad= \frac{1}{T^{N-1}}\int_{TQ_\nu} \big[a(y)W(q\circ h_\nu) + \big|\nabla(q\circ h_\nu)\big|^2\big]\, dy \\
&\qquad+ \frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu \cdot \nabla\big(\phi(u_T) - \phi(q\circ h_\nu)\big)\, dy,
\end{aligned} \tag{5.8}$$

where in the last line we used the fact that the function $q \circ h_\nu$ achieves equipartition of energy. Indeed, by the definition of h_ν, we have

$$\big|\nabla(q\circ h_\nu)(y)\big|^2 = \big(q'\big(h_\nu(y)\big)\big)^2 \big|\nabla h_\nu(y)\big|^2 = a(y)W\big(q\big(h_\nu(y)\big)\big).$$

Provided we can control the error term

$$\limsup_{T\to\infty}\left|\frac{1}{T^{N-1}}\int_{TQ_\nu} \nabla h_\nu(y) \cdot \nabla\big(\phi(u_T) - \phi(q\circ h_\nu)\big)\, dy\right| := \lambda_0(\nu),$$

we observe that the test function $q \circ h_\nu$ gives two-sided bounds on $\sigma(\nu)$. Controlling the term λ_0 is complicated by the fact that it couples a product of weakly converging sequences (on expanding domains). Indeed, rescaling using $y = Tx$ in order to work in a fixed domain Q_ν, the two weakly converging factors making up the above product are

(1) the oscillatory factor: by (5.4) and (5.3), the term $\{\nabla h_\nu(T\,\cdot)\}_T$, which is bounded in L^∞, converges weakly-*; and

(2) the concentration factor: the terms $\nabla\phi(u_T(T\,\cdot))$ and $\nabla\phi(q \circ h_\nu(T\,\cdot)$ converge weakly-* to measures (see Section 5.2.2 for precise statements).

In particular, as one of the factors converges to a measure, standard tools such as compensated compactness, used traditionally to pass to the limit in products of weakly converging sequences, are unavailable, and we must control this term "by hand." In Section 5.2.2 below, we obtain fine information on the concentration effects, and in Section 5.2.3 we deduce partial results concerning the oscillatory effects. Finally, we put these together in Section 5.2.4, where we obtain bounds on $\lambda_0(\nu)$.

5.2.2. Structure of minimizers of the cell formula

For fixed $T \gg 1$, let $u_T \in C^2(TQ_\nu)$ (by elliptic regularity) a minimizer of the energy

$$\int_{TQ_\nu} \left[a(y)W(u) + |\nabla u|^2\right] dy,$$

among competitors that equal $q \circ h_\nu$ along the boundary $\partial(TQ_\nu)$, and set

$$v_T(x) := u_T(Tx), \quad x \in Q_\nu.$$

which minimizes the energy.

Lemma 5.9. *The functions v_T converge in L^1 to $u_{0,\nu} : Q_\nu \to \{\pm 1\}$.*

The proof of this lemma (see [32, LEMMA 3.1]) is a nice application of the convexity of the one-homogeneous extension of σ (see Remark 5.8), using Jensen's inequality. The argument, without any changes, holds in the complete generality of the setting of [36] on the potential (vectorial, coupled, measurable dependence on the fast variable), and does not rely on the specific structure requested in (5.2). Combining Lemma 5.9 with the results of Caffarelli–Cordoba [29], we find that the level sets of v_T, for T sufficiently large, converge uniformly to $\Sigma_\nu \cap Q_\nu$.

Restricting ourselves to the scalar setting of (5.2), an argument using the strong maximum principle yields that, for all $T < \infty$, we have

$$-1 < u_T(y) < 1,$$

(see [32, LEMMA 3.2]). In particular, $w_T := \frac{1}{\sqrt{2}} \tanh^{-1} u_T$ is well defined, finite, and smooth in TQ_ν. Further, the function w_T verifies the elliptic boundary value problem

$$\begin{cases} \Delta w_T = \frac{4}{\sqrt{2}} \tanh w_T(|\nabla w_T|^2 - a(y)), & y \in TQ_\nu, \\ w_T(y) = h_\nu(y), & y \in \partial(TQ_\nu). \end{cases}$$

Proposition 5.10. *Let w_T be as above, and let $T \gg 1$. There exist universal constants α_0 and $\eta_0 > 0$ such that the following holds:*

$$\begin{cases} \sqrt{\Theta}(y \cdot \nu) - \alpha_0 \geqslant w_T(y) \geqslant \sqrt{\theta}(y \cdot \nu) - \eta_0, & \text{if } w_T(y) > 0, \\ -\sqrt{\theta}(y \cdot \nu) + \eta_0 \geqslant w_T(y) \geqslant -\sqrt{\Theta}(y \cdot \nu) + \alpha_0, & \text{if } w_T(y) < 0. \end{cases} \tag{5.9}$$

Proposition 5.10 asserts that, up to universal constants, the function w_T satisfies exactly the same growth rates as the function h_ν, see (5.5). To prove Proposition 5.10, consider, for instance, the lower bound in the first of the two inequalities in (5.9). The main observation is that the function $y \mapsto \zeta_T(y) := \frac{y \cdot \nu}{w_T(y) + \eta_0}$ satisfies an elliptic PDE that verifies a maximum principle. The remaining inequalities follow from similar arguments, and we refer the reader to [32, PROPOSITION 3.4] for details.

5.2.3. The planar metric problem

Our results on the distance function h_ν concern its large-scale behavior. The bounds on σ that we discuss in Section 5.2.4 below, depend solely on the large-scale behavior of the distance functions h_ν for which one can readily invoke efficient numerical algorithms, for example fast marching and sweeping methods [85].

A natural question concerns the large-scale homogenized behavior of h_ν, i.e., the characterization of the limit

$$\lim_{T\to\infty} \frac{h_\nu(Ty)}{T}, \quad y \in \mathbb{R}^N,$$

in a suitable topology of functions. We completely answer this question.

Theorem 5.11. *Let $\nu \in \mathbb{S}^{N-1}$. There exists a real number $c(\nu) \in [\sqrt{\theta}, \sqrt{\Theta}]$ such that for each $K \subseteq \mathbb{R}^N$ compact, we have*

$$\lim_{T\to\infty} \sup_{y\in K} \left| \frac{1}{T} h_\nu(Ty) - c(\nu)(y\cdot\nu) \right| = 0.$$

Moreover, for all compact subsets K of $\mathbb{R}^N \setminus \Sigma_\nu$, we have

$$\lim_{T\to\infty} \sup_{y\in K} \left| \frac{1}{T(y\cdot\nu)} h_\nu(Ty) - c(\nu) \right| = 0.$$

We can interpret Theorem 5.11 as a homogenization result for the eikonal equation in half-spaces. Indeed, it is well known (see, for example, [77]) that for each fixed $\nu \in \mathbb{S}^{N-1}$, the functions $k_T(y) := T^{-1}h_\nu(Ty)$ and $\ell(y) := c(\nu)(y\cdot\nu)$ are the unique viscosity solutions to

$$\begin{cases} |\nabla k_T| = \sqrt{a(Ty)} & \text{in } \{y\cdot\nu \geqslant 0\}, \\ k_T = 0 & \text{on } \Sigma_\nu, \end{cases} \quad \text{and} \quad \begin{cases} |\nabla \ell| = c(\nu) & \text{in } \{y\cdot\nu \geqslant 0\}, \\ \ell = 0 & \text{on } \Sigma_\nu. \end{cases} \tag{5.10}$$

Theorem 5.11 shows that viscosity solutions of the heterogeneous eikonal equations, i.e., k_T in (5.10), converge locally uniformly to ℓ. A viscous and stochastic version of these equations (termed the "planar metric problem") was introduced by Armstrong and Cardaliaguet [7], and studied by others [55, 56] in the context of stochastic homogenization of geometric flows. Small modifications of our arguments, in fact, yield homogenization theorems for first order Hamilton–Jacobi equations in almost periodic media in half-spaces, with Lipschitz dependence on the fast variable, and convex dependence on the gradient variable.

5.2.4. Bounds on the anisotropic surface tension

As explained in the string of inequalities (5.8), the function $q \circ h_\nu$ provides tight upper and lower bounds for the effective anisotropy $\sigma(\nu)$. To be precise, we have

Theorem 5.12. *Let $\sigma : \mathbb{S}^{N-1} \to [0,\infty)$ be the anisotropic surface energy as in (5.2). Let $q : \mathbb{R} \to \mathbb{R}$ be defined by*

$$q(z) := \tanh(z), \quad z \in \mathbb{R}.$$

For $\nu \in \mathbb{S}^{N-1}$, define

$$\underline{\lambda}(\nu) := \liminf_{T\to\infty} \frac{1}{T^{N-1}} \int_{TQ_\nu} \left[a(y)W(q \circ h_\nu) + \left|\nabla(q \circ h_\nu)\right|^2\right] dy,$$

$$\overline{\lambda}(\nu) := \limsup_{T\to\infty} \frac{1}{T^{N-1}} \int_{TQ_\nu} \left[a(y)W(q \circ h_\nu) + \left|\nabla(q \circ h_\nu)\right|^2\right] dy.$$

There exist $\Lambda_0 > 0$ and $\lambda_0 : \mathbb{S}^{N-1} \to [0, \Lambda_0]$ such that

$$\overline{\lambda}(\nu) - \lambda_0(\nu) \leqslant \sigma(\nu) \leqslant \underline{\lambda}(\nu).$$

We remark that in general, $\lambda_0(\nu)$ is never zero, unless $a \equiv 1$.

FUNDING

The research of Rita Ferreira was partially supported by baseline and start-up funds from King Abdullah University of Science and Technology (KAUST). The research of Irene Fonseca was partially funded by NSF grant DMS-19006238. The research of Raghav Venkatraman was partially funded by the Simons Foundation Postdoctoral Award, and an AMS-Simons Travel Award.

REFERENCES

[1] G. Allaire, Homogenization and two-scale convergence. *SIAM J. Math. Anal.* **23** (1992), no. 6, 1482–1518.

[2] G. Allaire, *Shape optimization by the homogenization method*. Appl. Math. Sci. 146, Springer, New York, 2002.

[3] M. Amar, Two-scale convergence and homogenization on BV(Ω). *Asymptot. Anal.* **16** (1998), no. 1, 65–84.

[4] M. P. Anderson, Characterization of geological heterogeneity. In *Subsurface flow and transport: a stochastic approach*, pp. 23–43, Cambridge University Press, 1997.

[5] N. Ansini, A. Braides, and V. C. Piat, Interactions between homogenization and phase-transition processes. *Tr. Mat. Inst. Steklova* **236** (2002), 386–398.

[6] N. Ansini, A. Braides, and V. C. Piat, Gradient theory of phase transitions in composite media. *Proc. Roy. Soc. Edinburgh Sect. A* **133** (2003), no. 2, 265–296.

[7] S. Armstrong and P. Cardaliaguet, Stochastic homogenization of quasilinear Hamilton–Jacobi equations and geometric motions. *J. Eur. Math. Soc. (JEMS)* **20** (2018), no. 4, 797–864.

[8] S. Armstrong, T. Kuusi, and J.-C. Mourrat, *Quantitative stochastic homogenization and large-scale regularity*. Grundlehren Math. Wiss. 352, Springer, Cham, 2019.

[9] A. A. Azuha, A. Klimkowicz, and A. Takasaki, Quaternary quasicrystal alloys for hydrogen storage technology. *MRS Adv.* **5** (2020), no. 20, 1071–1083.

[10] J.-F. Babadjian and M. Barchiesi, A variational approach to the local character of G-closure: the convex case. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **26** (2009), no. 2, 351–373.

[11] M. Baía and I. Fonseca, The limit behavior of a family of variational multiscale problems. *Indiana Univ. Math. J.* **56** (2007), no. 1, 1–50.

[12] S. Baldo, Minimal interface criterion for phase transitions in mixtures of Cahn-Hilliard fluids. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **7** (1990), no. 2, 67–90.

[13] A. C. Barroso and I. Fonseca, Anisotropic singular perturbations–the vectorial case. *Proc. Roy. Soc. Edinburgh Sect. A* **124** (1994), no. 3, 527–571.

[14] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*. Stud. Appl. Math. 5, North-Holland Publishing Co., Amsterdam–New York, 1978.

[15] H. Berestycki, F. Hamel, and L. Roques, Analysis of the periodically fragmented environment model: I–species persistence. *J. Math. Biol.* **51** (2005), no. 1, 75–113.

[16] H. Berestycki, F. Hamel, and L. Roques, Analysis of the periodically fragmented environment model: II—biological invasions and pulsating travelling fronts. *J. Math. Pures Appl.* **84** (2005), no. 8, 1101–1146.

[17] K. Bhattacharya, Phase boundary propagation in a heterogeneous body. *Proc. R. Soc. Lond. A* **455** (1999), 757–766.

[18] P. D. Bloom, K. G. Baikerikar, J. U. Otaigbe, and V. V. Sheares, Development of novel polymer/quasicrystal composite materials. *Mater. Sci. Eng. A* **294–296** (2000), 156–159.

[19] G. Bouchitté, Homogénéisation sur $\mathrm{BV}(\Omega)$ de fonctionnelles intégrales à croissance linéaire. Application à un problème d'analyse limite en plasticité. *C. R. Acad. Sci. Paris Sér. I Math.* **301** (1985), no. 17, 785–788.

[20] G. Bouchitté, I. Fonseca, and L. Mascarenhas, A global method for relaxation. *Arch. Ration. Mech. Anal.* **145** (1998), no. 1, 51–98.

[21] G. Bouchitté, S. Guenneau, and F. Zolla, Homogenization of dielectric photonic quasi crystals. *Multiscale Model. Simul.* **8** (2010), no. 5, 1862–1881.

[22] A. Braides, Almost periodic methods in the theory of homogenization. *Appl. Anal.* **47** (1992), no. 4, 259–277.

[23] A. Braides, A. Defranceschi, and E. Vitali, Homogenization of free discontinuity problems. *Arch. Ration. Mech. Anal.* **135** (1996), no. 4, 297–356.

[24] A. Braides, I. Fonseca, and G. Francfort, 3D–2D asymptotic analysis for inhomogeneous thin films. *Indiana Univ. Math. J.* **49** (2000), no. 4, 1367–1404.

[25] A. Braides, I. Fonseca, and G. Leoni, $\mathcal{A}$-quasiconvexity: relaxation and homogenization. *ESAIM Control Optim. Calc. Var.* **5** (2000), 539–577.

[26] A. Braides and C. I. Zeppieri, Multiscale analysis of a prototypical model for the interaction between microstructure and surface energy. *Interfaces Free Bound.* **11** (2009), no. 1, 61–118.

[27] L. Bufford, E. Davoli, and I. Fonseca, Multiscale homogenization in Kirchhoff's nonlinear plate theory. *Math. Models Methods Appl. Sci.* **25** (2015), no. 9, 1765–1812.

[28] L. Bufford and I. Fonseca, A note on two scale compactness for $p = 1$. *Port. Math.* **72** (2015), no. 2–3, 101–117.

[29] L. A. Caffarelli and A. Córdoba, Uniform convergence of a singular perturbation problem. *Comm. Pure Appl. Math.* **48** (1995), no. 1, 1–12.

[30] J. W. Cahn and J. E. Taylor, An introduction to quasicrystals. In *The legacy of Sonya Kovalevskaya (Cambridge, Mass., and Amherst, Mass., 1985)*, pp. 265–286 Contemp. Math. 64, Amer. Math. Soc., Providence, RI, 1987.

[31] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*. Lecture Notes in Math. 580, Springer, Berlin, 1977.

[32] R. Choksi, I. Fonseca, J. Lin, and R. Venkatraman, *Homogenization of an Allen–Cahn equation in periodic media: a variational approach*. 2020.

[33] R. Choksi, I. Fonseca, J. Lin, and R. Venkatraman, *Homogenization for an Allen–Cahn equation in periodic media: a variational approach*. 2021.

[34] C. Conca and M. Vanninathan, Homogenization of periodic structures via Bloch decomposition. *SIAM J. Appl. Math.* **57** (1997), no. 6, 1639–1659.

[35] A. C. Costa and A. Soares, Homogenization of climate data: review and new perspectives using geostatistics. *Math. Geosci.* **41** (2009), no. 3, 291–305.

[36] R. Cristoferi, I. Fonseca, A. Hagerty, and C. Popovici, A homogenization result in the gradient theory of phase transitions. *Interfaces Free Bound.* **21** (2019), 367–408.

[37] R. Cristoferi, I. Fonseca, A. Hagerty, and C. Popovici, Erratum to: A homogenization result in the gradient theory of phase transitions. *Interfaces Free Bound.* **22** (2020), 245–250.

[38] R. Cristoferi and G. Gravina, Sharp interface limit of a multi-phase transitions model under nonisothermal conditions. *Calc. Var. Partial Differential Equations* **60** (2021).

[39] G. Dal Maso, *An introduction to* Γ*-convergence*. Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser Boston Inc., Boston, MA, 1993.

[40] G. Dal Maso, I. Fonseca, and G. Leoni, Nonlocal character of the reduced theory of thin films with higher order perturbations. *Adv. Calc. Var.* **3** (2010), no. 3, 287–319.

[41] G. Dal Maso and L. Modica, Nonlinear stochastic homogenization and ergodic theory. *J. Reine Angew. Math.* **368** (1986), 28–42.

[42] E. Davoli and I. Fonseca, Homogenization of integral energies under periodically oscillating differential constraints. *Calc. Var. Partial Differential Equations* **55** (2016), no. 3, 69.

[43] E. Davoli and I. Fonseca, Periodic homogenization of integral energies under space-dependent differential constraints. *Port. Math.* **73** (2016), no. 4, 279–317.

[44] N. G. de Bruijn, Algebraic theory of Penrose's non-periodic tilings of the plane. I, II: dedicated to G. Pólya. *Indag. Math. (N.S.)* **43** (1981), no. 1, 39–66.

[45] R. De Arcangelis and G. Gargiulo, Homogenization of integral functionals with linear growth defined on vector-valued functions. *NoDEA Nonlinear Differential Equations Appl.* **2** (1995), no. 3, 371–416.

[46] E. De Giorgi, Sulla convergenza di alcune successioni d'integrali del tipo dell'area. *Rend. Mat. (6)* **8** (1975), 277–294. Collection of articles dedicated to Mauro Picone on the occasion of his ninetieth birthday.

[47] E. De Giorgi and G. Letta, Une notion générale de convergence faible pour des fonctions croissantes d'ensemble. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (4)* **4** (1977), no. 1, 61–99.

[48] E. De Giorgi and S. Spagnolo, Sulla convergenza degli integrali dell'energia per operatori ellittici del secondo ordine. *Boll. Unione Mat. Ital.* **4** (1973), no. 8, 391–411.

[49] E. De Giorgi and S. Spagnolo, Sulla convergenza degli integrali dell'energia per operatori ellittici del secondo ordine. *Boll. Unione Mat. Ital.* **4** (1973), no. 8, 391–411.

[50] N. Dirr, M. Lucia, and M. Novaga, Γ-convergence of the Allen–Cahn energy with an oscillating forcing term. *Interfaces Free Bound.* **8** (2006), no. 1, 47–78.

[51] N. Dirr, M. Lucia, and M. Novaga, Gradient theory of phase transitions with a rapidly oscillating forcing term. *Asymptot. Anal.* **60** (2008), no. 1–2, 29–59.

[52] M. Duneau and A. Katz, Quasiperiodic patterns. *Phys. Rev. Lett.* **54** (1985), no. 25, 2688–2691.

[53] M. Engel, P. F. Damasceno, C. L. Phillips, and S. C. Glotzer, Computational self-assembly of a one-component icosahedral quasicrystal. *Nat. Mater.* **14** (2015), no. 1, 109–116.

[54] A. Fannjiang and G. Papanicolaou, Convection enhanced diffusion for periodic flows. *SIAM J. Appl. Math.* **54** (1994), no. 2, 333–408.

[55] W. M. Feldman, Mean curvature flow with positive random forcing in 2-d. 2019, arXiv:1911.00488.

[56] W. M. Feldman and P. E. Souganidis, Homogenization and non-homogenization of certain non-convex Hamilton–Jacobi equations. *J. Math. Pures Appl. (9)* **108** (2017), no. 5, 751–782.

[57] R. Ferreira and I. Fonseca, Characterization of the multiscale limit associated with bounded sequences in BV. *J. Convex Anal.* **19** (2012), no. 2, 403–452.

[58] R. Ferreira and I. Fonseca, Reiterated homogenization in BV via multiscale convergence. *SIAM J. Math. Anal.* **44** (2012), no. 3, 2053–2098.

[59] R. Ferreira, I. Fonseca, and R. Venkatraman, Homogenization of quasi-crystalline functionals via two-scale-cut-and-project convergence. *SIAM J. Math. Anal.* **53** (2021), no. 2, 1785–1817.

[60] I. Fonseca and G. A. Francfort, Relaxation in BV versus quasiconvexification in $W^{1,p}$; a model for the interaction between fracture and damage. *Calc. Var. Partial Differential Equations* **3** (1995), no. 4, 407–446.

[61] I. Fonseca and S. Krömer, Multiple integrals under differential constraints: two-scale convergence and homogenization. *Indiana Univ. Math. J.* **59** (2010), no. 2, 427–457.

[62] I. Fonseca and S. Müller, Quasi-convex integrands and lower semicontinuity in L^1. *SIAM J. Math. Anal.* **23** (1992), no. 5, 1081–1098.

[63] I. Fonseca and S. Müller, A-quasiconvexity, lower semicontinuity, and Young measures. *SIAM J. Math. Anal.* **30** (1999), no. 6, 1355–1390.

[64] I. Fonseca and C. Popovici, Coupled singular perturbations for phase transitions. *Asymptot. Anal.* **44** (2005), no. 3–4, 299–325.

[65] I. Fonseca and L. Tartar, The gradient theory of phase transitions for systems with two potential wells. *Proc. Roy. Soc. Edinburgh Sect. A* **111** (1989), no. 1–2, 89–102.

[66] I. Fonseca and E. Zappale, Multiscale relaxation of convex functionals. *J. Convex Anal.* **10** (2003), no. 2, 325–350.

[67] G. A. Francfort and S. Müller, Combined effects of homogenization and singular perturbations in elasticity. *J. Reine Angew. Math.* **454** (1994), 1–35.

[68] J. L. Gevertz, G. T. Gillies, and S. Torquato, Simulating tumor growth in confined heterogeneous environments. *Phys. Biol.* **5** (2008), no. 3, 036010.

[69] A. Gloria and F. Otto, Quantitative results on the corrector equation in stochastic homogenization. *J. Eur. Math. Soc. (JEMS)* **19** (2017), no. 11, 3489–3548.

[70] J. Guo, T. Sun, and E. Pan, Three-dimensional nonlocal buckling of composite nanoplates with coated one-dimensional quasicrystal in an elastic medium. *Int. J. Solids Struct.* **185–186** (2020), 272–280.

[71] M. Heida, S. Neukamm, and M. Varga, Stochastic homogenization of Λ-convex gradient *Discrete Contin. Dyn. Syst. Ser. S* **14** (2021), 427–453.

[72] P. Hornung, S. Neukamm, and I. Velčić, Derivation of a homogenized non-linear plate theory from 3d elasticity. *Calc. Var. Partial Differential Equations* **51** (2014), no. 3–4, 677–699.

[73] K. Kamiya, T. Takeuchi, N. Kabeya, N. Wada, T. Ishimasa, A. Ochiai, K. Deguchi, K. Imura, and N. K. Sato, Discovery of superconductivity in quasicrystal. *Nat. Commun.* **9** (2018), no. 1, 154.

[74] S. Kenzari, D. Bonina, J. M. Dubois, and V. Fournée, Quasicrystal–polymer composites for selective laser sintering technology. *Mater. Des.* **35** (2012), 691–695. New Rubber Materials, Test Methods and Processes.

[75] S. M. Kozlov, The averaging of random operators. *Mat. Sb.* **109** (1979), no. 151(2), 188–202, 327.

[76] M. L. Levin and M. A. Miller, Maxwell a treatise on electricity and magnetism. *Usp. Fiz. Nauk* **135** (1981), no. 3, 425–440.

[77] C. Mantegazza and A. C. Mennucci, Hamilton-Jacobi equations and distance functions on Riemannian manifolds. *Appl. Math. Optim.* **47** (2003), no. 1, 1–25.

[78] L. Modica, The gradient theory of phase transitions and the minimal interface criterion. *Arch. Ration. Mech. Anal.* **98** (1987), no. 2.

[79] L. Modica, The gradient theory of phase transitions and the minimal interface criterion. *Arch. Ration. Mech. Anal.* **98** (1987), no. 2, 123–142.

[80] F. Murat, Compacité par compensation: condition nécessaire et suffisante de continuité faible sous une hypothèse de rang constant. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (4)* **8** (1981), no. 1, 69–102.

[81] Y. Nagaoka, H. Zhu, D. Eggert, and O. Chen, Single-component quasicrystalline nanocrystal superlattices through flexible polygon tiling rule. *Science* **362** (2018), no. 6421, 1396–1400.

[82] G. Nguetseng, A general convergence result for a functional related to the theory of homogenization. *SIAM J. Math. Anal.* **20** (1989), no. 3, 608–623.

[83] R. Penrose, Pentaplexity: a class of nonperiodic tilings of the plane. *Math. Intelligencer* **2** (1979/1980), no. 1, 32–37.

[84] L. Rayleigh, Lvi. on the influence of obstacles arranged in rectangular order upon the properties of a medium. *Philos. Mag.* **34** (1892), no. 211, 481–502.

[85] J. A. Sethian, Fast marching methods. *SIAM Rev.* **41** (1999), no. 2, 199–235.

[86] P. Sternberg, The effect of a singular perturbation on nonconvex variational problems. *Arch. Ration. Mech. Anal.* **101** (1988), no. 3, 209–260.

[87] L. Tartar, Compensated compactness and applications to partial differential equations. In *Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV*, pp. 136–212, Res. Notes Math. 39, Pitman, Boston, Mass.-London, 1979.

[88] S. Torquato, *Random heterogeneous materials*. Interdiscip. Appl. Math. 16, Springer, New York, 2002.

[89] N. G. Trillos and D. Slepčev, Continuum limit of total variation on point clouds. *Arch. Ration. Mech. Anal.* **220** (2016), no. 1, 193–241.

[90] I. Velčić, On the derivation of homogenized bending plate model. *Calc. Var. Partial Differential Equations* **53** (2015), no. 3–4, 561–586.

[91] E. Weinan, *Principles of multiscale modeling*. Cambridge University Press, 2011.

[92] N. Wellander, S. Guenneau, and E. Cherkaev, Homogenization of quasiperiodic structures and two-scale cut-and-projection convergence. 2019, arXiv:1911.03560.

[93] M. Xu, X. Teng, and J. Geng, Effect of cooling rates on solidification and microstructure of rapidly solidified $\mathrm{Mg}_{57}\mathrm{Zn}_{37}\mathrm{Y}_6$ quasicrystal alloy. *J. Mater. Res.* **30** (2015), no. 21, 3324–3330.

RITA FERREIRA

King Abdullah University of Science and Technology (KAUST), CEMSE Division, Thuwal 23955-6900, Saudi Arabia, rita.ferreira@kaust.edu.sa

IRENE FONSECA

Department of Mathematical Sciences, Carnegie Mellon University (CMU), Forbes Avenue, Pittsburgh, PA 15213, USA, fonseca@andrew.cmu.edu

RAGHAVENDRA VENKATRAMAN

Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012, USA, raghav@cims.nyu.edu

LIEB–THIRRING INEQUALITIES AND OTHER FUNCTIONAL INEQUALITIES FOR ORTHONORMAL SYSTEMS

RUPERT L. FRANK

ABSTRACT

We review recent results on functional inequalities for systems of orthonormal functions. The key finding is that for various operators the orthonormality leads to a gain over a simple application of the triangle inequality. The operators under consideration are either related to Sobolev-type inequalities or to Fourier restriction-type inequalities.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 42B37; Secondary 35Q40, 46E35, 47B10, 47N50, 81Q10

KEYWORDS

Lieb–Thirring inequality, orthonormal systems, Sobolev inequality, Strichartz inequality, Fourier restriction inequality

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3756–3774
DOI 10.4171/ICM2022/105

Published by EMS Press

1. INTRODUCTION

For more than four decades, Lieb–Thirring inequalities have played an important role in various areas of mathematical physics and analysis. The progress that has been made towards the conjectures in the area and many extensions and generalizations of the original inequalities have been reviewed in the surveys [5,19,38,42,43,49], the textbooks [51,52], as well as in the forthcoming book [27]. In order to avoid too large an overlap with these existing works, the present contribution, which was invited by the organizers of the International Congress of Mathematicians 2022, to whom the author is most grateful, focuses only on one single aspect of these inequalities. Namely, we will consider Lieb–Thirring inequalities from the point of view of Sobolev-type inequalities for systems of orthonormal functions, and we discuss recent extensions, in particular, to the Strichartz and Stein–Tomas inequalities from harmonic analysis. We will also briefly comment on selected applications of these newly obtained bounds.

1.1. The general setup

Let $\mathcal{H}$ be a (typically complex) Hilbert space with norm denoted by $\|\cdot\|$ and let X be a measure space, with measure denoted simply by dx and with corresponding Lebesgue spaces $L^q(X)$. Assume that T is a bounded linear operator from $\mathcal{H}$ to $L^q(X)$ for some $q > 2$. That is, for all $f \in \mathcal{H}$,

$$\int_X |Tf|^q \, dx \lesssim \|f\|^q. \tag{1.1}$$

As a consequence, if $f_1, \dots, f_N$ are normalized in $\mathcal{H}$, then

$$\int_X \left(\sum_{n=1}^N |Tf_n|^2 \right)^{q/2} dx \lesssim N^{q/2}.$$

This is a consequence of (1.1) and the triangle inequality in $L^{q/2}$. The power $N^{q/2}$ is best possible, as can be seen by taking all f_n to be equal.

The question that interests us here is whether for a given operator T there is a power

$$\alpha < q/2$$

such that for all N and all $f_1, \dots, f_N \in \mathcal{H}$ satisfying the *orthonormality constraint*

$$(f_n, f_m) = \delta_{n,m} \quad \text{for all } 1 \le n, m \le N$$

one has

$$\int_X \left(\sum_{n=1}^N |Tf_n|^2 \right)^{q/2} dx \lesssim N^{\alpha}. \tag{1.2}$$

As explained, for instance, in [19,50,52], such bounds, if true, have important consequences in the mathematical physics of large fermionic quantum systems, density functional theory, and the theory of nonlinear evolution equations. Their study is also interesting from a purely analytical point of view and reveals aspects of the underlying operator T which go beyond its boundedness.

At the moment there is no general principle that determines the exponent α directly from the operator T. Rather, inequalities of the form (1.2), if true, have been verified on a case by case basis. Most of the existing results concern the case where T is (at least approximately) translation invariant. Finding a regime of orthonormal functions $f_1, \dots, f_N$ with $N \to \infty$ where the power α in the bound (1.2) is saturated often relies on techniques of semiclassical and microlocal analysis.

1.2. Example: the HLS inequality

The above principle is most clearly illustrated on the example of the Hardy–Littlewood–Sobolev (HLS) inequality, also know as the weak Young inequality or the theorem of fractional integration; see, e.g., [51, THEOREM 4.3]. A particular case of this inequality states that for $0 < s < d/2$ the operator of convolution with $|x|^{-d+s}$ is bounded from $L^2(\mathbb{R}^d)$ to $L^q(\mathbb{R}^d)$ with $q = 2d/(d-2s)$.

Its extension to systems of orthonormal functions is due to Lieb [48] and reads as follows.

Theorem 1. *Let* $0 < s < d/2$. *Then, if* $f_1, \dots, f_N$ *are orthonormal in* $L^2(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \left(\sum_{n=1}^N \big| |x|^{-d+s} * f_n \big|^2 \right)^{d/(d-2s)} dx \lesssim N.$$

Remarks 2. (a) The power 1 of N on the right side is best possible.

(b) The bound is equivalent (in a certain weak sense) to the Cwikel–Lieb–Rozenblum (CLR) bound

$$N\big((-\Delta)^s + V\big) \lesssim \int_{\mathbb{R}^d} V_-^{d/(2s)} dx$$

on the number $N((-\Delta)^s + V)$ of negative eigenvalues of the generalized Schrödinger operator $(-\Delta)^s + V$ in $L^2(\mathbb{R}^d)$. Here $V(x)_- = \max\{-V(x), 0\}$. The meaning of "equivalent" will be explained in the next subsection. It is a "weak" form of equivalence, because this argument does *not* mean that the sharp constant in Theorem 1 is in one-to-one correspondence with the sharp constant in the CLR bound. This is in contrast to a form of duality that we will encounter later.

(c) The proof of Theorem 1 in [48] proceeds by reducing it to Cwikel's theorem [11]. Alternative, direct proofs of Theorem 1 were given in [17, 58]. We present a different, unpublished proof in Subsection 1.5 below.

(d) Just like the HLS inequality, the bound in Theorem 1 is conformally invariant. This leads to a natural conjecture for its optimal constant [17].

1.3. The duality argument

Let us return to the general setting described in Subsection 1.1 and consider a bounded operator $T : \mathcal{H} \to L^q(X)$ for some $q > 2$. By Hölder's inequality, this boundedness is equivalent to having, for any $W \in L^{2q/(q-2)}(X)$ and any $f \in \mathcal{H}$,

$$\int_X |W|^2 |Tf|^2 \, dx \lesssim \|W\|_{2q/(q-2)}^2 \|f\|^2,$$

which, in turn, is equivalent to the boundedness of the operator WT from $\mathcal{H}$ to $L^2(X)$ with norm

$$\|WT\| \lesssim \|W\|_{2q/(q-2)}.$$

Here, as usual, we do not distinguish in the notation between the function W and the operator of multiplication by W. Moreover, $\|\cdot\|$ on the left side denotes the operator norm.

Let us now reformulate the desired inequality (1.2) in terms of the operator WT. We assume that $\alpha < q/2$. Again by Hölder's inequality, we see that (1.2) is equivalent to

$$\sum_{n=1}^{N} \int_X |W|^2 |Tf_n|^2 \, dx \lesssim N^{2\alpha/q} \|W\|_{2q/(q-2)}^2. \tag{1.3}$$

In order to state this previous inequality succinctly, we recall the notion of Schatten spaces. Background can be found, for instance, in [36,62]. For a compact operator K between two Hilbert spaces, we denote by $(s_n(K))_{n\in\mathbb{N}}$ the sequence of its singular values, that is, the square roots of the eigenvalues of the operator K^*K in nonincreasing order, repeated according to multiplicities. Then, by definition, for any $0 < r < \infty$, the Schatten class $\mathcal{S}^r$ consists of all compact operators K with $s.(K) \in \ell^r$. This is a normed linear space with respect to

$$\|K\|_r := \left(\sum_{n\in\mathbb{N}} s_n(K)^r\right)^{1/r}.$$

Also, we will need the weak variant of this space, $\mathcal{S}^r_{\text{weak}}$, consisting of all compact K with $s.(K) \in \ell^r_{\text{weak}}$. For $2 < r < \infty$, the following norm will appear naturally in our analysis:

$$\|K\|_{r,\mathrm{w}} := \sup_{N\in\mathbb{N}} N^{-1/2+1/r} \left(\sum_{n=1}^{N} s_n(K)^2\right)^{1/2}.$$

It follows from the variational principle for sums of eigenvalues that

$$\sum_{n=1}^{N} s_n(K)^2 = \sup\left\{\sum_{n=1}^{N} \|Kf_n\|^2 : f_1, \ldots, f_N \text{ orthonormal}\right\}.$$

From this and the triangle inequality in $\mathbb{R}^N$ it follows that $\|\cdot\|_{r,\mathrm{w}}$ defines, indeed, a norm. It is also easy to see that $\|\cdot\|_{r,\mathrm{w}}$ is equivalent to the more standard quasinorm in $\mathcal{S}^r_{\text{weak}}$ defined by

$$\|K\|'_{r,\mathrm{w}} := \sup_{n\in\mathbb{N}} n^{1/r} s_n(K).$$

The constants in this equivalence depend on $r > 2$ and their explicit values can be found, for instance, in [17, LEMMA 2.3], where another expression for $\|K\|_{r,\mathrm{w}}$ is used.

Returning to the above setting, we now see that (1.3) is equivalent to the fact that WT belongs to the weak Schatten class $\mathcal{S}^{2q/(q-2\alpha)}_{\text{weak}}$ with

$$\|WT\|_{2q/(q-2\alpha),\mathrm{w}} \lesssim \|W\|_{2q/(q-2)}. \tag{1.4}$$

To summarize, we have seen that the desired inequality (1.2) is equivalent to a quantitative compactness property of the operator WT, expressed in terms of a weak Schatten norm. The exponent α in (1.2) is in one-to-one correspondence with the Schatten exponent. What we have gained through this reformulation is, for instance, that we can use interpolation methods to prove trace ideal properties of the operators TW.

At this point we can present Lieb's proof of Theorem 1. Namely, Cwikel's theorem [11] says that, for $2 < p < \infty$,

$$\left\| a(X)b(-i\nabla) \right\|_{p,\mathrm{w}} \lesssim \|a\|_p \|b\|_{p,\mathrm{w}}. \tag{1.5}$$

Here $a(X)$ denotes the operator of multiplication by a function $a \in L^p(\mathbb{R}^d)$ in position space and $b(-i\nabla)$ denotes the operator of multiplication by a function $b \in L^p_{\mathrm{weak}}(\mathbb{R}^d)$ in momentum space. The operator T relevant for Theorem 1 is convolution with $|x|^{-d+s}$ which corresponds to multiplication by (a constant times) $|\xi|^{-s}$ in Fourier space. The latter function belongs to $L^{d/s}_{\mathrm{weak}}(\mathbb{R}^d)$. Thus, (1.5) implies (1.4) with $\alpha = 1$ and $q = 2d/(d-2s)$, as claimed.

The proof of Cwikel's theorem in [17] goes in some sense the other way around. Namely, first Theorem 1 (or rather a slight generalization of it) is established, using the method of [58], and then the above duality argument is used to deduced (1.5).

We can now also explain the notion of weak equivalence in Remark 2 (b). Namely, by the Birman–Schwinger principle, the bounds there for negative eigenvalues of generalized Schrödinger operators are the same as bounds on the operator $W(-\Delta)^{-s/2}$ in the quasinorm $\|\cdot\|'_{d/s,\mathrm{w}}$, whereas by the above argument the bound in Theorem 1 is the same as a bound on this operator in the norm $\|\cdot\|_{d/s,\mathrm{w}}$.

1.4. A generalization

There is a far reaching generalization of Theorem 1. Namely, if X is a sigma-finite measure space and A is a nonnegative operator in $L^2(X)$ with heat semigroup satisfying, for some $\nu > 2$,

$$\left\| \exp(-tA) \right\|_{L^2 \to L^\infty} \lesssim t^{-\nu/4} \quad \text{for all } t > 0,$$

then for all $u_1, \dots, u_N \in \operatorname{dom} A^{1/2}$ satisfying $(A^{1/2}u_n, A^{1/2}u_m) = \delta_{n,m}$ for $1 \le n, m \le N$,

$$\int_X \left(\sum_{n=1}^N |u_n|^2 \right)^{\nu/(\nu-2)} dx \lesssim N.$$

This is shown in [17], improving earlier results in [30, 45] that require nonnegativity of the heat kernel.

This more general result reduces to Theorem 1 with $A = (-\Delta)^s$ and $u_n =$ a constant times $(-\Delta)^{-s/2} f_n$. Another application concerns the case where A is the Laplace–Beltrami operator on certain noncompact manifolds. For a compact manifold, the above assumption on the semigroup is not satisfied because of the zero eigenvalue, but one can add a positive constant to the Laplace–Beltrami operator.

1.5. Appendix: proof of Theorem 1

We present here an unpublished proof of Theorem 1. It is neither the most elementary one, nor one giving particularly good constants, but we think it is conceptually rather clear and might allow for interesting generalizations. In view of the previous subsection it provides an alternative proof of the CLR inequality and is based on some ideas of Conlon [9].

By the duality argument in Subsection 1.3, we need to prove (1.4) with T equal to convolution with $|x|^{-d+s}$, $\alpha = 1$, and $q = 2d/(d-2s)$. Since the weak Schatten norm of WT equals that of $(WT)^* = T\overline{W}$, it suffices to consider the latter operator. We have, using $\int |x-z|^{-d+s}|z-y|^{-d+s}\,dz = \text{const}\,|x-y|^{-d+2s}$ and the Fefferman–de la Llave decomposition [13],

$$\begin{aligned}\|T\overline{W}f\|^2 &= \text{const} \iint_{\mathbb{R}^d\times\mathbb{R}^d} \frac{\overline{f(x)}\,W(x)\,\overline{W(y)}\,f(y)}{|x-y|^{d-2s}}\,dx\,dy \\ &= \text{const} \int_{\mathbb{R}^d} da \int_0^\infty \frac{dr}{r^{2d+1-2s}} \iint_{B_r(a)\times B_r(a)} \overline{f(x)}\,W(x)\,\overline{W(y)}\,f(y)\,dx\,dy.\end{aligned}$$

We apply this with $f = f_n$ for some orthonormal f_n in $L^2(\mathbb{R}^d)$ and sum over n. For fixed a and r, we estimate the double integral over x and y in two different ways. First, since the operator γ with kernel $\sum_n \overline{f_n(x)}f_n(y)$ has operator norm one, we have

$$\begin{aligned}\left|\sum_{n=1}^N \iint_{B_r(a)\times B_r(a)} \overline{f_n(x)}\,W(x)\,\overline{W(y)}\,f_n(y)\,dx\,dy\right| &= \left|(\overline{W}\mathbb{1}_{B_r(a)}, \gamma\overline{W}\mathbb{1}_{B_r(a)})\right| \\ &\le \int_{B_r(a)} \left|W(x)\right|^2 dx \\ &\lesssim r^d\left(M\left(|W|^2\right)\right)(a),\end{aligned}$$

where M is the maximal function. Second, since $\gamma \ge 0$,

$$\left|\sum_{n=1}^N \overline{f_n(x)}f_n(y)\right| \le \sqrt{\rho(x)}\sqrt{\rho(y)}, \quad \text{where } \rho(x) := \sum_{n=1}^N \left|f_n(x)\right|^2,$$

so

$$\begin{aligned}\left|\sum_{n=1}^N \iint_{B_r(a)\times B_r(a)} \overline{f_n(x)}\,W(x)\,\overline{W(y)}\,f_n(y)\,dx\,dy\right| &\le \left(\int_{B_r(a)} \sqrt{\rho(x)}\,\left|W(x)\right|dx\right)^2 \\ &\lesssim r^{2d}\left(M\left(|W|\sqrt{\rho}\right)\right)(a)^2.\end{aligned}$$

Inserting this into the above formula, we find

$$\begin{aligned}\sum_{n=1}^N \|T\overline{W}f_n\|^2 &\lesssim \int_{\mathbb{R}^d} da \int_0^\infty \frac{dr}{r^{2d+1-2s}} \min\{r^d\left(M\left(|W|^2\right)\right)(a), r^{2d}\left(M\left(|W|\sqrt{\rho}\right)\right)(a)^2\} \\ &= \text{const} \int_{\mathbb{R}^d} da\,\left(M\left(|W|\sqrt{\rho}\right)\right)(a)^{2(1-2s/d)}\left(M\left(|W|^2\right)\right)(a)^{2s/d} \\ &\le \text{const}\left(\int_{\mathbb{R}^d} \left(M\left(|W|\sqrt{\rho}\right)\right)(a)^{2d/(d+2s)}da\right)^{1-(2s/d)^2} \\ &\quad\times\left(\int_{\mathbb{R}^d} \left(M\left(|W|^2\right)\right)(a)^{d/(2s)}\,da\right)^{(2s/d)^2}.\end{aligned}$$

By the boundedness of the maximal function on L^p, $1 < p < \infty$, this is bounded by a constant times

$$\left(\int_{\mathbb{R}^d} |W|^{2d/(d+2s)} \rho^{d/(d+2s)} da\right)^{1-(2s/d)^2} \left(\int_{\mathbb{R}^d} |W|^{d/s}\, da\right)^{(2s/d)^2}$$
$$\leq \left(\int_{\mathbb{R}^d} |W|^{d/s}\, da\right)^{2s/d} \left(\int_{\mathbb{R}^d} \rho\, da\right)^{1-2s/d}.$$

To summarize, we have shown that

$$\sum_{n=1}^{N} \|T\overline{W} f_n\|^2 \lesssim \|W\|_{d/s}^2 N^{1-2s/d}.$$

If we take the f_n to be the eigenfunctions of $WT^2\overline{W}$ corresponding to its N largest eigenvalues, the previous inequality becomes

$$\sum_{n=1}^{N} s_n(T\overline{W})^2 \lesssim \|W\|_{d/s}^2 N^{1-2s/d}.$$

This is the claimed bound on $T\overline{W}$ in the Schatten space $\mathcal{S}^{d/s}_{\text{weak}}$.

2. SOBOLEV-TYPE INEQUALITIES FOR ORTHONORMAL FUNCTIONS

Before turning to the more recent bounds related to Fourier restriction, in this section we review some classical inequalities for orthonormal functions that are related to Sobolev inequalities. Those include, in particular, the classical Lieb–Thirring inequality in Theorem 4 below.

2.1. Bessel-potential bounds

The bounds in Theorem 1 concern $|x|^{-d+s} * f$, which is a constant multiple of the Riesz potential $(-\Delta)^{-s/2} f$. We present a generalization, due to Lieb [48], of these bounds to the Bessel potentials $(-\Delta + m^2)^{-s/2} f$ with $m > 0$.

Theorem 3. *Let $s > 0$ and let*

$$\begin{cases} 2 \leq q \leq \infty & \text{if } s > d/2, \\ 2 \leq q < \infty & \text{if } s = d/2, \\ 2 \leq q \leq 2d/(d-2s) & \text{if } s < d/2. \end{cases}$$

Then, if $f_1, \dots, f_N$ are orthonormal in $L^2(\mathbb{R}^d)$,

$$\left\| \sum_{n=1}^{N} \left| (-\Delta + m^2)^{-s/2} f_n \right|^2 \right\|_{q/2} \lesssim m^{d-2s-2d/q} N^{2/q}.$$

The bound for $q = 2d/(d-2s)$ if $s < d/2$ follows as before using Cwikel's theorem (1.5). The remaining bounds follow similarly, but using the simpler bound

$$\left\| a(X) b(-i\nabla) \right\|_p \leq (2\pi)^{-d/p} \|a\|_p \|b\|_p \tag{2.1}$$

for $2 \leq p \leq \infty$. The latter bound is due to Kato, Seiler, and Simon (see, e.g., [62, THEOREM 4.1]) and can also be inferred from the Lieb–Thirring matrix inequality [54].

Since (2.1), in contrast to (1.5), involves a strong instead of a weak Schatten norm, a generalization of the bound in Theorem 3 to sums of the form $\sum_n \nu_n |(-\Delta + m^2)^{-s/2} f_n|^2$ is possible provided, if $s > d/2$, $q < 2d/(d-2s)$. We discuss this in the next section.

Using bounds due to Solomyak [64] (and their natural extension to odd dimensions) it seems plausible that in the case $s = d/2$ there is an endpoint bound in the Orlicz space $\exp L$ in the spirit of a Moser–Trudinger inequality. For instance, the bounds in [24] can be dualized to yield that, if $\Omega \subset \mathbb{R}^2$ is open and of finite measure, then for any $u_1, \dots, u_N \in H_0^1(\Omega)$ satisfying $\int_\Omega \nabla \overline{u_n} \cdot \nabla u_m \, dx = \delta_{n,m}$ for all $1 \leq n, m \leq N$,

$$\int_\Omega \mathcal{A}\left((C L_N)^{-1} \sum_{n=1}^N |u_n|^2 \right) dx \leq |\Omega|,$$

where $\mathcal{A}(t) = e^t - 1 - t$, $L_N = \sum_{n=1}^N n^{-1}$, and where C is a universal constant.

2.2. The Lieb–Thirring inequality

The original LT inequality in its form for orthonormal functions reads as follows.

Theorem 4. *Let $d \geq 1$ and $s > 0$. Then, if $u_1, \dots, u_N \in H^s(\mathbb{R}^d)$ are orthonormal in $L^2(\mathbb{R}^d)$,*

$$\sum_{n=1}^N \int_{\mathbb{R}^d} \left|(-\Delta)^{s/2} u_n\right|^2 dx \gtrsim \int_{\mathbb{R}^d} \left(\sum_{n=1}^N |u_n|^2 \right)^{1+2s/d} dx.$$

Remarks 5. (a) The main point is that the implicit constant can be chosen independently of N.

(b) The bound is equivalent to the bound

$$\sum_j |E_j| \lesssim \int_{\mathbb{R}^d} V_-^{1+d/(2s)} dx$$

on the sum of the negative eigenvalues (counted with multiplicities) of the generalized Schrödinger operator $(-\Delta)^s + V$ in $L^2(\mathbb{R}^d)$. This equivalence is, for instance, in the sense that the sharp constants in the two inequalities are in one-to-one correspondence.

(c) It is important in applications that the inequality in Theorem 4 extends to density matrices, namely, for any sequence $0 \leq \nu \in \ell^\infty$,

$$\sum_n \nu_n \int_{\mathbb{R}^d} \left|(-\Delta)^{s/2} u_n\right|^2 dx \gtrsim \left(\sup_n \nu_n \right)^{-2s/d} \int_{\mathbb{R}^d} \left(\sum_n \nu_n |u_n|^2 \right)^{1+2s/d} dx.$$

(d) Theorem 4 in $d = 3$ with $s = 1$ is due to Lieb and Thirring [53] and was the crucial ingredient in their proof of stability of matter; see also [52]. Their proof of Theorem 1 in [54] for $s = 1$ extends to general s. Alternative proofs are due to [59], [56] and [61].

(e) Lieb and Thirring [54] made a famous conjecture about the optimal constant in the inequality in Theorem 4 for $s = 1$; see, for instance, [19] for details. This predicts, in particular, that there is a fundamental difference between dimensions $d \leq 2$ and $d \geq 3$. This conjecture is *open* in any dimension.

(f) The currently best constants in Theorem 4 are due to [23]. A bound with "almost" the semiclassical constant and a gradient remainder term appears in [57].

(g) As a step towards the Lieb–Thirring conjecture, one can study the best constant in the inequality in Theorem 4 with fixed N. For $s = 1$, it is shown in [20] that in dimensions $d \geq 3$ this constant is always *strictly less* than the optimal constant that works for arbitrary N. This is consistent with the Lieb–Thirring conjecture. For further results in this direction, see also [21,22].

In the spirit of the generalization discussed in Subsection 1.4, Theorem 4 has been extended to abstract operators satisfying certain heat kernel bounds or Sobolev inequalities; see [30].

2.3. A more general Lieb–Thirring inequality

The following theorem provides a Sobolev inequality with exponent q less than $2(1 + 2s/d)$, the exponent in Theorem 4. The bound is deduced in [55] via a duality argument from a bound of Lieb and Thirring [54]. Note that the functions here are orthogonal, not necessarily orthonormal.

Theorem 6. *Let $d \geq 1$, $s > 0$ and $2 < q < 2(1 + 2s/d)$. Then, if $u_1, \dots, u_N \in H^s(\mathbb{R}^d)$ are orthogonal in $L^2(\mathbb{R}^d)$,*

$$\sum_{n=1}^{N} \int_{\mathbb{R}^d} \left|(-\Delta)^{s/2} u_n\right|^2 dx$$

$$\gtrsim \left(\sum_{n=1}^{N} \|u_n\|_2^{\frac{2(2d-(d-2s)q)}{2d+4s-dq}} \right)^{-\frac{2d+4s-dq}{d(q-2)}} \left(\int_{\mathbb{R}^d} \left(\sum_{n=1}^{N} |u_n|^2 \right)^{\frac{q}{2}} dx \right)^{\frac{4s}{d(q-2)}}.$$

Remarks 7. (a) The implicit constant can be chosen independently of N.

(b) The bound is equivalent to the bound

$$\sum_j |E_j|^\gamma \lesssim \int_{\mathbb{R}^d} V_-^{\gamma + d/(2s)} dx$$

on the sum of the negative eigenvalues (counted with multiplicities) of the generalized Schrödinger operator $(-\Delta)^s + V$ in $L^2(\mathbb{R}^d)$. Here $\gamma > 1$ and $q < 2(1 + 2s/d)$ are related by

$$q = \frac{2(\gamma + \frac{d}{2s})}{\gamma + \frac{d}{2s} - 1}, \quad \gamma = \frac{2d - (d-2s)q}{2s(q-2)}.$$

(c) For $s = 1$, Lieb and Thirring [54] made a famous conjecture about the optimal constant in the eigenvalue inequality in (b), which translates into a conjecture for the constant in Theorem 4. This conjecture was proved by Laptev and Weidl [44] for $\gamma \geq 3/2$, that is, $q \leq 2(d+3)/(d+1)$.

(d) The analysis mentioned in Remark 5 (f) concerning truncated versions of the inequality is applicable as well in the situation of Theorem 6 with $s = 1$; see [20–22].

3. FOURIER RESTRICTION INEQUALITIES FOR ORTHONORMAL FUNCTIONS

We now turn to inequalities for systems of orthonormal functions that are mathematically related to the question of restricting the Fourier transform to hypersurfaces. Such a restriction is possible under certain curvature assumptions on the hypersurface and has important applications to partial differential equations.

3.1. Strichartz inequality for orthonormal functions

The Strichartz inequality [39, 66] concerns solutions $e^{it\Delta}\psi$ of the free Schrödinger equation and quantifies their dispersive behavior. It states that if $d \geq 1$, $2 \leq q \leq \infty$ if $d = 1$, $2 \leq q < \infty$ if $d = 2$, and $2 \leq q \leq 2d/(d-2)$ if $d \geq 3$, and $2/p + d/q = d/2$, then for all $\psi \in L^2(\mathbb{R}^d)$,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} \left| e^{it\Delta}\psi \right|^q dx \right)^{p/q} dt \lesssim \|\psi\|_2^p. \tag{3.1}$$

Here is a version of this inequality for systems of orthonormal functions.

Theorem 8. *Let $d \geq 1$, $2 \leq q < 2(d+1)/(d-1)$ and $2/p + d/q = d/2$. Then, if $\psi_1, \dots, \psi_N \in L^2(\mathbb{R}^d)$ are orthonormal in $L^2(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} \left(\sum_{n=1}^{N} \left| e^{it\Delta}\psi_n \right|^2 \right)^{q/2} dx \right)^{p/q} dt \lesssim N^{p(q+2)/(4q)}.$$

Remarks 9. (a) The power of N is best possible, as can be deduced from [28].

(b) The bound in Theorem 8 can be slightly improved, namely, for any sequence $0 \leq \nu \in \ell^{2q/(q+2)}$,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} \left(\sum_n \nu_n \left| e^{it\Delta}\psi_n \right|^2 \right)^{q/2} dx \right)^{p/q} dt \lesssim \left(\sum_n \nu_n^{2q/(q+2)} \right)^{p(q+2)/(4q)}.$$

The bound in the theorem corresponds to the case $\nu_n \in \{0, 1\}$ and is equivalent to a bound for ν in the Lorentz space $\ell^{2q/(q+2),1}$. It is remarkable that, while in the extension of the bound in Theorem 1 the Lorentz space $\ell^{s,1}$ is optimal, here it can be improved to the space ℓ^s. The assumption $\nu \in \ell^{2q/(q+2)}$ cannot be relaxed to $\nu \in \ell^s$ for any $s > 2q/(q+2)$ [28].

(c) Theorem 8 appears in [28] for $q \leq 2(d+2)/d$ and in [31] in the full range. The proof in [31] uses a duality argument, similarly to that in Subsection 1.3. In fact, it is slightly simpler, since the duality between $\mathcal{S}^r$ and $\mathcal{S}^{r'}$ is more straightforward than that between $\mathcal{S}^r_{\text{weak}}$ and the Lorentz space $\mathcal{S}^{r',1}$, which is at the core of Subsection 1.3. On the other hand, the fact that here we work in a mixed norm space $L_t^{p/2} L_x^{q/2}$ does not really complicate the argument.

(d) It is conjectured in [28] that Theorem 8 remains valid for $q = 2(d+1)/(d-1)$. At the same time it is shown there that the strengthening in (b) with the $\ell^{2q/(q+2)}$-norm fails at $q = 2(d+1)/(d-1)$. This conjecture was disproved in [2] in dimension $d = 1$, but is still *open* for $d \geq 2$.

(e) There is a "semiclassical" version of the inequality where the Schrödinger equation is replaced by a transport equation for densities on the phase space. The proof in [28] can be adapted to this setting, as shown in [1]. For more on the connection between the two equations, see [60]. The disproof of the conjecture mentioned in (d) for $d = 1$ was by disproving the corresponding conjecture in this simpler setting. It uses the existence of a Kakeya set of arbitrary small measure. The validity of this analogue conjecture for $d \geq 2$ is still *open*.

(f) There is a natural "one-particle constant," namely the sharp constant in (3.1). This was determined in the diagonal case $q = p$ in [14] for $d = 1, 2$. Besides, there is a semiclassical constant related to the inequality in (e). To which extent these two constants play a role for the sharp constant in Theorem 8, in analogy with the Lieb–Thirring conjecture, has not been investigated.

The restriction $q < 2(d+1)/(d-1)$ in Theorem 8 is not present for the single function inequality (3.1). It is known that for orthonormal functions the case $q \geq 2(d+1)/(d-1)$ behaves differently, but there are several open questions. The following is known.

Theorem 10. *Let $d \geq 2$, $2(d+1)/(d-1) \leq q < 2d/(d-2)$, and $2/p + d/q = d/2$. Let $\beta < 2q/(d(q-2))$. Then, if $\psi_1, \dots, \psi_N \in L^2(\mathbb{R}^d)$ are orthonormal in $L^2(\mathbb{R}^d)$,*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} \left(\sum_{n=1}^N \left| e^{it\Delta} \psi_n \right|^2 \right)^{q/2} dx \right)^{p/q} dt \lesssim N^{p/(2\beta)}.$$

Remarks 11. (a) It is known that the bound in the theorem does not hold with an exponent $\beta > 2q/(d(q-2))$, as can be deduced from [33], but it is not known whether or not it holds with exponent $\beta = 2q/(d(q-2))$.

(b) Similarly as in the case of Theorem 8, the bound in Theorem 10 can be slightly improved, namely, for any sequence $0 \leq \nu \in \ell^\beta$ with $\beta < 2q/(d(q-2))$,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}^d} \left(\sum_n \nu_n \left| e^{it\Delta} \psi_n \right|^2 \right)^{q/2} dx \right)^{p/q} dt \lesssim \left(\sum_n \nu_n^\beta \right)^{p/(2\beta)}.$$

The bound is known to fail, in general, if $\nu \in \ell^\beta$ for $\beta > 2q/(d(q-2))$ [33].

(c) Theorem 10 appears in [33] (see the discussion there after Proposition 1). It is obtained by interpolation between Theorem 8 with q near $2(d+1)/(d-1)$ and the bound (3.1) with q near its maximal value $2d/(d-2)$.

(d) If the conjecture mentioned in Remark 9 (d) is true, an interpolation argument (at least in dimensions $d \geq 3$) might yield Theorem 10 with $\beta = 2q/(d(q-2))$.

(e) Let us discuss the endpoints $q = 2d/(d-2)$ in $d \geq 3$ and $q = \infty$ in $d = 1$, which are excluded in Theorem 10. At the endpoint $q = 2d/(d-2)$ in $d \geq 3$, it is known that there is no gain due to orthonormality over the triangle inequality, that is, the bound in (b) holds with $\beta = 1$ and not with any larger power [33]. At the endpoint $q = \infty$, in $d = 1$ it is known that the bound in (b) does not hold for $\beta \geq 2$ (see [28] and also [33, PROPOSITION 1]) and one may wonder whether it holds for $\beta < 2$. In [3] it is shown that the bound holds for $\beta \leq 4/3$ and that the slightly weaker inequality with the $L_t^{2,\infty} L_x^\infty$-norm instead of the $L_t^2 L_x^\infty$-norm holds for all $\beta < 2$.

Strichartz inequalities for orthonormal system have been proved for more general operators than $-\Delta$ (see, e.g., [4,31]) and for more regular functions (see, e.g., [2–4]).

One application of the Strichartz inequality for orthonormal functions concerns the nonlinear, time-dependent Hartree equation for infinite quantum systems. (Here "infinite" means that the initial data are allowed to have infinite trace.) Using Theorem 8, one can show global well-posedness and, for small initial data, dispersion for large time; see [31,60]. For the more involved case of a positive background density, see [46,47].

Another application of the Strichartz inequality for orthonormal functions concerns Besov-space improvements of inequality (3.1) for single functions; see [31, COROLLARY 9]. While these bounds can be derived using deep results from bilinear restriction theory, it is interesting to note that the proof via Theorem 8 is much more elementary.

3.2. Stein–Tomas inequality for orthonormal functions

The Fourier restriction problem is whether the Fourier transform of a function on $\mathbb{R}^d$ has a well-defined restriction to a hypersurface and, if so, to establish corresponding L^p bounds. Sometimes it is helpful to study the equivalent, adjoint problem of Fourier extensions. From a harmonic analysis perspective, the Strichartz inequality corresponds to a Fourier extension inequality for the hypersurface $\{(\xi, -|\xi|^2) : \xi \in \mathbb{R}^d\}$ in $\mathbb{R}^{d+1}$ endowed with a natural measure. Another paradigmatic case concerns the Fourier extension for the sphere. The corresponding result, due to Tomas [67] and Stein [65], states that, if $f \in L^2(\mathbb{S}^{d-1})$, then

$$\int_{\mathbb{R}^d} \left| \int_{\mathbb{S}^{d-1}} e^{i\omega\cdot x} f(\omega)\, d\omega \right|^{2(d+1)/(d-1)} dx \lesssim \|f\|_{L^2(\mathbb{S}^{d-1})}^{2(d+1)/(d-1)}. \tag{3.2}$$

The inequality extends trivially to exponents greater than $2(d+1)/(d-1)$ on the left side, but a counterexample due to Knapp shows that $2(d+1)/(d-1)$ is the smallest possible exponent. Here is a version for orthonormal functions.

Theorem 12. *Let $d \geq 2$. Then, if $f_1, \ldots, f_N$ are orthonormal in $L^2(\mathbb{S}^{d-1})$,*

$$\int_{\mathbb{R}^d} \left(\sum_{n=1}^{N} \left| \int_{\mathbb{S}^{d-1}} e^{i\omega \cdot x} f_n(\omega)\, d\omega \right|^2 \right)^{(d+1)/(d-1)} dx \lesssim N^{d/(d-1)}.$$

Remarks 13. (a) The power of N is best possible, as can be deduced from [31].

(b) Similarly as Theorem 8, the bound in Theorem 12 can be slightly improved, namely, for any $0 \leq \nu \in \ell^{(d+1)/d}$,

$$\int_{\mathbb{R}^d} \left(\sum_n \nu_n \left| \int_{\mathbb{S}^{d-1}} e^{i\omega \cdot x} f_n(\omega)\, d\omega \right|^2 \right)^{(d+1)/(d-1)} dx \lesssim \left(\sum_n \nu_n^{(d+1)/d} \right)^{d/(d-1)}.$$

The assumption $\nu \in \ell^{(d+1)/d}$ cannot be relaxed to $\nu \in \ell^r$ for any $r > (d+1)/d$ [31].

(c) Theorem 12 appears in [31], where it is proved using a duality argument similarly as in Subsection 1.3.

(d) In analogy to Remark 9 (f), the optimal constant in (3.2) is only known for $d = 3$ [15]; see also [29] for a connection between the optimal constants in (3.2) and (3.1). As far as we know, a "semiclassical inequality" corresponding to that in Theorem 12 has not been investigated.

One application of Theorem 12 concerns trace ideal bounds for the scattering matrix for Schrödinger operators $-\Delta + V$ in $L^2(\mathbb{R}^d)$ [31]. These bounds are universal in the sense that they only depend on the "energy" parameter and an L^p norm of V, and the trace ideal exponent is shown to be optimal.

To motivate the discussion in the following subsection, we note that by the duality argument in Subsection 1.3 and by scaling, the bound in Theorem 12 (or rather in Remark 13 (b)) can be written as

$$\|\mathcal{R}_k W\|_{2(d+1)} \lesssim k^{\frac{d-1}{2(d+1)}} \|W\|_{d+1},$$

where $\mathcal{R}_k$ denotes restriction of the Fourier transform to the sphere $\{|\xi| = k\}$. Integrating this bound with respect to k between λ and $\lambda + 1$, we obtain, in terms of the spectral projection $\Pi_\lambda = \mathbb{1}(\lambda^2 \leq -\Delta \leq (\lambda+1)^2)$ with $\lambda \geq 1$,

$$\left\| \Pi_\lambda |W|^2 \Pi_\lambda \right\|_{d+1} = \|\overline{W} \Pi_\lambda W\|_{d+1} \leq \int_\lambda^{\lambda+1} \|\overline{W} \mathcal{R}_k W\|_{d+1}\, dk \lesssim \lambda^{\frac{d-1}{d+1}} \|W\|_{d+1}^2.$$

Dualizing back, we find that if (f_n) are orthonormal in $L^2(\mathbb{R}^d)$ and satisfy $\operatorname{supp} \hat{f}_n \subset \{\lambda \leq |\xi| \leq \lambda + 1\}$ with $\lambda \geq 1$ and if $0 \leq \nu \in \ell^{(d+1)/d}$, then

$$\left(\int_{\mathbb{R}^d} \left(\sum_n \nu_n |f_n|^2 \right)^{\frac{d+1}{d-1}} dx \right)^{\frac{d-1}{d+1}} \lesssim \lambda^{\frac{d-1}{d+1}} \left(\sum_n \nu_n^{\frac{d+1}{d}} \right)^{\frac{d}{d+1}}. \tag{3.3}$$

3.3. Spectral cluster bounds

As shown by Sogge [63], the version (3.3) of the Stein–Tomas inequality has a generalization to closed manifolds. Here is a generalization of this theorem to the case of orthonormal functions from [32].

Theorem 14. *Let (M, g) be a smooth compact Riemannian manifold without boundary of dimension $d \geq 2$. Denote by $-\Delta_g$ the Laplace–Beltrami operator on M and, for any $\lambda \geq 1$, let $\Pi_\lambda := \mathbb{1}(\lambda^2 \leq -\Delta_g < (\lambda+1)^2)$. Then, if $(f_n) \subset \Pi_\lambda L^2(M)$ are orthonormal in $L^2(M)$ and if $(\nu_n) \subset [0, \infty)$,*

$$\Big\| \sum_n \nu_n |f_n|^2 \Big\|_{L^{q/2}(M)} \lesssim \lambda^{2s(q)} \Big(\sum_n \nu_n^{\alpha(q)} \Big)^{1/\alpha(q)},$$

where

$$\begin{cases} s(q) := d(\frac{1}{2} - \frac{1}{q}) - \frac{1}{2}, & \alpha(q) = \frac{q(d-1)}{2d} \quad \text{if } \frac{2(d+1)}{d-1} \leq q \leq \infty, \\ s(q) := \frac{d-1}{2}(\frac{1}{2} - \frac{1}{q}), & \alpha(q) = \frac{2q}{q+2} \quad \text{if } 2 \leq q \leq \frac{2(d+1)}{d-1}. \end{cases}$$

Remarks 15.

(a) If there is a single nonzero ν_n, the bound in Theorem 14 reduces to Sogge's bound [63]. Therefore, according to known results about this inequality, for each (M, g) the exponent $2s(q)$ of λ is best possible. As shown in [32], for each (M, g) the exponent $\alpha(q)$ is also best possible. Moreover, on $\mathbb{S}^2$ with its standard metric it can be shown that the inequality can be saturated even with $\nu_n \in \{0, 1\}$ and an arbitrary prescribed sequence $\#\{n : \nu_n = 1\}$ [32].

(b) The proof of Theorem 14 relies on Schatten norm bounds for oscillatory integral operators satisfying the Carleson–Sjölin condition, which are of independent interest, but somewhat technical to state. They imply, for instance, Theorem 12.

3.4. Kenig–Ruiz–Sogge inequalities

In this final subsection we discuss resolvent bounds that are close in spirit to the Stein–Tomas theorem. The original result due to Kenig, Ruiz, and Sogge [40] states that, if $2d/(d+2) \leq p \leq 2(d+1)/(d+3)$ (and $p > 1$ if $d = 2$), then for all $z \in \mathbb{C} \setminus [0, \infty)$,

$$\big\| (-\Delta - z)^{-1} f \big\|_{p'} \lesssim |z|^{-d/2+d/p-1} \|f\|_p. \tag{3.4}$$

For the case $d = 2$, see, e.g., [16]. We mention that similar inequalities are valid also on Riemannian manifolds; see, e.g., [8,12,34]. A notable feature of the bounds (3.4) is that they do not deteriorate as z approaches the positive real half-line and, for that reason, they are also known as "uniform" Sobolev inequalities. Note that the endpoint exponent $p = 2(d+1)/(d+3)$ is the dual of the exponent in the Stein–Tomas Fourier extension inequality (3.2) and, in fact, (3.2) is an easy consequence of (3.4).

For p greater than this exponent, the uniformity is lost, in general. It can be restored, up to $p = 2d/(d+1)$ by using mixed norms involving an L^2-norm over angular variables [35]. A nonuniform inequality valid for $2(d+1)/(d+3) < p \leq 2$ is

$$\big\| (-\Delta - z)^{-1} f \big\|_{p'} \lesssim |z|^{-(\frac{1}{p}-\frac{1}{2})} \operatorname{dist}(z, \mathbb{R}_+)^{-1+(d+1)(\frac{1}{p}-\frac{1}{2})} \|f\|_p.$$

This bound follows by interpolation between the case $p = 2(d+1)/(d+3)$ and the trivial bound at $p = 2$. It appeared in an equivalent, dual form in [18]. Remarkably, it is best possible [41].

Inequality (3.4) is somewhat different from the others treated in this paper since it does not involve a Hilbert space norm and, since the operator $(-\Delta - z)^{-1}$ for $z \notin (-\infty, 0]$ is not positive definite, it cannot be rewritten in such a form. Consequently, we cannot state a version for orthonormal functions, but we will directly state trace ideal bounds, similar to what is behind the proofs of the other bounds in this paper. The following two theorems are from [31] and [18], respectively.

Theorem 16. *Let $d \geq 2$ and let $8/3 \leq q \leq 3$ if $d = 2$ and $d \leq q \leq d+1$ if $d \geq 3$. Then, for all $z \in \mathbb{C} \setminus [0, \infty)$,*

$$\left\| W_1(-\Delta - z)^{-1} W_2 \right\|_{(d-1)q/(d-q)} \lesssim |z|^{-1+d/q} \|W_1\|_q \|W_2\|_q.$$

Theorem 17. *Let $d \geq 1$ and let $q > d+1$. Then, for all $z \in \mathbb{C} \setminus [0, \infty)$,*

$$\left\| W_1(-\Delta - z)^{-1} W_2 \right\|_q \lesssim |z|^{-1/q} \operatorname{dist}\bigl(z, [0, \infty)\bigr)^{-1+(d+1)/q} \|W_1\|_q \|W_2\|_q.$$

The trace ideal exponent in Theorem 16 is best possible, as follows from the corresponding result for the Stein–Tomas inequality [31]. The optimal form of Theorem 16 for $d = 2$ and $2 < q < 8/3$ is not known and we refer to [31] for some partial results.

The main application of Theorems 16 and 17 is to Lieb–Thirring inequalities for eigenvalues of Schrödinger operators with complex-valued potentials; see, e.g., [18, 31]. This is an active area of research with many open question and we refer, for instance, to [6, 7, 10, 25, 26, 37] for more on this.

FUNDING

Partial support through U. S. NSF grant DMS-1954995 and through Germany's Excellence Strategy EXC-2111-390814868 is acknowledged.

REFERENCES

[1] J. Bennett, N. Bez, S. Gutiérrez, and S. Lee, On the Strichartz estimates for the kinetic transport equation. *Comm. Partial Differential Equations* **39** (2014), no. 10, 1821–1826.

[2] N. Bez, Y. Hong, S. Lee, S. Nakamura, and Y. Sawano, On the Strichartz estimates for orthonormal systems of initial data with regularity. *Adv. Math.* **354** (2019), 106736, 37 pp.

[3] N. Bez, S. Lee, and S. Nakamura, Maximal estimates for the Schrödinger equation with orthonormal initial data. *Selecta Math. (N.S.)* **26** (2020), no. 4, 52, 24 pp.

[4] N. Bez, S. Lee, and S. Nakamura, Strichartz estimates for orthonormal families of initial data and weighted oscillatory integral estimates. *Forum Math. Sigma* **9** (2021), e1, 52 pp.

[5] Ph. Blanchard and J. Stubbe, Bound states for Schrödinger Hamiltonians: phase space methods and applications. *Rev. Math. Phys.* **35** (1996), 504–547.

[6] S. Bögli and F. Stampach, On Lieb–Thirring inequalities for one-dimensional non-self-adjoint Jacobi and Schrödinger operators. *J. Spectr. Theory*, to appear. arXiv:2004.09794.

[7] A. Borichev, R. L. Frank, and A. Volberg, Counting eigenvalues of Schrödinger operator with complex fast decreasing potential. *Adv. Math.*, to appear. arXiv:1811.05591.

[8] J. Bourgain, P. Shao, C. D. Sogge, and X. Yao, On L^p-resolvent estimates and the density of eigenvalues for compact Riemannian manifolds. *Comm. Math. Phys.* **333** (2015), no. 3, 1483–1527.

[9] J. G. Conlon, A new proof of the Cwikel–Lieb–Rosenbljum bound. *Rocky Mountain J. Math.* **15** (1985), no. 1, 117–122.

[10] J.-C. Cuenin, Schrödinger operators with complex sparse potentials. *Comm. Math. Phys.*, to appear. arXiv:2102.12706.

[11] M. Cwikel, Weak type estimates for singular values and the number of bound states of Schrödinger operators. *Ann. of Math. (2)* **106** (1977), no. 1, 93–100.

[12] D. Dos Santos Ferreira, C. E. Kenig, and M. Salo, On L^p resolvent estimates for Laplace–Beltrami operators on compact manifolds. *Forum Math.* **26** (2014), no. 3, 815–849.

[13] C. L. Fefferman and R. de la Llave, Relativistic stability of matter. I. *Rev. Mat. Iberoam.* **2** (1986), 119–161.

[14] D. Foschi, Maximizers for the Strichartz inequality. *J. Eur. Math. Soc. (JEMS)* **9** (2007), no. 4, 739–774.

[15] D. Foschi, Global maximizers for the sphere adjoint Fourier restriction inequality. *J. Funct. Anal.* **268** (2015), no. 3, 690–702.

[16] R. L. Frank, Eigenvalue bounds for Schrödinger operators with complex potentials. *Bull. Lond. Math. Soc.* **43** (2011), no. 4, 745–750.

[17] R. L. Frank, Cwikel's theorem and the CLR inequality. *J. Spectr. Theory* **4** (2014), no. 1, 1–21.

[18] R. L. Frank, Eigenvalue bounds for Schrödinger operators with complex potentials. III. *Trans. Amer. Math. Soc.* **370** (2018), no. 1, 219–240.

[19] R. L. Frank, The Lieb–Thirring inequalities: recent results and open problems. In *Nine mathematical challenges: an Elucidation*, edited by A. Kechris et al., Proc. Symp. Pure Math. Amer. Math. Soc. 104, Amer. Math. Soc., Providence, RI, 2021.

[20] R. L. Frank, D. Gontier, and M. Lewin, The nonlinear Schrödinger equation for orthonormal functions: II. Application to Lieb–Thirring inequalities. *Comm. Math. Phys.* **384** (2021), 1783–1828.

[21] R. L. Frank, D. Gontier, and M. Lewin, The periodic Lieb–Thirring inequality. In *Partial differential equations, spectral theory, and mathematical physics. The Ari Laptev anniversary volume*, edited by P. Exner et al., pp. 135–154, EMS Ser. Congr. Rep. 18, EMS Publishing House, 2021.

[22] R. L. Frank, D. Gontier, and M. Lewin, Optimizers for the finite rank Lieb–Thirring inequality. 2021, arXiv:2109.05984.

[23] R. L. Frank, D. Hundertmark, M. Jex, and P. T. Nam, The Lieb–Thirring inequality revisited. *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 8, 2583–2600.

[24] R. L. Frank and A. Laptev, Bound on the number of negative eigenvalues of two-dimensional Schrödinger operators on domains. *Algebra i Analiz* **30** (2018), no. 3, 250–272. Reprinted in *St. Petersburg Math. J.* **30** (2019), no. 3, 573–589.

[25] R. L. Frank, A. Laptev, E. H. Lieb, and R. Seiringer, Lieb–Thirring inequalities for Schrödinger operators with complex-valued potentials. *Lett. Math. Phys.* **77** (2006), no. 3, 309–316.

[26] R. L. Frank, A. Laptev, and O. Safronov, On the number of eigenvalues of Schrödinger operators with complex potentials. *J. Lond. Math. Soc. (2)* **94** (2016), no. 2, 377–390.

[27] R. L. Frank, A. Laptev, and T. Weidl, *Schrödinger operators: eigenvalues and Lieb–Thirring inequalities*. Cambridge University Press, Cambridge, to appear.

[28] R. L. Frank, M. Lewin, E. H. Lieb, and R. Seiringer, Strichartz inequality for orthonormal functions. *J. Eur. Math. Soc. (JEMS)* **16** (2014), no. 7, 1507–1526.

[29] R. L. Frank, E. H. Lieb, and J. Sabin, Maximizers for the Stein–Tomas inequality. *Geom. Funct. Anal.* **26** (2016), no. 4, 1095–1134.

[30] R. L. Frank, E. H. Lieb, and R. Seiringer, Equivalence of Sobolev inequalities and Lieb–Thirring inequalities. In *XVIth International Congress on Mathematical Physics*, pp. 523–535, World Sci. Publ., Hackensack, NJ, 2010.

[31] R. L. Frank and J. Sabin, Restriction theorems for orthonormal functions, Strichartz inequalities, and uniform Sobolev estimates. *Amer. J. Math.* **139** (2017), no. 6, 1649–1691.

[32] R. L. Frank and J. Sabin, Spectral cluster bounds for orthonormal systems and oscillatory integral operators in Schatten spaces. *Adv. Math.* **317** (2017), 157–192.

[33] R. L. Frank and J. Sabin, The Stein–Tomas inequality in trace ideals. In *Séminaire Laurent Schwartz–Équations aux dérivées partielles et applications*. Année 2015–2016, Exp. No. XV, 12 pp., Ed. Éc. Polytech., Palaiseau, 2017.

[34] R. L. Frank and L. Schimmer, Endpoint resolvent estimates for compact Riemannian manifolds. *J. Funct. Anal.* **272** (2017), no. 9, 3904–3918.

[35] R. L. Frank and B. Simon, Eigenvalue bounds for Schrödinger operators with complex potentials. II. *J. Spectr. Theory* **7** (2017), no. 3, 633–658.

[36] I. C. Gohberg and M. G. Kreĭn, *Introduction to the theory of linear nonselfadjoint operators*. Transl. Math. Monogr. 18, American Mathematical Society, Providence, RI, 1969.

[37] M. Hansmann, An eigenvalue estimate and its application to non-selfadjoint Jacobi and Schrödinger operators. *Lett. Math. Phys.* **98** (2011), no. 1, 79–95.

[38] D. Hundertmark, Some bound state problems in quantum mechanics. In *Spectral theory and mathematical physics: a Festschrift in honor of Barry Simon's 60th birthday*, pp. 463–496, Proc. Sympos. Pure Math. 76, Part 1, Amer. Math. Soc., Providence, RI, 2007.

[39] M. Keel and T. Tao, Endpoint Strichartz estimates. *Amer. J. Math.* **120** (1998), no. 5, 955–980.

[40] C. E. Kenig, A. Ruiz, and C. D. Sogge, Uniform Sobolev inequalities and unique continuation for second order constant coefficient differential operators. *Duke Math. J.* **55** (1987), no. 2, 329–347.

[41] Y. Kwon and S. Lee, Sharp resolvent estimates outside of the uniform boundedness range. *Comm. Math. Phys.* **374** (2020), no. 3, 1417–1467.

[42] A. Laptev, Spectral inequalities for partial differential equations and their applications. In *Fifth international congress of Chinese mathematicians*, pp. 629–643, AMS/IP Stud. Adv. Math. 51, pt. 1, Amer. Math. Soc., Providence, RI, 2012.

[43] A. Laptev and T. Weidl, Recent results on Lieb–Thirring inequalities. *Journ. Equ. Dériv. Partielles* (2000), Exp. No. 20. Univ. Nantes, Nantes, 2000.

[44] A. Laptev and T. Weidl, Sharp Lieb–Thirring inequalities in high dimensions. *Acta Math.* **184** (2000), 87–111.

[45] D. Levin and M. Solomyak, The Rozenblum–Lieb–Cwikel inequality for Markov generators. *J. Anal. Math.* **71** (1997), 173–193.

[46] M. Lewin and J. Sabin, The Hartree equation for infinitely many particles, II: Dispersion and scattering in 2D. *Anal. PDE* **7** (2014), no. 6, 1339–1363.

[47] M. Lewin and J. Sabin, The Hartree equation for infinitely many particles I. Well-posedness theory. *Comm. Math. Phys.* **334** (2015), no. 1, 117–170.

[48] E. H. Lieb, An L^p bound for the Riesz and Bessel potentials of orthonormal functions. *J. Funct. Anal.* **51** (1983), no. 2, 159–165.

[49] E. H. Lieb, Kinetic energy bounds and their application to the stability of matter. In *Schrödinger operators (Sønderborg, 1988)*, pp. 371–382, Lecture Notes in Phys. 345, Springer, Berlin, 1989.

[50] E. H. Lieb, M. Lewin, and R. Seiringer, Universal functionals in density functional theory. 2019, arXiv:1912.10424.

[51] E. H. Lieb and M. Loss, *Analysis. 2nd edn.*, Grad. Stud. Math. 14, American Mathematical Society, Providence, RI, 2001.

[52] E. H. Lieb and R. Seiringer, *The stability of matter in quantum mechanics*. Cambridge University Press, Cambridge, 2010.

[53] E. H. Lieb and W. E. Thirring, Bound on kinetic energy of fermions which proves stability of matter. *Phys. Rev. Lett.* **35** (1975), 687–689.

[54] E. H. Lieb and W. E. Thirring, Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities. In *Studies in mathematical physics*, pp. 269–303, Princeton University Press, 1976.

[55] P.-L. Lions and T. Paul, Sur les mesures de Wigner. *Rev. Mat. Iberoam.* **9** (1993), no. 3, 553–618.

[56] D. Lundholm and J. P. Solovej, Hardy and Lieb–Thirring inequalities for anyons. *Comm. Math. Phys.* **322** (2013), no. 3, 883–908.

[57] P. T. Nam, Lieb–Thirring inequality with semiclassical constant and gradient error term. *J. Funct. Anal.* **274** (2018), no. 6, 1739–1746.

[58] M. Rumin, Spectral density and Sobolev inequalities for pure and mixed states. *Geom. Funct. Anal.* **20** (2010), no. 3, 817–844.

[59] M. Rumin, Balanced distribution-energy inequalities and related entropy bounds. *Duke Math. J.* **160** (2011), no. 3, 567–597.

[60] J. Sabin, The Hartree equation for infinite quantum systems. *Journ. Equ. Dériv. Partielles* (2014), Exp. No. 8.

[61] J. Sabin, Littlewood–Paley decomposition of operator densities and application to a new proof of the Lieb–Thirring inequality. *Math. Phys. Anal. Geom.* **19** (2016), no. 2, 11, 11 pp.

[62] B. Simon, *Trace ideals and their applications*. 2nd edn., Math. Surveys Monogr. 120, American Mathematical Society, Providence, RI, 2005.

[63] C. D. Sogge, Concerning the L^p norm of spectral clusters for second-order elliptic operators on compact manifolds. *J. Funct. Anal.* **77** (1988), no. 1, 123–138.

[64] M. Solomyak, Spectral problems related to the critical exponent in the Sobolev embedding theorem. *Proc. Lond. Math. Soc. (3)* **71** (1995), no. 1, 53–75.

[65] E. M. Stein, Oscillatory integrals in Fourier analysis. In *Beijing lectures in harmonic analysis (Beijing, 1984)*, pp. 307–355, Ann. of Math. Stud. 112, Princeton Univ. Press, Princeton, NJ, 1986.

[66] R. S. Strichartz, Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations. *Duke Math. J.* **44** (1977), no. 3, 705–714.

[67] P. A. Tomas, A restriction theorem for the Fourier transform. *Bull. Amer. Math. Soc.* **81** (1975), 477–478.

RUPERT L. FRANK

Mathematisches Institut, Ludwig-Maximilans Universität München, Theresienstr. 39, 80333 München, Germany, and Munich Center for Quantum Science and Technology, Schellingstr. 4, 80799 München, Germany, and Mathematics 253-37, Caltech, Pasadena, CA 91125, USA, r.frank@lmu.de

ON THE NONLINEAR STABILITY OF SHEAR FLOWS AND VORTICES

ALEXANDRU D. IONESCU AND HAO JIA

ABSTRACT

In this article we present some of the main ideas in our recent work on the asymptotic stability of shear flows and vortices among solutions of the Euler equations in two dimensions. More precisely, we discuss the following results:

(1) a theorem on the nonlinear asymptotic stability of a large class of shear flows $(b(y), 0)$ in the finite channel $\mathbb{T} \times [0, 1]$, defined by strictly increasing Gevrey smooth functions b, which are linear outside a compact subset of the interval $(0, 1)$ and satisfy suitable spectral conditions;

(2) a theorem on the nonlinear asymptotic stability of point vortex solutions of the Euler equation in $\mathbb{R}^2$;

(3) heuristic analysis showing that the mechanism of inviscid damping is unlikely to work to produce global solutions of the α-generalized SQG equation in two dimensions, for any parameter $\alpha > 0$.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 35Q31; Secondary 35B40, 35B20

KEYWORDS

Incompressible flows, asymptotic stability, nonlinear inviscid damping, Gevrey spaces

INTERNATIONAL CONGRESS OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3776–3799
DOI 10.4171/ICM2022/1

1. INTRODUCTION

In this paper we present some of our recent results on the asymptotic stability of solutions of the two-dimensional incompressible Euler equation. More precisely, we consider solutions $u : [0, \infty) \times \mathcal{D} \to \mathbb{R}^2$ of the equation

$$\partial_t u + u \cdot \nabla u + \nabla p = 0, \quad \operatorname{div} u = 0, \tag{1.1}$$

where the domain $\mathcal{D}$ is either the entire plane $\mathcal{D} = \mathbb{R}^2$ or the finite channel $\mathcal{D} = \mathbb{T} \times [0, 1]$. Letting $\omega := -\partial_y u^x + \partial_x u^y$ denote the vorticity field, equation (1.1) can be written as

$$\partial_t \omega + u \cdot \nabla \omega = 0, \quad u = \nabla^{\perp} \psi = (-\partial_y \psi, \partial_x \psi), \quad \Delta \psi = \omega. \tag{1.2}$$

In the case of the finite channel $\mathcal{D} = \mathbb{T} \times [0, 1]$, we impose also the boundary conditions

$$\psi(x, 0) \equiv 0, \quad \psi(x, 1) \equiv C_0, \tag{1.3}$$

where C_0 is a constant preserved by the flow.

The two-dimensional incompressible Euler equation is globally well posed for smooth initial data, by the classical result of Wolibner [48]. The long-time behavior of general solutions is, however, very difficult to understand, due to the lack of a global relaxation mechanism. A more realistic goal is to study the global nonlinear dynamics of solutions that are close to steady states of the 2D Euler equation. Coherent structures, such as shear flows and vortices, are particularly important in the study of the 2D Euler equation, since numerical simulations and physical experiments, such as those of [2,3,9,21,34,35,40,41], show that they tend to form dynamically and become the dominant feature of solutions.

The main topic in this article is the study of asymptotic stability of shear flows and vortices. This is a classical subject and a fundamental problem in hydrodynamics. Early investigations were started by Rayleigh [38], Kelvin [27], Orr [37], Taylor [44], among many others, with a focus on mode stability. More detailed understanding of general spectral properties and suitable linear decay estimates were obtained later, see, for example, [8,10,17,42]. In the direction of nonlinear results, Arnold [1] proved a general stability theorem, using the energy method, but this does not give asymptotic information on the global dynamics.

The full nonlinear asymptotic stability problem has only been investigated in recent years, starting with the work of Bedrossian–Masmoudi [7], who proved nonlinear stability in the simplest case of perturbations of the Couette flow, i.e., showing that small perturbations of the Couette flow on the infinite cylinder $\mathbb{T} \times \mathbb{R}$ converge weakly to nearby shear flows. This result was extended by the authors [23] to the finite channel $\mathbb{T} \times [0, 1]$, in order to be able to consider solutions with finite energy. In [24] the authors also proved asymptotic stability of point vortex solutions in $\mathbb{R}^2$, showing that small perturbations converge to a radial profile, and the position of the point vortex stabilizes rapidly at the center of the final radial profile. Finally, in [25] the authors proved nonlinear asymptotic stability of a large family of monotonic shear flows (a similar theorem was proved slightly later and independently by Masmoudi–Zhao [33]). In this article we discuss the main ideas of our papers [23–25].

The linearized equations around other stationary solutions were also investigated intensely in the last few years, and linear inviscid damping and decay was proved in many cases of physical interest, see, for example, [4, 15, 20, 45–47, 49, 50]. However, it also became clear that there are major conceptual difficulties in passing from linear to nonlinear stability, such as the presence of “resonant times” in the nonlinear problem, which require refined Fourier analysis techniques, and the fact that the final state of the flow is determined dynamically by the global evolution and cannot be described in terms of the initial data.

Nonlinear inviscid damping is a very subtle mechanism of stability that has only been established in 2 dimensions and for Euler-type equations. In fact, the heuristic analysis we present in Section 3 of this article suggests that this mechanism fails to produce global solutions of the α-generalized SQG equation in 2 dimensions, for any parameter $\alpha > 0$.

The Euler equations can also be viewed as the limiting case of the Navier–Stokes equations with small viscosity $\nu > 0$. In the presence of viscosity, one can have much more robust stability results, both in 2 and 3 dimensions, for initial data that is sufficiently small relative to ν. See the recent papers [5, 6, 12, 18] and references therein.

We note also that the problem of nonlinear inviscid damping is connected to the well-known Landau damping effect for the Vlasov–Poisson equations. We refer to the celebrated work of Mouhot–Villani [36] for the physical background and more references.

1.1. Monotonic shear flows

We consider a perturbative regime for the Euler equation (1.1), with velocity field given by $(b(y), 0)) + u(x, y)$ and vorticity given by $-b'(y) + \omega$. We define the Gevrey spaces $\mathcal{G}^{\lambda,s}(\mathbb{T}\times\mathbb{R})$ as the spaces of L^2 functions f on $\mathbb{T}\times\mathbb{R}$ defined by the norm

$$\|f\|_{\mathcal{G}^{\lambda,s}(\mathbb{T}\times\mathbb{R})} := \left\|e^{\lambda\langle k,\xi\rangle^s}\tilde{f}(k,\xi)\right\|_{L^2_{k,\xi}} < \infty, \quad s \in (0,1], \lambda > 0. \tag{1.4}$$

In the above $(k, \xi) \in \mathbb{Z}\times\mathbb{R}$ and $\tilde{f}$ denotes the Fourier transform of f in (x, y). More generally, for any interval $I \subseteq \mathbb{R}$ we define the Gevrey spaces $\mathcal{G}^{\lambda,s}(\mathbb{T}\times I)$ by

$$\|f\|_{\mathcal{G}^{\lambda,s}(\mathbb{T}\times I)} := \|Ef\|_{\mathcal{G}^{\lambda,s}(\mathbb{T}\times\mathbb{R})}, \tag{1.5}$$

where $Ef(x) := f(x)$ if $x \in I$ and $Ef(x) := 0$ if $x \notin I$. The use of Gevrey spaces is necessary in the context of inviscid damping, mainly due to loss of regularity during the flow.

We will assume that the background shear flow $b \in C^\infty(\mathbb{R})$ satisfies the following:

(A) For some $\vartheta_0 \in (0, 1/10]$ and $\beta_0 > 0$

$$\vartheta_0 \le b'(y) \le 1/\vartheta_0 \quad \text{for } y \in [0,1] \quad \text{and} \quad b''(y) \equiv 0 \quad \text{for } y \notin [2\vartheta_0, 1-2\vartheta_0], \tag{1.6}$$

and

$$\|b\|_{L^\infty(0,1)} + \left\|b''\right\|_{\mathcal{G}^{\beta_0,1/2}} \le 1/\vartheta_0. \tag{1.7}$$

(B) The associated linear operator $L_k : L^2(0,1) \to L^2(0,1)$, $k \in \mathbb{Z}\backslash\{0\}$, given by

$$L_k f = b(y)f - b''(y)\varphi_k, \quad \text{where } \partial_y^2\varphi_k - k^2\varphi_k = f,\ \varphi_k(0) = \varphi_k(1) = 0, \tag{1.8}$$

has no discrete eigenvalues and, therefore, by the general theory of Fredholm operators, the spectrum of L_k is purely continuous spectrum $[b(0), b(1)]$ for all $k \in \mathbb{Z}\backslash\{0\}$.

For any function $H = H(x, y) : \mathbb{T} \times \mathbb{R} \to \mathbb{C}$, let $\langle H\rangle = \langle H\rangle(y)$ denote the average of H in x. Our main result in [25] is the following theorem:

Theorem 1.1. *Assume that $\beta_0, \vartheta_0 > 0$ and b satisfies the assumptions above. Then there are constants $\beta_1 > 0$ and $\overline{\varepsilon} > 0$ such that the following statement is true:*

Assume that ω_0 has compact support in $\mathbb{T} \times [2\vartheta_0, 1 - 2\vartheta_0]$, and satisfies

$$\|\omega_0\|_{\mathcal{G}^{\beta_0,1/2}(\mathbb{T}\times\mathbb{R})} = \varepsilon \le \overline{\varepsilon}, \qquad \int_{\mathbb{T}} \omega_0(x, y)\,dx = 0 \quad \text{for any } y \in [0, 1]. \tag{1.9}$$

Let $\omega : [0, \infty) \times \mathbb{T} \times [0, 1] \to \mathbb{R}$ denote the global smooth solution to the Euler equation

$$\begin{cases} \partial_t\omega + b(y)\partial_x\omega - b''(y)\partial_x\psi + u \cdot \nabla\omega = 0, \\ u = (u^x, u^y) = (-\partial_y\psi, \partial_x\psi), \quad \Delta\psi = \omega, \quad \psi(t, x, 0) = \psi(t, x, 1) = 0, \end{cases} \tag{1.10}$$

with initial data ω_0. Then we have the following conclusions:

(i) *For all $t \ge 0$, $\operatorname{supp}\omega(t) \subseteq \mathbb{T} \times [\vartheta_0, 1 - \vartheta_0]$.*

(ii) *There exists $F_\infty(x, y) \in \mathcal{G}^{\beta_1,1/2}$ with $\operatorname{supp} F_\infty \subseteq \mathbb{T} \times [\vartheta_0, 1 - \vartheta_0]$ such that*

$$\big\|\omega\big(t, x + tb(y) + \Phi(t, y), y\big) - F_\infty(x, y)\big\|_{\mathcal{G}^{\beta_1,1/2}(\mathbb{T}\times[0,1])} \lesssim_{\beta_0,\vartheta_0,\kappa} \frac{\varepsilon}{\langle t\rangle} \tag{1.11}$$

for all $t \ge 0$, where

$$\Phi(t, y) := \int_0^t \big\langle u^x\big\rangle(\tau, y)\,d\tau. \tag{1.12}$$

(iii) *We define the smooth functions $\psi_\infty, u_\infty : [0, 1] \to \mathbb{R}$ by*

$$\partial_y^2\psi_\infty = \langle F_\infty\rangle, \quad \psi_\infty(0) = \psi_\infty(1) = 1, \quad u_\infty(y) := -\partial_y\psi_\infty(y). \tag{1.13}$$

Then the velocity field $u = (u^x, u^y)$ satisfies the bounds

$$\big\|\big\langle u^x\big\rangle(t, y) - u_\infty(y)\big\|_{\mathcal{G}^{\beta_1,1/2}(\mathbb{T}\times[0,1])} \lesssim \frac{\varepsilon}{\langle t\rangle^2} \tag{1.14}$$

and

$$\langle t\rangle\big\|u^x(t, x, y) - \big\langle u^x\big\rangle(t, y)\big\|_{L^\infty(\mathbb{T}\times[0,1])} + \langle t\rangle^2\big\|u^y(t, x, y)\big\|_{L^\infty(\mathbb{T}\times[0,1])} \lesssim \varepsilon. \tag{1.15}$$

1.1.1. Remarks

The simplest case $b(y) = y$ (the Couette flow) was treated earlier in [7, 23]. We discuss now some of the assumptions and conclusions of our main theorem.

(1) Equation (1.10) for the vorticity deviation is equivalent to the original Euler equations (1.1)–(1.3). The condition $\int_{\mathbb{T}} \omega_0(x,y)\,dx = 0$ can be imposed without loss of generality, by replacing the shear flow $b(y)$ with the nearby shear flow $b(y) + \langle u_0^x \rangle(y)$.

(2) The assumption on the compact support of ω_0 is likely necessary to prove scattering in Gevrey spaces. Indeed, Zillinger [49] showed that scattering does not hold in high Sobolev spaces unless one assumes that the vorticity vanishes at high order at the boundary. This is due to what is called "boundary effect," which is not consistent with inviscid damping. Investigating the boundary effect in the context of asymptotic stability of Euler or Navier–Stokes equations is an interesting topic by itself, but we will not address it here.

(3) The assumption on the support of b'' is necessary to preserve the compact support of $\omega(t)$ in $\mathbb{T} \times [\vartheta_0, 1-\vartheta_0]$, due to the nonlocal term $b''(y)\partial_x \psi$ in (1.10). Assumption (1.6) on the uniform monotonicity of the function b is also important for our proof, to ensure a uniform rate of inviscid damping. It is an open question to investigate what happens in the case of nonmonotone shear flows which are linearly stable, such as Kolmogorov or Poiseuille flows.

(4) There is a large class of shear flows b satisfying our assumptions, for instance, functions $b(y)$ satisfying $b'(y) \geq 1$ and $|b'''(y)| < 1$, $y \in [0,1]$.

(5) The Gevrey regularity assumption (1.9) on the initial data ω_0 is likely sharp. See the recent construction of nonlinear instability of Deng–Masmoudi [16] for the Couette flow in slightly larger Gevrey spaces, and the more definitive counterexamples to inviscid damping in low Sobolev spaces by Lin–Zeng [30].

(6) The most important statement in Theorem 1.1 is (1.11), which provides strong control on the "profile" of the vorticity and from which the other statements follow easily. We note that the convergence (1.11) of the profile for vorticity holds in a slightly weaker Gevrey space, since $\beta_1 < \beta_0$. This is connected with the use of energy functionals with decreasing time-dependent weights to control the profile, and is a reflection of the phenomenon that "decay costs regularity" in inviscid damping.

(7) At the qualitative level, our main conclusion (1.11) shows that the vorticity ω converges weakly to the function $\langle F_\infty \rangle(y)$. This is consistent with a far-reaching conjecture regarding the long-time behavior of solutions of the $2D$ Euler equation, see [43], which predicts that for general generic solutions the vorticity field converges, as $t \to \infty$, weakly but not strongly in L^2_{loc} to a steady state. Proving such a conjecture for general solutions is, of course, well beyond the current PDE techniques, but the nonlinear asymptotic stability results we have so far in [7,23–25] are consistent with this conjecture.

(8) One can gain some intuition and explain the more technical conclusions in Theorem 1.1 by examining a simple explicit case, corresponding to the Couette flow $b(y) = y$. In this case $b''(y) = 0$ and the linearization of the main equation (1.10) is

$$\partial_t \omega + y \partial_x \omega = 0, \tag{1.16}$$

which was studied by Orr in a pioneering work [37]. To simplify the discussion, assume that $x \in \mathbb{T}$, $y \in \mathbb{R}$ (to avoid the boundary issue which is not our main concern here). By direct calculation, we have $\omega(t,x,y) = \omega_0(x - yt, y)$. The stream function is given by

$\Delta\psi(t,x,y) = \omega(t,x,y)$ for $(x,y) \in \mathbb{T} \times \mathbb{R}$, so in the Fourier space we have the formulas

$$\widetilde{\omega}(t,k,\xi) = \widetilde{\omega_0}(k,\xi+kt), \quad \widetilde{\psi}(t,k,\xi) = -\frac{\widetilde{\omega_0}(k,\xi+kt)}{k^2+|\xi|^2}. \tag{1.17}$$

We remark that the conclusions in the full nonlinear Theorem 1.1 are consistent with these explicit formulas. Indeed, if ω_0 is smooth, so $\widetilde{\omega_0}(k,\xi)$ decays fast in (k,ξ), then:

(i) The main contribution comes from the frequencies $\xi = -kt + O(1)$, therefore $\widetilde{\psi}(t,k,\xi)$ decays like $|k|^{-2}\langle t\rangle^{-2}$ if $k \neq 0$. Similarly, since $u^x = -\partial_y\psi$ and $u^y = \partial_x\psi$, we see that $\widetilde{u^x}$ decays like $|k|^{-1}\langle t\rangle^{-1}$ and $\widetilde{u^y}$ decays like $|k|^{-1}\langle t\rangle^{-2}$, as claimed in (1.15).

(ii) It can be seen from (1.17) that the functions $\omega(t,x,y)$ and $\psi(t,x,y)$ are not uniformly smooth as $t \to \infty$, in the coordinates x, y. To identify smooth "profiles," we need to make changes of coordinates, i.e., we define

$$z = x - tv, \quad v = y, \quad F(t,z,v) = \omega(t,x,y), \quad \phi(t,z,v) = \psi(t,x,y). \tag{1.18}$$

Notice that $F(t,z,v) = \omega_0(z,v)$ (independent of t), while $\phi(t,z,v)$ is uniformly smooth for all t provided that ω_0 is smooth. Taking the Fourier transform in z, v, we have the formula

$$\widetilde{\phi}(t,k,\xi) = -\frac{\widetilde{\omega_0}(k,\xi)}{k^2+|\xi-kt|^2}. \tag{1.19}$$

An important observation of Orr is that for $k \neq 0$ and large ξ, the normalized stream function ϕ (as well as the velocity field) experiences a *transient growth* as t approaches the "critical time" $t_c = \xi/k$ before decaying to zero. This can be seen easily from the formula (1.19). This transient growth on the linearized level turns out to be crucial for the nonlinear analysis as well, and leads to the high regularity assumptions (Gevrey spaces) that are required for the nonlinear perturbation theory.

1.2. Point vortices

Vortices (radial functions) are stationary solutions of the Euler equation in $\mathbb{R}^2$ in vorticity formulation (1.2). The stability of vortices is a major open problem for 2D Euler equations, which is challenging even at the linear level as shown in [4] in the case of radially decreasing vortices.

In [24] we initiated the rigorous study of the full nonlinear asymptotic stability problem for vortices of the Euler equation in $\mathbb{R}^2$. We consider the simplest class of vortices, called *point vortices*, which are δ-functions centered at points in $\mathbb{R}^2$. Such solutions (and more generally the so called N-vortex solutions) are models of general solutions with vorticity concentrated sharply in small neighborhoods, and have been studied by many authors. See, for instance, the classical work of Kirchhoff [28], C. C. Lin [29], and the book of Majda–Bertozi [31] for more references.

To state our main conclusions, consider solutions of the form

$$\text{vorticity field} = \kappa\,\delta\big(P(t)\big) + \omega, \quad \text{velocity field} = \nabla^\perp\Delta^{-1}\delta\big(P(t)\big) + u, \tag{1.20}$$

where $\kappa \in \mathbb{R}\backslash\{0\}$ is the strength of the point vortex, $\delta(P(t))$ is the Dirac mass centered at $P(t) = (P_1(t), P_2(t)) \in \mathbb{R}^2$. We assume that $P(t)$ is not in the support of ω, which will be satisfied as part of our analysis. Then the perturbation ω satisfies the equation

$$\partial_t\omega + U\cdot\nabla\omega + u\cdot\nabla\omega = 0, \quad \text{for } (x,y,t) \in \mathbb{R}^2\times[0,\infty), \tag{1.21}$$

where

$$U = \nabla^\perp\Delta^{-1}\delta\big(P(t)\big) = \frac{\kappa}{2\pi}\nabla^\perp\log\big|(x,y) - P(t)\big|. \tag{1.22}$$

The velocity field u and the stream function ψ are determined through

$$\begin{aligned} &u = \nabla^\perp\psi = (-\partial_y\psi, \partial_x\psi),\\ &\Delta\psi = \omega, \qquad \lim_{|(x,y)|\to\infty}\Big\{\psi(x,y) - \frac{c_0}{2\pi}\log\big|(x,y)\big|\Big\} = 0, \end{aligned} \tag{1.23}$$

where

$$c_0 := \int_{\mathbb{R}^2}\omega(t,x,y)dxdy \tag{1.24}$$

is a constant preserved by the flow for all times (as long as the support of $\omega(t)$ is away from $P(t)$). In addition, the center $P(t)$ satisfies the transport ODE

$$P'(t) = \nabla^\perp\psi\big(t, P(t)\big). \tag{1.25}$$

Equations (1.21)–(1.25) can be derived rigorously when the vortex $P(t)$ lies outside of the support of $\omega(t)$, see, for example, [32]. In our case, this support condition is propagated dynamically by the flow, as a consequence of the proof of stability.

In [24] we prove axisymmetrization around a point vortex. More precisely, we prove that small, Gevrey smooth, and compactly supported perturbations symmetrize around the point vortex whose location changes in time and converges fast as $t\to\infty$.

Theorem 1.2. *Assume that* $\kappa \in \mathbb{R}\backslash\{0\}$, $\lambda \in (0,\infty)$, $M \in (1,\infty)$, *and* $\omega_0 \in C_0^\infty(\mathbb{R}^2)$ *satisfies the support property* $\operatorname{supp}\omega_0 \subseteq \{x \in \mathbb{R}^2 : |x| \in [1/M, M]\}$. *Assume that*

$$\int_{\mathbb{R}^2} e^{\lambda\langle\xi,\eta\rangle^{1/2}}\big|\widetilde{\omega_0}(\xi,\eta)\big|^2\,d\xi d\eta \le \varepsilon^2, \tag{1.26}$$

for a sufficiently small constant $\varepsilon \le \varepsilon(\kappa, M, \lambda)$, *where* $\widetilde{\omega_0}$ *denotes the Fourier transform of* ω_0. *Then there is a unique smooth global solution* (ω, P) *of the system* (1.21)–(1.25) *such that* $P(t)$ *stays outside of the support of* $\omega(t)$ *for all* $t \ge 0$. *Moreover,*

$$\big|P(t) - P_\infty\big| \lesssim \varepsilon\, e^{-c\langle t\rangle^{1/2}} \quad \textit{for all } t \ge 0, \tag{1.27}$$

for some $P_\infty \in \mathbb{R}^2$ *and* $c = c(\kappa, M, \lambda) > 0$, *and the vorticity* $\omega(t)$ *converges weakly to a Gevrey-2 regular function* $\omega_\infty \in C^\infty(\mathbb{R}^2)$ *which is radial with respect to* P_∞, *as* $t\to\infty$.

1.2.1. Adapted polar coordinates and precise results

To understand the mechanism of convergence in Theorem 1.2, we need to analyze the Euler equations in the polar coordinates, recentered around the moving point vortex $P(t)$. Let

$$(x, y) = P(t) + r(\cos\theta, \sin\theta). \tag{1.28}$$

In (r, θ) coordinates, we set the functions u'_r, u'_θ, ψ', ω' as follows:

$$\begin{aligned} &\omega'(t,\theta,r) = \omega(t,x,y), \quad \psi'(t,\theta,r) = \psi(t,x,y), \\ &u'_r(t,\theta,r)e_r + u'_\theta(t,\theta,r)e_\theta = u(t,x,y), \end{aligned} \tag{1.29}$$

where $e_r := (\cos\theta, \sin\theta)$, $e_\theta := (-\sin\theta, \cos\theta)$. Equation (1.21) can be rewritten as

$$\partial_t\omega' - \big(P'(t), e_r\big)\partial_r\omega' - \frac{1}{r}\big(P'(t), e_\theta\big)\partial_\theta\omega' + \frac{\kappa}{2\pi r^2}\partial_\theta\omega' - \frac{\partial_\theta\psi'\partial_r\omega' - \partial_r\psi'\partial_\theta\omega'}{r} = 0, \tag{1.30}$$

where the stream function $\psi'(t,\theta,r)$ can be calculated through

$$\partial_r^2\psi' + \frac{1}{r}\partial_r\psi' + \frac{1}{r^2}\partial_\theta^2\psi' = \omega', \quad \lim_{r\to\infty}\Big\{\psi'(t,r,\theta) - \frac{c_0}{2\pi}\log r\Big\} = 0. \tag{1.31}$$

In the above,

$$P'(t) = \frac{1}{2\pi}\int_0^\infty\int_0^{2\pi}(\sin\theta, -\cos\theta)\,\omega'(t,\theta,r)d\theta dr, \tag{1.32}$$

and $(P'(t), e_r)$, $(P'(t), e_\theta)$ denote the scalar products between the vectors $P'(t)$, e_r, and e_θ. The velocity field (u'_θ, u'_r) can be calculated according to the formulas

$$u'_\theta(t,\theta,r) = \partial_r\psi', \quad u'_r(t,\theta,r) = -(1/r)\partial_\theta\psi'. \tag{1.33}$$

The following theorem is the full quantitative version of our main result in [24]:

Theorem 1.3. *Assume that $\beta_0, \vartheta_0 \in (0, 1/8]$, $\kappa \in (0,\infty)$, and assume ω'_0 is smooth initial data, satisfying the support condition* $\operatorname{supp}\omega'_0 \subseteq \mathbb{T}\times[\vartheta_0, 1/\vartheta_0]$ *and the smallness condition*

$$\|\omega'_0\|_{\mathcal{G}^{\beta_0,1/2}(\mathbb{T}\times\mathbb{R})} = \varepsilon \le \overline{\varepsilon}, \tag{1.34}$$

where $\overline{\varepsilon} = \overline{\varepsilon}(\beta_0, \vartheta_0, \kappa) > 0$ is sufficiently small and the Gevrey spaces $\mathcal{G}^{\beta_0,1/2}(\mathbb{T}\times\mathbb{R})$ are defined as in (1.4). We have the following conclusions:

(i) (global regularity) *There exist $\beta_1 = \beta_1(\beta_0, \vartheta_0, \kappa) > 0$ and a unique global solution $\omega' \in C([0,\infty) : \mathcal{G}^{\beta_1,1/2}(\mathbb{T}\times\mathbb{R}))$ of the system (1.30)–(1.32) with initial data $\omega'(0) = \omega'_0$ such that* $\operatorname{supp}\omega'(t) \subseteq \mathbb{T}\times[\vartheta_0/2, 2/\vartheta_0]$ *and $|P(t)| < \vartheta_0/100$ for any $t \in [0,\infty)$.*

(ii) (asymptotic stability) *There exist $\Omega_\infty \in \mathcal{G}^{\beta_1,1/2}(\mathbb{T}\times\mathbb{R})$ and $P_\infty = (P^1_\infty, P^2_\infty) \in \mathbb{R}^2$ with* $\operatorname{supp}\Omega_\infty \subseteq \mathbb{T}\times[\vartheta_0/2, 2/\vartheta_0]$ *and $|P_\infty| \le \vartheta_0/100$ such that*

$$\big\|\omega'\big(t, \theta + \kappa t/(2\pi r^2) + \Phi(t,r), r\big) - \Omega_\infty(\theta, r)\big\|_{\mathcal{G}^{\beta_1,1/2}(\mathbb{T}\times\mathbb{R})} \lesssim \varepsilon\langle t\rangle^{-1}, \tag{1.35}$$

$$\big|P(t) - P_\infty\big| \lesssim \varepsilon\, e^{-\beta_1 t^{1/2}}, \tag{1.36}$$

for any $t \geq 0$. Here

$$\Phi(t,r) := \int_0^t \frac{\langle u'_\theta \rangle(\tau,r)}{r}\,d\tau = \int_0^t \frac{\langle \partial_r \psi' \rangle(\tau,r)}{r}\,d\tau. \tag{1.37}$$

(iii) (control of the velocity field) *The velocity field u' satisfies the asymptotic bounds*

$$\big\| \langle u'_\theta \rangle(t,r) - u'_\infty(r) \big\|_{\mathcal{G}^{\beta_1,1/2}(\mathbb{R})} \lesssim \varepsilon \langle t \rangle^{-2}, \tag{1.38}$$

$$\langle t \rangle \big\| u'_\theta(t,\theta,r) - \langle u'_\theta \rangle(t,r) \big\|_{L^\infty(\mathbb{T}\times\mathbb{R})} + \langle t \rangle^2 \big\| u'_r(t,\theta,r) \big\|_{L^\infty(\mathbb{T}\times\mathbb{R})} \lesssim \varepsilon, \tag{1.39}$$

where the function $u'_\infty \in \mathcal{G}^{\beta_1,1/2}(\mathbb{R})$ is defined by

$$\partial_r\big(r u'_\infty(r)\big) = r\Omega_\infty(r), \quad u'_\infty(r) = \begin{cases} 0 & \text{if } r \leq \vartheta_0/2, \\ c_0/(2\pi) & \text{if } r \geq 2/\vartheta_0. \end{cases}$$

1.2.2. Remarks

(1) We notice the similarities between Theorem 1.1 (in the Couette case $b(y) = y$) and Theorem 1.3. In the point vortex case, the inviscid damping is generated by the term $\frac{\kappa}{2\pi r^2}\partial_\theta \omega'$ in (1.30), $\kappa \neq 0$. Indeed, at the linearized level, equation (1.30) is

$$\partial_t \omega^{\text{lin}} + \frac{\kappa}{2\pi r^2}\partial_\theta \omega^{\text{lin}} = 0, \tag{1.40}$$

with the explicit solution

$$\omega^{\text{lin}}(t,\theta,r) = \omega_0^{\text{lin}}\big(\theta - \kappa t/(2\pi r^2), r\big). \tag{1.41}$$

Using now (1.31), we can express ψ_k^{lin}, $k \in \mathbb{Z}\backslash\{0\}$, as

$$\psi_k^{\text{lin}}(t,r) = \int_{\mathbb{R}} G_k(r,\rho)\omega_{0,k}^{\text{lin}}(\rho) e^{-ik\kappa t/(2\pi\rho^2)}\,d\rho, \tag{1.42}$$

where ψ_k^{lin} and $\omega_{0,k}^{\text{lin}}$ denote the kth Fourier modes of the functions ψ^{lin} and ω_0^{lin} in θ and G_k is the associated Green function for the operator $\partial_r^2 + \partial_r/r - k^2/r^2$. These formulas and integration by parts in ρ lead to pointwise decay in time for the velocity field $u^{\text{lin}} = (u_\theta^{\text{lin}}, u_r^{\text{lin}}) = (\partial_r \psi^{\text{lin}}, -\partial_\theta \psi^{\text{lin}}/r)$, consistent with the bounds (1.39). In other words, the main conclusions of Theorem 1.3 can be verified for the linearized flow as a consequence of the explicit formulas (1.41)–(1.42), as expected.

(2) The main difference between Theorems 1.1 and 1.3 comes from the global shift caused by the movement of the vortex $P(t)$. It is very important to prove that the point vortex stabilizes rapidly, according to (1.36), which gives just the right amount of decay to compensate for the loss of regularity caused by changes of variables and mixing.

(3) Finally, we note that the assumption that the point vortex lies outside the support of the perturbation is necessary for inviscid damping in Gevrey spaces. This is analogous to the “boundary effect” discussed earlier in the context of shear flows.

1.3. Organization

The rest of this paper is organized as follows: in Section 2 we discuss the main ideas in the proofs of Theorems 1.1 and 1.3. In Section 3 we discuss the limitations of the mechanism of inviscid damping, showing that it cannot be used to prove global regularity of solutions of the generalized SQG equations.

2. MAIN IDEAS

In this section we discuss some of the main ideas involved in the proofs of Theorems 1.1 and 1.3. Most of our discussion will be focused on the harder case of general monotonic shear flows, but some of the key ideas apply also in the case of point vortices.

2.1. Renormalization and the new equations

We introduce now a nonlinear change of variables and define the main quantities we need to control uniformly in time. We need to unwind the transportation in x. Assume that $\omega : [0,T] \times \mathbb{T} \times [0,1]$ is a sufficiently smooth solution of the system (1.10),

$$\begin{aligned} &\partial_t \omega + b(y)\partial_x \omega - b''(y)\partial_x \psi + u \cdot \nabla \omega = 0, \\ &(u^x, u^y) = (-\partial_y \psi, \partial_x \psi), \quad \Delta \psi = \omega, \quad \psi(t,x,1) = \psi(t,x,0) = 0, \end{aligned} \tag{2.1}$$

which is supported in $\mathbb{T} \times [\vartheta_0, 1-\vartheta_0]$ at all times $t \in [0,T]$, satisfying $\|\langle \omega \rangle(t)\|_{H^{10}} \ll 1$. We make the nonlinear change of variables

$$v = b(y) + \frac{1}{t}\int_0^t \langle u^x \rangle(\tau, y)\, d\tau, \quad z = x - tv. \tag{2.2}$$

The point of this change of variables is to eliminate two of the nondecaying terms in the evolution equation in (2.1), namely the terms $b(y)\partial_x \omega$ and $\langle u^x \rangle \partial_x \omega$. The change of variable $y \to v$ is crucial for our analysis, and it allows us to link the renormalized stream function ϕ to the profile F using the elliptic equation (2.7). The point is that this equation has constant coefficients at the linear level, so it is compatible with Fourier analysis.

Then we define the functions

$$F(t,z,v) := \omega(t,x,y), \quad \phi(t,z,v) := \psi(t,x,y), \tag{2.3}$$

$$V'(t,v) := \partial_y v(t,y), \quad V''(t,v) := \partial_{yy} v(t,y), \quad \dot{V}(t,v) := \partial_t v(t,y), \tag{2.4}$$

$$B'(t,v) := \partial_y b(y), \quad B''(t,v) := \partial_{yy} b(y). \tag{2.5}$$

The evolution equation in (2.1) becomes

$$\partial_t F - B''\partial_z \phi - V'\partial_v P_{\neq 0}\phi\, \partial_z F + (\dot{V} + V'\partial_z \phi)\, \partial_v F = 0, \tag{2.6}$$

where $P_{\neq 0}$ is projection off the zero mode, $P_{\neq 0} H(t,z,v) = H(t,z,v) - \langle H \rangle(t,v)$. The renormalized vorticity ϕ satisfies the elliptic-type equation

$$\partial_z^2 \phi + (V')^2(\partial_v - t\partial_z)^2 \phi + V''(\partial_v - t\partial_z)\phi = F, \tag{2.7}$$

The functions V', V'', B', B'', $\dot{V}$ also satisfy suitable evolution or elliptic equations in the new variables (t, v), which can be derived from (2.1) and the definitions, such as

$$\partial_t B'(t,v) + \dot{V}\partial_v B'(t,v) = \partial_t B''(t,v) + \dot{V}\partial_v B''(t,v) = 0, \tag{2.8}$$

$$\partial_t(V' - B') + \dot{V}\partial_v(V' - B') = \mathcal{H}/t, \tag{2.9}$$

$$\partial_t \mathcal{H} + \dot{V}\partial_v \mathcal{H} = -\mathcal{H}/t - V'\langle \partial_v P_{\neq 0}\phi\, \partial_z F\rangle + V'\langle \partial_z\phi\, \partial_v F\rangle, \tag{2.10}$$

where

$$\mathcal{H}(t,v) := tV'(t,v)\partial_v \dot{V}(t,v) = B'(t,v) - V'(t,v) - \langle F\rangle(t,v). \tag{2.11}$$

Equations (2.6)–(2.11) are the main equations we analyze in our proof.

2.2. Energy functionals and imbalanced weights

The main idea is to control the regularity of F for all $t \geq 0$, as well as other quantities such as ϕ, V', V'', B', B'', $\dot{V}$, using a bootstrap argument involving energy functionals and space-time norms. These norms depend on families of weights $A_k(t,\xi)$, $A_{NR}(t,\xi)$, $A_R(t,\xi)$, $k \in \mathbb{Z}$, $\xi \in \mathbb{R}$, which have to be designed carefully to control the nonlinearities.

To identify the main issue and motivate the choice of weights, assume first that F and ϕ satisfy the simplified closed system

$$\partial_t F - \partial_v P_{\neq 0}\phi\, \partial_z F = 0, \quad \partial_z^2\phi + (\partial_v - t\partial_z)^2\phi = F, \tag{2.12}$$

for $(z, v, t) \in \mathbb{T} \times \mathbb{R} \times [0, \infty)$. Compared to the original equations (2.6)–(2.7), we assume that $b'' \equiv 0$ (the Couette flow) and keep only one nonlinear term, the “reaction term” $\partial_v P_{\neq 0}\phi \cdot \partial_z F$. We would like to control, uniformly in time, an energy functional of the form

$$\mathcal{E}(t) := \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} A_k^2(t,\xi)\big|\tilde{F}(t,k,\xi)\big|^2\, d\xi, \tag{2.13}$$

where $\tilde{F}$ denotes the spacial Fourier transform of F, for a suitable weight $A_k(t,\xi)$ which decreases in t. Let $\mathbb{Z}^* = \mathbb{Z} \setminus \{0\}$ and notice that

$$\widetilde{\partial_v P_{\neq 0}\phi}(t,k,\xi) = -\frac{i\xi}{k^2}\frac{\tilde{F}(t,k,\xi)}{1 + |t - \xi/k|^2}\mathbf{1}_{\mathbb{Z}^*}(k). \tag{2.14}$$

When $|\xi| \gg k^2$, the factor ξ/k^2 in (2.14) indicates a loss of one full derivative in v in the resonant region $\{(t,k,\xi) : |t - \xi/k| \ll |\xi|/k^2,\ k^2 \ll |\xi|\}$. This is a major obstruction to proving stability, which cannot be removed by standard symmetrization techniques.

The key original idea of Bedrossian–Masmoudi [7] is to use *imbalanced weights* $A_k(t,\xi)$ to absorb this derivative loss, taking advantage of the favorable structure of the nonlinearity that does not allow for contributions to the resonant region to come from bilinear interactions of small frequencies and frequencies in the resonant region (due to the factor $\partial_z F$ in the reaction term). More precisely, the weights have to satisfy the unusual property

$$\frac{A_\ell(t,\eta)}{A_k(t,\xi)} \approx \left|\frac{\eta}{\ell^2}\right|\frac{1}{1 + |t - \eta/\ell|}, \tag{2.15}$$

when $k \neq \ell, \ell \neq 0, \xi = \eta + O(1), k = \ell + O(1)$, and $1 + |t - \eta/\ell| \ll |\eta|/\ell^2$. In addition, these weights have to decrease in time, in the quantitative form,

$$-\frac{\partial_t A_\ell(t,\xi)}{A_\ell(t,\xi)} \gtrsim \frac{1}{\langle t - \xi/k \rangle}, \tag{2.16}$$

if $k \in \mathbb{Z}\backslash\{0\}$, $\langle t - \xi/k \rangle \lesssim |\xi|/k^2$, and $|\ell| \leq \langle \xi \rangle$, in order to be able to control some of the nonlinear terms using the Cauchy–Kowalevski terms coming from time differentiation of the energy functional $\mathcal{E}$. This leads to loss of regularity of the profile F during the evolution, which is the price to pay to prove nonlinear decay of the stream function ϕ.

2.2.1. The weights A_{NR}, A_R, A_k

For the sake of completeness, we summarize here the construction of our main imbalanced weights A_R, A_{NR}, A_k in [23–25]. Given $\delta_0 = \delta_0(\beta_0, \vartheta_0) > 0$, we define first the decreasing function $\lambda : [0,\infty) \to [\delta_0, 3\delta_0/2]$ by

$$\lambda(0) = \frac{3}{2}\delta_0, \quad \lambda'(t) = -\frac{\delta_0 \sigma_0^2}{\langle t \rangle^{1+\sigma_0}}, \tag{2.17}$$

for small positive constant σ_0 (say $\sigma_0 = 0.01$). Then we define

$$A_R(t,\xi) := \frac{e^{\lambda(t)\langle\xi\rangle^{1/2}}}{b_R(t,\xi)} e^{\sqrt{\delta}\langle\xi\rangle^{1/2}}, \quad A_{NR}(t,\xi) := \frac{e^{\lambda(t)\langle\xi\rangle^{1/2}}}{b_{NR}(t,\xi)} e^{\sqrt{\delta}\langle\xi\rangle^{1/2}}, \tag{2.18}$$

$$A_k(t,\xi) := e^{\lambda(t)\langle k,\xi\rangle^{1/2}} \left(\frac{e^{\sqrt{\delta}\langle\xi\rangle^{1/2}}}{b_k(t,\xi)} + e^{\sqrt{\delta}|k|^{1/2}} \right), \tag{2.19}$$

where $\delta > 0$ is a small constant and $k \in \mathbb{Z}$.

To construct the main functions b_k, b_{NR}, b_R that appear in (2.18)–(2.19), we start by defining two functions $w_{NR}, w_R : [0,\infty) \times \mathbb{R} \to [0,1]$, which distinguish between resonant and nonresonant regions and play a key role in the analysis. Resonance is measured in terms of the size of the denominators $\langle t - \xi/k \rangle$, which appear in formula (2.14). The intervals $I_{k,\eta}$ defined below, where this factor is small are called "resonant" intervals. Notice the imbalance in (2.24) between the weights $w_R(t,\eta)$ and $w_{NR}(t,\eta)$, especially around the center of the resonant intervals, consistent with the loss of derivative discussed earlier.

Assume that $\delta > 0$ is small, $\delta \ll \delta_0$. For $|\eta| \leq \delta^{-10}$, we define simply

$$w_{NR}(t,\eta) := 1, \quad w_R(t,\eta) := 1. \tag{2.20}$$

For $\eta > \delta^{-10}$, we define $k_0(\eta) := \lfloor \sqrt{\delta^3 \eta} \rfloor$. For $l \in \{1,\ldots,k_0(\eta)\}$, we define

$$t_{l,\eta} := \frac{1}{2}\left(\frac{\eta}{l+1} + \frac{\eta}{l} \right), \quad t_{0,\eta} := 2\eta, \quad I_{l,\eta} := [t_{l,\eta}, t_{l-1,\eta}]. \tag{2.21}$$

Notice that $|I_{l,\eta}| \approx \eta/l^2$ and

$$\delta^{-3/2}\sqrt{\eta}/2 \leq t_{k_0(\eta),\eta} \leq \cdots \leq t_{l,\eta} \leq \eta/l \leq t_{l-1,\eta} \leq \cdots \leq t_{0,\eta} = 2\eta.$$

We define

$$w_{NR}(t,\eta) := 1, \; w_R(t,\eta) := 1 \quad \text{if } t \geq t_{0,\eta} = 2\eta. \tag{2.22}$$

Then we define, for $k \in \{1, \dots, k_0(\eta)\}$,

$$w_{NR}(t,\eta) := \begin{cases} \left(\frac{1+\delta^2|t-\eta/k|}{1+\delta^2|t_{k-1,\eta}-\eta/k|}\right)^{\delta_0} w_{NR}(t_{k-1,\eta},\eta) & \text{if } t \in [\eta/k, t_{k-1,\eta}], \\ \left(\frac{1}{1+\delta^2|t-\eta/k|}\right)^{1+\delta_0} w_{NR}(\eta/k,\eta) & \text{if } t \in [t_{k,\eta}, \eta/k]. \end{cases} \tag{2.23}$$

We define also the weight w_R by the formula

$$w_R(t,\eta) := \begin{cases} w_{NR}(t,\eta)\frac{1+\delta^2|t-\eta/k|}{1+\delta^2\eta/(8k^2)} & \text{if } |t-\eta/k| \le \eta/(8k^2), \\ w_{NR}(t,\eta) & \text{if } t \in I_{k,\eta},\ |t-\eta/k| \ge \eta/(8k^2), \end{cases} \tag{2.24}$$

for any $k \in \{1, \dots, k_0(\eta)\}$ and notice that for $t \in I_{k,\eta}$,

$$\frac{\partial_t w_{NR}(t,\eta)}{w_{NR}(t,\eta)} \approx \frac{\partial_t w_R(t,\eta)}{w_R(t,\eta)} \approx \frac{\delta^2}{1+\delta^2|t-\eta/k|}. \tag{2.25}$$

For small values of $t = (1-\beta)t_{k_0(\eta),\eta}$, $\beta \in [0,1]$, we define w_{NR} and w_R by the formulas

$$w_{NR}(t,\eta) = w_R(t,\eta) := \left(e^{-\delta\sqrt{\eta}}\right)^{\beta} w_{NR}(t_{k_0(\eta),\eta},\eta)^{1-\beta}. \tag{2.26}$$

If $\eta < -\delta^{-10}$, then we define $w_R(t,\eta) := w_R(t,|\eta|)$, $w_{NR}(t,\eta) := w_{NR}(t,|\eta|)$, and $I_{k,\eta} := I_{-k,-\eta}$. To summarize, the resonant intervals $I_{k,\eta}$ are defined for $(k,\eta) \in \mathbb{Z} \times \mathbb{R}$ satisfying $|\eta| > \delta^{-10}$, $1 \le |k| \le \sqrt{\delta^3|\eta|}$, and $\eta/k > 0$.

Finally, we define the weights $w_k(t,\eta)$ by the formula

$$w_k(t,\eta) := \begin{cases} w_{NR}(t,\eta) & \text{if } t \notin I_{k,\eta}, \\ w_R(t,\eta) & \text{if } t \in I_{k,\eta}. \end{cases} \tag{2.27}$$

If particular, $w_k(t,\eta) = w_{NR}(t,\eta)$ unless $|\eta| > \delta^{-10}$, $1 \le |k| \le \sqrt{\delta^3|\eta|}$, $\eta/k > 0$, and $t \in I_{k,\eta}$.

The functions w_{NR}, w_R, and w_k have the right size but lack optimal smoothness in the frequency parameter η, mainly due to the jump discontinuities of the function $k_0(\eta)$. This smoothness is important in symmetrization arguments (energy control of the transport terms) and in commutator arguments. To correct this problem, we fix $\varphi : \mathbb{R} \to [0,1]$ an even smooth function supported in $[-8/5, 8/5]$ and equal to 1 in $[-5/4, 5/4]$, and let $d_0 := \int_{\mathbb{R}} \varphi(x)\,dx$. For $k \in \mathbb{Z}$ and $Y \in \{NR, R, k\}$, let

$$\begin{aligned} b_Y(t,\xi) &:= \int_{\mathbb{R}} w_Y(t,\rho)\varphi\left(\frac{\xi-\rho}{L_{\delta'}(t,\xi)}\right)\frac{1}{d_0 L_{\delta'}(t,\xi)}\,d\rho, \\ L_{\delta'}(t,\xi) &:= 1 + \frac{\delta'\langle\xi\rangle}{\langle\xi\rangle^{1/2}+\delta' t}, \quad \delta' \in [0,1]. \end{aligned} \tag{2.28}$$

The length $L_{\delta'}(t,\xi)$ in (2.28) is chosen to optimize the smoothness in ξ of the functions $b_Y(t,\cdot)$, while not changing significantly the size of the weights. The parameter δ' is fixed sufficiently small, depending only on δ.

These definitions can be used to prove the key properties (2.15)–(2.16), as well as many other properties needed in the nonlinear analysis. We notice also that

$$\begin{aligned} e^{\lambda(t)\langle\xi\rangle^{1/2}} &\le A_{NR}(t,\xi) \le A_R(t,\xi) \le e^{\lambda(t)\langle\xi\rangle^{1/2}} e^{2\sqrt{\delta}\langle\xi\rangle^{1/2}}, \\ e^{\lambda(t)\langle k,\xi\rangle^{1/2}} &\le A_k(t,\xi) \le 2e^{\lambda(t)\langle k,\xi\rangle^{1/2}} e^{2\sqrt{\delta}\langle k,\xi\rangle^{1/2}}, \end{aligned} \tag{2.29}$$

for any $k \in \mathbb{Z}$, $t \geq 0$, and $\xi \in \mathbb{R}$. Finally, to prove commutator estimates in the context of our problem, we need to know that the weights vary sufficiently slowly in ξ. In our case the weights satisfy the key inequalities

$$\big|A_k(t,\xi) - A_k(t,\eta)\big| \lesssim \left[\frac{C(\delta)}{\langle k,\xi\rangle^{1/2}} + \sqrt{\delta}\right] \max\big\{A_k(t,\xi), A_k(t,\eta)\big\} \tag{2.30}$$

if $\langle \xi - \eta\rangle \lesssim 1 \ll \min\{\langle k,\xi\rangle, \langle k,\eta\rangle\}$. Such bounds are suitable to control the commutators by letting δ small enough, due to the gain of $\sqrt{\delta}$ at large frequencies.

2.3. The auxiliary nonlinear profile

In the case of general shear flows, an essential new difficulty that is not present in the Couette case, is the additional linear term $B''\partial_z\phi$ in (2.6). This linear term cannot be treated as a perturbation if b'' is not assumed small. On the linearized level, one can understand the evolution by using spectral analysis, especially the regularity analysis of generalized eigenfunctions corresponding to the continuous spectrum. However, it is still a challenge to combine the linear spectral analysis with the more sophisticated Fourier analysis tools needed for controlling the nonlinearity. We deal with this basic issue in two steps: first, we define an auxiliary nonlinear profile $F^*(t)$ given by

$$F^*(t,z,v) = F(t,z,v) - \int_0^t B''(0,v)\partial_z\phi'(s,z,v)\,ds. \tag{2.31}$$

Thus F^* takes into account the linear effect accumulated up to time t and can be bounded perturbatively, using the methods outlined in the previous subsection. The function ϕ' is a small but crucial modification of ϕ, defined as the unique solution to the elliptic equation

$$\begin{aligned} &\partial_z^2\phi' + (B_0')^2(\partial_v - t\partial_z)^2\phi' + B_0''(\partial_v - t\partial_z)\phi' = F,\\ &\phi'\big(t,b(0)\big) = \phi'\big(t,b(1)\big) = 0, \end{aligned} \tag{2.32}$$

on $\mathbb{T} \times [b(0), b(1)]$. This equation is obtained by freezing the coefficients of the main elliptic equation (2.7) at time $t = 0$ to gain additional smoothness.

On a heuristic level, we expect that the full evolution of F consists of two contributions: the main, linear evolution that changes the size of the profile most significantly, and a small but rough (compared with the linear evolution) nonlinear correction. We can view (2.31) as a bounded linear transformation in both space and time from F to F^* which takes into account the bulk linear evolution. The key point is that this transformation can be inverted to get bounds on the full profile F from bounds on F^*.

2.4. Control of the full profile

We still need to recover the bounds on F and the improved bounds on $F - F^*$. This is a critical step where we need to use our main spectral assumption and the precise estimates on the linearized flow. To link $F - F^*$ with the linearized flow, we define an auxiliary function ϕ^*, which can be approximately viewed as a stream function associated with F^*, and set $g = F - F^*$, $\varphi := \phi' - \phi^*$. The functions g and φ satisfy the inhomogeneous linear system with trivial initial data

$$\begin{aligned}&\partial_t g - B_0''(v)\partial_z\varphi = H, \quad g(0,z,v) = 0,\\ &B_0'(v)^2(\partial_v - t\partial_z)^2\varphi + B_0''(v)(\partial_v - t\partial_z)\varphi + \partial_z^2\varphi = g(t,z,v),\end{aligned} \tag{2.33}$$

where $(t,z,v) \in [0,\infty) \times \mathbb{T} \times [b(0), b(1)]$. The functions $B_0'(v) = B'(0,v)$ and $B_0''(v) = B''(0,v)$ are time-independent, very smooth, and can be expressed in terms of the original shear flow b. The source term H is given by $H = B_0''(v)\partial_z\phi^*$.

The function ϕ^* is determined by the auxiliary profile F^*. Since we have already proved quadratic bounds on the profile F^*, we can use elliptic estimates to prove quadratic bounds on ϕ^*, and then on the source term H. Therefore, we can think of (2.33) as a linear inhomogeneous system with trivial initial data, and adapt the linear theory to our situation.

Decomposing in modes, conjugating by e^{-ikvt}, and using Duhamel's formula, we can further reduce to the study of the homogeneous initial-value problem

$$\begin{aligned}&\partial_t g_k + ikvg_k - ikB_0''\varphi_k = 0, \quad g_k(0,v) = X_k(v)e^{-ikav},\\ &(B_0')^2\partial_v^2\varphi_k + B_0''(v)\partial_v\varphi_k - k^2\varphi_k = g_k, \quad \varphi_k\big(b(0)\big) = \varphi_k\big(b(1)\big) = 0.\end{aligned} \tag{2.34}$$

for $(t,v) \in [0,\infty) \times [b(0), b(1)]$, where $k \in \mathbb{Z} \setminus \{0\}$ and $a \in \mathbb{R}$.

2.5. Analysis of the linearized flow

Equation (2.34) was analyzed, at least when $a = 0$, by Wei–Zhang–Zhao in [45] and by the second author in [26]. We follow the approach in [26]. The main idea is to use the spectral representation formula and reduce the analysis of the linearized flow to the analysis of generalized eigenfunctions corresponding to the continuous spectrum.

More precisely, using general spectral theory, we can express the stream function as an oscillatory integral of the spectral density function (which depends both on the physical and the spectral variables). As a consequence, given data X_k smooth and satisfying $\operatorname{supp} X_k \subseteq [b(\vartheta_0), b(1-\vartheta_0)]$ we find a representation formula

$$\begin{aligned}\widetilde{g_k}(t,\xi) &= \widetilde{X_k}(\xi + kt + ka)\\ &\quad + ik\int_0^t\int_{\mathbb{R}} \widetilde{B_0''}(\zeta)\widetilde{\Pi_k'}(\xi + kt - \zeta - k\tau, \xi + kt - \zeta, a)\,d\zeta\,d\tau\end{aligned} \tag{2.35}$$

for the solution g_k of the linear evolution equation (2.34), where $\Pi_k'(\xi,\eta,a)$ can be expressed in terms of a family of generalized eigenfunctions. As proved in [26], these eigenfunctions cannot be calculated explicitly, but can be estimated very precisely in the Fourier space,

$$\big\|\big(|k| + |\xi|\big)W_k(\eta + ka)\widetilde{\Pi_k'}(\xi,\eta,a)\big\|_{L^2_{\xi,\eta}} \lesssim_\delta \big\|W_k(\eta)\widetilde{X_k}(\eta)\big\|_{L^2_\eta}, \tag{2.36}$$

for any $a \in \mathbb{R}$, for a large family of weights W_k that satisfy a slow variation property similar to (2.30). This leads to suitable control on the functions $g_k = F_k - F_k^*$, which allows us to close the bootstrap argument.

2.6. Energy functionals and the bootstrap proposition

We are now ready to summarize our main argument: given a solution $\omega : [0,T] \times \mathbb{T} \times [0,1] \to \mathbb{R}$ of equation (2.1), we define first the functions F, ϕ, V', V'', $\dot{V}$, B', B'', $\mathcal{H}$ as in (2.3)–(2.5) and (2.11). To construct useful energy functionals, we need to modify the functions V', B', B'' which are not "small," so we define the new variables

$$\begin{aligned} &B_0'(v) := B'(0,v) = (\partial_y b)\big(b^{-1}(v)\big), \quad B_0''(v) := B''(0,v) = (\partial_y^2 b)\big(b^{-1}(v)\big),\\ &V_*' := V' - B_0', \quad B_*' := B' - B_0', \quad B_*'' := B'' - B_0''. \end{aligned} \tag{2.37}$$

Our main goal is to control the functions F and ϕ. For this we need to consider two auxiliary functions F^* and ϕ', defined as in (2.31)–(2.32). Then we define the renormalized elliptic profiles

$$\begin{aligned} \Theta(t,z,v) &:= \big(\partial_z^2 + (\partial_v - t\partial_z)^2\big)\big(\Psi(v)\,\phi(t,z,v)\big),\\ \Theta^*(t,z,v) &:= \big(\partial_z^2 + (\partial_v - t\partial_z)^2\big)\big(\Psi(v)\,\big(\phi(t,z,v) - \phi'(t,z,v)\big)\big), \end{aligned} \tag{2.38}$$

where $\Psi : \mathbb{R} \to [0,1]$ is a Gevrey class cut-off function, satisfying

$$\begin{aligned} &\|e^{\langle\xi\rangle^{3/4}}\widetilde{\Psi}(\xi)\|_{L^\infty} \lesssim 1,\\ &\operatorname{supp}\Psi \subseteq \big[b(\vartheta_0/4), b(1-\vartheta_0/4)\big], \quad \Psi \equiv 1 \text{ in } \big[b(\vartheta_0/3), b(1-\vartheta_0/3)\big]. \end{aligned} \tag{2.39}$$

Our bootstrap argument is based on controlling simultaneously energy functionals and space-time integrals. For this we need carefully chosen weights A_{NR}, A_R, and A_k, defined as in Section 2.2.1. Let $\dot{A}_Y(t,\xi) := (\partial_t A_Y)(t,\xi) \leq 0$, $Y \in \{NR, R, k\}$, and define, for any $t \in [0,T]$,

$$\begin{aligned} \mathcal{E}_f(t) &:= \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} A_k^2(t,\xi)\big|\tilde{f}(t,k,\xi)\big|^2\,d\xi, \quad f \in \{F, F^*\},\\ \mathcal{B}_f(t) &:= \int_1^t \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \big|\dot{A}_k(s,\xi)\big|A_k(s,\xi)\big|\tilde{f}(s,k,\xi)\big|^2\,d\xi ds, \end{aligned} \tag{2.40}$$

$$\begin{aligned} \mathcal{E}_{F-F^*}(t) &:= \sum_{k\in\mathbb{Z}^*} \int_{\mathbb{R}} \big(1 + \langle k,\xi\rangle/\langle t\rangle\big)A_k^2(t,\xi)\big|\widetilde{(F-F^*)}(t,k,\xi)\big|^2\,d\xi,\\ \mathcal{B}_{F-F^*}(t) &:= \int_1^t \sum_{k\in\mathbb{Z}^*} \int_{\mathbb{R}} \big(1 + \langle k,\xi\rangle/\langle s\rangle\big)\big|\dot{A}_k(s,\xi)\big|A_k(s,\xi)\big|\widetilde{(F-F^*)}(s,k,\xi)\big|^2\,d\xi ds, \end{aligned} \tag{2.41}$$

$$\begin{aligned} \mathcal{E}_\Phi(t) &:= \sum_{k\in\mathbb{Z}^*} \int_{\mathbb{R}} A_k^2(t,\xi)\frac{|k|^2\langle t\rangle^2}{|\xi|^2 + |k|^2\langle t\rangle^2}\big|\widetilde{\Phi}(t,k,\xi)\big|^2\,d\xi, \quad \Phi \in \{\Theta, \Theta^*\},\\ \mathcal{B}_\Phi(t) &:= \int_1^t \sum_{k\in\mathbb{Z}^*} \int_{\mathbb{R}} \big|\dot{A}_k(s,\xi)\big|A_k(s,\xi)\frac{|k|^2\langle s\rangle^2}{|\xi|^2 + |k|^2\langle s\rangle^2}\big|\widetilde{\Phi}(s,k,\xi)\big|^2\,d\xi ds, \end{aligned} \tag{2.42}$$

$$\begin{aligned} \mathcal{E}_g(t) &:= \int_{\mathbb{R}} A_R^2(t,\xi)\big|\tilde{g}(t,\xi)\big|^2\,d\xi, \quad g \in \{V_*', B_*', B_*''\},\\ \mathcal{B}_g(t) &:= \int_1^t \int_{\mathbb{R}} \big|\dot{A}_R(s,\xi)\big|A_R(s,\xi)\big|\tilde{g}(s,\xi)\big|^2\,d\xi ds, \end{aligned} \tag{2.43}$$

$$
\begin{aligned}
\mathcal{E}_{\mathcal{H}}(t) &:= \mathcal{K}^2 \int_{\mathbb{R}} A_{NR}^2(t,\xi)\big(\langle t\rangle/\langle\xi\rangle\big)^{3/2}\big|\tilde{\mathcal{H}}(t,\xi)\big|^2\,d\xi,\\
\mathcal{B}_{\mathcal{H}}(t) &:= \mathcal{K}^2 \int_1^t \int_{\mathbb{R}} \big|\dot{A}_{NR}(s,\xi)\big|A_{NR}(s,\xi)\big(\langle s\rangle/\langle\xi\rangle\big)^{3/2}\big|\tilde{\mathcal{H}}(s,\xi)\big|^2\,d\xi ds,
\end{aligned}
\tag{2.44}
$$

where $\mathbb{Z}^* := \mathbb{Z}\setminus\{0\}$ and $\mathcal{K}\geq 1$ is a large constant that depends only on δ.

Our main bootstrap proposition is the following:

Proposition 2.1. *Assume $T\geq 1$ and $\omega\in C([0,T]:\mathcal{G}^{2\delta_0,1/2})$ is a sufficiently smooth solution of the system (2.1), with the property that $\omega(t)$ is supported in $\mathbb{T}\times[\vartheta_0,1-\vartheta_0]$ and that $\|\langle\omega\rangle(t)\|_{H^{10}}\ll 1$ for all $t\in[0,T]$. Define F, F^*, Θ, Θ^* B'_*, B''_*, V'_*, $\mathcal{H}$ as above. Assume that ε_1 is sufficiently small (depending on δ),*

$$
\sum_{g\in\{F,F^*,F-F^*,\Theta,\Theta^*,V'_*,B'_*,B''_*,\mathcal{H}\}} \mathcal{E}_g(t)\leq\varepsilon_1^3 \quad \textit{for any } t\in[0,1], \tag{2.45}
$$

and

$$
\sum_{g\in\{F,F^*,F-F^*,\Theta,\Theta^*,V'_*,B'_*,B''_*,\mathcal{H}\}} \big[\mathcal{E}_g(t)+\mathcal{B}_g(t)\big]\leq\varepsilon_1^2 \quad \textit{for any } t\in[1,T]. \tag{2.46}
$$

Then for any $t\in[1,T]$, we have the improved bounds

$$
\sum_{g\in\{F,F^*,F-F^*,\Theta,\Theta^*,V'_*,B'_*,B''_*,\mathcal{H}\}} \big[\mathcal{E}_g(t)+\mathcal{B}_g(t)\big]\leq\varepsilon_1^2/2. \tag{2.47}
$$

Moreover, we also have the stronger bounds for $t\in[1,T]$, namely

$$
\sum_{g\in\{F,\Theta\}} \big[\mathcal{E}_g(t)+\mathcal{B}_g(t)\big]\lesssim_\delta \varepsilon_1^3. \tag{2.48}
$$

This proposition is the main ingredient in the proof of Theorem 1.1 in [25]. Its proof is based on implementing the steps outlined in Sections 2.2–2.5. It is important to control not only the main variables F, Θ, F^* and Θ^*, but also the variables V'_*, B'_*, and B''_* which are connected to the change of variables $y\to v$. These variables appear in many nonlinear terms, so it is important to control their smoothness precisely, as part of a combined bootstrap argument, in a way that is consistent with the smoothness of the functions F and Θ.

The function $\mathcal{H}$ plays a different role, as it is the only variable that decays in time and encodes the convergence of the system as $t\to\infty$. This function decays at a rate of $\langle t\rangle^{-3/4}$, in a weaker topology, which shows that the function $\partial_v\dot{V}$ decays fast at an integrable rate of $\langle t\rangle^{-7/4}$, again in a weaker topology. We remark also that the bootstrap control on the variable $F-F^*$ is slightly stronger than on the variables F and F^* separately, which is needed to compensate for the lack of symmetry in some of the transport terms.

3. AN UNSTABLE MODEL: THE GENERALIZED SQG EQUATION

We consider now the generalized surface quasigeostrophic equations (gSQG)

$$
\begin{cases}
\partial_t\theta+u\cdot\nabla\theta=0, \quad (t,x)\in[0,T)\times\mathcal{D},\\
u=-\nabla^\perp(-\Delta)^{-1+\alpha/2}\theta,
\end{cases}
\tag{3.1}
$$

where $\alpha \in [0,2]$ and $\mathcal{D}$ is a domain in $\mathbb{R}^2$. The case $\alpha = 1$ corresponds to the surface quasi-geostrophic (SQG) equation, introduced by Constantin–Majda–Tabak [13] as a model for the full 3D Euler equations. Notice that the case $\alpha = 0$ corresponds to the 2D incompressible Euler equations and the case $\alpha = 2$ produces stationary solutions.

These are the so-called active scalar equations, which have been analyzed extensively both in the setting of smooth solutions θ and in the setting of the so-called α-patches, which are solutions for which θ is a step function. The local regularity theory is generally well understood: as expected, suitable initial data lead to local in time unique solutions that propagate the regularity of the initial data, both in the smooth and the patch setting (see, for example, [13,19,22,39] for regularity results of this type).

The construction of nontrivial global solutions for the gSQG equations is a very challenging open problem for all parameters $\alpha \in (0,2)$, both in the smooth and in the patch case (the construction of solutions that blow up in finite time is also a challenging open problem, but we will not discuss it here). In fact, the only known nonstationary global solutions of finite energy, both in the smooth and the patch setting, are special rotating solutions, periodic in time. See the recent work [11] for the construction of such solutions in the harder smooth case, and more references. See also [14] for the construction of a stable class of global solutions in the patch case, using the mechanism of dispersion, but which have infinite energy.

It is tempting to try to use the mechanism of inviscid damping to construct families of nontrivial global solutions of the gSQG equations, at least for some parameters $\alpha \in (0,2)$, by perturbing around stationary solutions. The easiest would be to perturb around shear flows on the finite channel domain $\mathcal{D} = \mathbb{T} \times [0,1]$, in particular around the Couette flow corresponding to $\theta(t,x,y) = -1$. The fractional Laplacian $(-\Delta)^{-1+\alpha/2}$ on the domain $\mathcal{D} = \mathbb{T} \times [0,1]$ can be defined using explicit spectral theory. The vorticity deviation $\omega = \theta + 1 : [0,T] \times \mathbb{T} \times [0,1] \to \mathbb{R}$ satisfies the system

$$\begin{aligned} &\partial_t \omega + \partial_y a(y) \partial_x \omega - \partial_y \psi \partial_x \omega + \partial_x \psi \partial_y \omega = 0, \\ &\psi = -(-\Delta)^{-1+\alpha/2} \omega, \quad \psi(t,x,1) = \psi(t,x,0) = 0, \end{aligned} \tag{3.2}$$

where $a = a(y)$ is given by $(-\partial_y^2)^{1-\alpha/2} a(y) = -1$, $a(0) = a(1) = 0$. Notice that if $\alpha = 0$ this is the same as the Euler equation (2.1) for the Couette flow $b(y) = y - 1/2$, as expected.

At first glance it seems plausible to adapt the ideas described in Sections 2.1–2.2 to prove global regularity of the system (3.2), at least for some $\alpha > 0$ small. One can still perform a nonlinear change of variables and derive a system of equations for a profile F, as in Section 2.1. A simplified version of this system is the closed equation

$$\partial_t F - \partial_v P_{\neq 0} \Phi \, \partial_z F = 0, \quad \widetilde{P_{\neq 0} \Phi}(t,k,\xi) = \frac{\tilde{F}(t,k,\xi)}{[k^2 + (\xi - tk)^2]^{1-\alpha/2}} \mathbf{1}_{\mathbb{Z}^*}(k) \tag{3.3}$$

for the smooth function $F : [0,T] \times \mathbb{T} \times \mathbb{R} \to \mathbb{R}$, which is analogous to the simplified equation (2.12) considered in Section 2.2.

Surprisingly, our analysis (in collaboration also with Javier Gómez-Serrano) reveals that the system (3.3) is unstable, for any $\alpha > 0$. To see this, let

$$\begin{aligned}\mathcal{E}_F(t) &:= \sum_{k\in\mathbb{Z}^*}\int_{\mathbb{R}} W_k^2(t,\xi)\big|\tilde{F}(t,k,\xi)\big|^2\,d\xi,\\ \mathcal{B}_F(t) &:= \sum_{k\in\mathbb{Z}^*}\int_0^t\int_{\mathbb{R}}\big|\dot{W}_k(s,\xi)\big|W_k(s,\xi)\big|\tilde{F}(s,k,\xi)\big|^2\,d\xi,\end{aligned} \tag{3.4}$$

where $\mathbb{Z}^* = \mathbb{Z}\setminus\{0\}$. We will show below that it is not possible to find a family of weights W_k, decreasing in t and compatible with nonlinear analysis, for which one could control the energy functional $\mathcal{E}_F$ for uniformly all times.

Indeed, we calculate

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_F(t) &= \sum_{k\in\mathbb{Z}^*}\int_{\mathbb{R}} 2\dot{W}_k(t,\xi)W_k(t,\xi)\big|\tilde{F}(t,k,\xi)\big|^2\,d\xi\\ &\quad + 2\Re\sum_{k\in\mathbb{Z}^*}\int_{\mathbb{R}} W_k^2(t,\xi)\partial_t\tilde{F}(t,k,\xi)\overline{\tilde{F}(t,k,\xi)}\,d\xi.\end{aligned} \tag{3.5}$$

Therefore, since $\partial_t W_k \le 0$, for any $t\in[1,T]$, we have

$$\mathcal{E}_F(t) + 2\mathcal{B}_F(t) = \mathcal{E}_F(0) + \int_0^t\bigg\{2\Re\sum_{k\in\mathbb{Z}^*}\int_{\mathbb{R}} W_k^2(s,\xi)\partial_s\tilde{F}(s,k,\xi)\overline{\tilde{F}(s,k,\xi)}\,d\xi\bigg\}ds. \tag{3.6}$$

Using equation (3.3), the cubic term on the right-hand side of (3.6) is equal to

$$\begin{aligned}&C\bigg|2\Re\bigg\{\sum_{k,\ell\in\mathbb{Z}^*}\int_0^t\int_{\mathbb{R}^2} W_k^2(s,\xi)i\eta\widetilde{\Phi}(s,\ell,\eta)i(k-\ell)\tilde{F}(s,k-\ell,\xi-\eta)\overline{\tilde{F}(s,k,\xi)}\,d\xi d\eta ds\bigg\}\bigg|\\ &\quad = C\bigg|2\Re\bigg\{\sum_{k,\ell\in\mathbb{Z}^*}\int_0^t\int_{\mathbb{R}^2} W_k^2(s,\xi)\frac{\eta\tilde{F}(s,\ell,\eta)\overline{\tilde{F}(s,k,\xi)}}{[\ell^2+(\eta-s\ell)^2]^{1-\alpha/2}}(k-\ell)\\ &\qquad\qquad \times\tilde{F}(s,k-\ell,\xi-\eta)\,d\xi d\eta ds\bigg\}\bigg|\\ &\quad = C\bigg|\sum_{k,\ell\in\mathbb{Z}^*}\int_0^t\int_{\mathbb{R}^2}\bigg[\frac{\eta W_k^2(s,\xi)}{[\ell^2+(\eta-s\ell)^2]^{1-\alpha/2}} - \frac{\xi W_\ell^2(s,\eta)}{[k^2+(\xi-sk)^2]^{1-\alpha/2}}\bigg]\\ &\qquad\quad \times\tilde{F}(s,\ell,\eta)\overline{\tilde{F}(s,k,\xi)}(k-\ell)\tilde{F}(s,k-\ell,\xi-\eta)\,d\xi d\eta ds\bigg|,\end{aligned} \tag{3.7}$$

where in the last identity we use symmetrization in (k,ξ) and (ℓ,η), based on the fact that F is real-valued.

We restrict ourselves to the range

$$\xi,\eta = N + O(1),\quad k=2,\ell=1, \tag{3.8}$$

where N is very large. This corresponds to the main "reaction term" in the original equation (3.3), where the frequency of Φ in the nonlinearity is large and the frequency of F is small.

To estimate the right-hand side of (3.6) using the bulk term $\mathcal{B}_F$ defined in (3.4), we need that the weights satisfy the inequality

$$\begin{aligned} &\left|\frac{\eta W_2^2(s,\xi)}{[1+(\eta-s)^2]^{1-\alpha/2}} - \frac{\xi W_1^2(s,\eta)}{[4+(\xi-2s)^2]^{1-\alpha/2}}\right| \\ &\qquad \lesssim \sqrt{|\dot{W}_2(s,\xi)| W_2(s,\xi)} \sqrt{|\dot{W}_1(s,\eta)| W_1(s,\eta)}, \end{aligned} \tag{3.9}$$

for all $\xi, \eta = N + O(1)$ and $s \in [0,\infty)$.

Assume that we are further restricting to a neighborhood of the largest resonant time $s = N + O(1)$. We notice that in this case the two terms on the left-hand side of (3.9) cannot have a meaningful cancelation because the denominator of the first term varies uniformly between 1 and C if ξ and s are fixed and $\eta = N + O(1)$, while all the other numerators and denominators vary much less. So we would need

$$\begin{aligned} &\frac{\eta W_2^2(s,\xi)}{[1+(\eta-s)^2]^{1-\alpha/2}} + \frac{\xi W_1^2(s,\eta)}{[4+(\xi-2s)^2]^{1-\alpha/2}} \\ &\qquad \lesssim \sqrt{|\dot{W}_2(s,\xi)| W_2(s,\xi)} \sqrt{|\dot{W}_1(s,\eta)| W_1(s,\eta)}, \end{aligned}$$

for all $\xi, \eta, s = N + O(1)$. In other words, the symmetrization performed in (3.7) does not help in the resonant case $\xi, \eta, s = N + O(1)$. In particular, for all $\eta, s = N + O(1)$,

$$N W_2^2(s,\eta) + N^{-1+\alpha} W_1^2(s,\eta) \lesssim W_2(s,\eta) W_1(s,\eta) \sqrt{\frac{|\dot{W}_2(s,\eta)|}{W_2(s,\eta)} \frac{|\dot{W}_1(s,\eta)|}{W_1(s,\eta)}}. \tag{3.10}$$

Using the mean inequality twice, this can only be satisfied if

$$N^{\alpha/2} \lesssim \frac{|\dot{W}_2(s,\eta)|}{W_2(s,\eta)} + \frac{|\dot{W}_1(s,\eta)|}{W_1(s,\eta)}, \quad \text{if } s, \eta = N + O(1). \tag{3.11}$$

Unfortunately, it is not possible to find suitable weights that satisfy a bound like (3.11), for any $\alpha > 0$. This is because the weights also need to satisfy basic bounds like

$$W_k(s,\xi) \approx W_k(s,\eta) \tag{3.12}$$

for any $s \in [0,\infty)$, $k \in \{1,2\}$, and $\xi, \eta \in \mathbb{R}$, $|\xi - \eta| \le 1$. These bounds are essential in order for the weights to be compatible with nonlinear analysis. Letting $W_k(s,\xi) = e^{\lambda_k(s,\xi)}$, $k \in \mathbb{Z}$, and $\lambda = \lambda_1 + \lambda_2$, it follows from (3.11)–(3.12) that $\lambda : [0,\infty) \times \mathbb{R} \to [0,\infty)$ is a decreasing function in s satisfying

$$\langle \eta \rangle^{\alpha/2} \lesssim \left|(\partial_s \lambda)(s,\eta)\right|, \quad \left|\lambda(s,\eta) - \lambda(s,\xi)\right| \lesssim 1 \tag{3.13}$$

if $\eta \gg 1$, $|\eta - s| \le 2$, and $|\xi - \eta| \le 2$. We use these inequalities with $s = \eta = N \gg 1$ and recall that $\alpha > 0$ to see that

$$\lambda(N-1, N-1) \ge \lambda(N,N) + cN^{\alpha/2}. \tag{3.14}$$

We can then apply this inductively to conclude that $\lambda(N-n, N-n) \ge \lambda(N,N) + cnN^{\alpha/2}$ for $n = 1, \ldots, N/2$. In particular, $\lambda(N/2, N/2) \ge cN^{1+\alpha/2}$, which would force $\lambda(0, N/2) \ge cN^{1+\alpha/2}$ (since λ is decreasing in s). However, this is not compatible with the bounds (3.12) when $s = 0$, giving the final contradiction.

Notice that most of this argument applies in the Euler case $\alpha = 0$, except that (3.13) does not imply (3.14) (in fact, our weights A_k constructed in Section 2.2.1 satisfy (3.13) but not (3.14)). To summarize, these calculations show that the main construction used in the proof of global stability of the Couette flow for the 2D Euler equations does not extend to any more singular generalized SQG equations.

REFERENCES

[1] V. Arnold and B. Khesin, *Topological methods in hydrodynamics*. Appl. Math. Sci., 125. Springer, New York, 1998.

[2] A. P. Bassom and A. D. Gilbert, The spiral wind-up of vorticity in an inviscid planar vortex. *J. Fluid Mech.* **371** (1998), 109–140.

[3] A. P. Bassom and A. D. Gilbert, The relaxation of vorticity fluctuations in approximately elliptical streamlines. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **456** (2000), 295–314.

[4] J. Bedrossian, M. Coti Zelati, and V. Vicol, Vortex axisymmetrization, inviscid damping, and vorticity depletion in the linearized 2D Euler equations. *Ann. PDE* **5** (2019), no. 1, Art. 4, 192 pp.

[5] J. Bedrossian, P. Germain, and N. Masmoudi, On the stability threshold for the 3D Couette flow in Sobolev regularity. *Ann. of Math. (2)* **185** (2017), 541–608.

[6] J. Bedrossian and S. He, Inviscid damping and enhanced dissipation of the boundary layer for 2D Navier–Stokes linearized around Couette flow in a channel. *Comm. Math. Phys.* **379** (2020), no. 1, 177–226.

[7] J. Bedrossian and N. Masmoudi, Inviscid damping and the asymptotic stability of planar shear flows in the 2D Euler equations. *Publ. Math. Inst. Hautes Études Sci.* **122** (2015), 195–300.

[8] F. Bouchet and H. Morita, Large time behavior and asymptotic stability of the 2D Euler and linearized Euler equations. *Phys. D* **239** (2010), 948–966.

[9] M. Brachet, M. Meneguzzi, H. Politano, and P. Sulem, The dynamics of freely decaying two-dimensional turbulence. *J. Fluid Mech.* **194** (1988), 333–349.

[10] K. M. Case, Stability of inviscid plane Couette flow. *Phys. Fluids* **3** (1960), 143–148.

[11] A. Castro, D. Córdoba, and J. Gómez-Serrano, Global smooth solutions for the inviscid SQG equation. *Mem. Amer. Math. Soc.* **266** (2020), no. 1292, v+89 pp.

[12] Q. Chen, T. Li, D. Wei, and Z. Zhang, Transition threshold for the 2-D Couette flow in a finite channel. *Arch. Ration. Mech. Anal.* **238** (2020), no. 1, 125–183.

[13] P. Constantin, A. Majda, and E. Tabak, Formation of strong fronts in the 2-D quasigeostrophic thermal active scalar. *Nonlinearity* **7** (1994), 1495–1533.

[14] D. Córdoba, J. Gómez-Serrano, and A. D. Ionescu, Global solutions for the generalized SQG patch equation. *Arch. Ration. Mech. Anal.* **233** (2019), 1211–1251.

[15] M. Coti Zelati and C. Zillinger, On degenerate circular and shear flows: the point vortex and power law circular flows. *Comm. Partial Differential Equations* **44** (2019), no. 2, 110–155.

[16] Y. Deng and N. Masmoudi, Long time instability of the Couette flow in low Gevrey spaces. Preprint (2018), arXiv:1803.01246.

[17] L. Faddeev, On the theory of the stability of stationary plane parallel flows of an ideal fluid. In *Boundary-value problems of mathematical physics and related problems of function theory*, Part 5, Zap. Nauchn. Sem. LOMI, 21, "Nauka," Leningrad. Otdel., Leningrad, 1971, 164–172.

[18] T. Gallay, Enhanced dissipation and axisymmetrization of two-dimensional viscous vortices. *Arch. Ration. Mech. Anal.* **230** (2018), 939–975.

[19] F. Gancedo, Existence for the α-patch model and the QG sharp front in Sobolev spaces. *Adv. Math.* **217** (2008), no. 6, 2569–2598.

[20] E. Grenier, T. Nguyen, F. Rousset, and A. Soffer, Linear inviscid damping and enhanced viscous dissipation of shear flows by using the conjugate operator method. *J. Funct. Anal.* **278** (2020), no. 3, 108339, 27 pp.

[21] I. Hall, A. Bassom, and A. Gilbert, The effect of fine structures on the stability of planar vortices. *Eur. J. Mech. B Fluids* **22** (2003), no. 2, 179–198.

[22] I. M. Held, R. T. Pierrehumbert, S. T. Garner, and K. L. Swanson, Surface quasi-geostrophic dynamics. *J. Fluid Mech.* **282** (1995), 1–20.

[23] A. Ionescu and H. Jia, Inviscid damping near the Couette flow in a channel. *Comm. Math. Phys.* **374** (2020), no. 3, 2015–2096.

[24] A. Ionescu and H. Jia, Axi-symmetrization near point vortex solutions for the 2D Euler equation, *Comm. Pure Appl. Math.* (to appear), arXiv:1904.09170.

[25] A. Ionescu and H. Jia, Nonlinear inviscid damping near monotonic shear flows. *Acta Math.* (to appear), arXiv:2001.03087.

[26] H. Jia, Linear inviscid damping in Gevrey spaces. *Arch. Ration. Mech. Anal.* **235** (2020), no. 2, 1327–1355.

[27] Lord Kelvin, Stability of fluid motion: rectilinear motion of viscous fluid between two plates. *Philos. Mag.* **24** (1887), 188–196.

[28] G. Kirchhoff, *Vorlesungen über mathematische Physik*. Leipzig, B. G. Teubner, 1876.

[29] C. C. Lin, On the motion of vortices in two dimensions, I. Existence of the Kirchhoff–Routh function. *Proc. Natl. Acad. Sci. USA* **27** (1941), 570–575.

[30] Z. Lin and C. Zeng, Inviscid dynamical structures near Couette flow. *Arch. Ration. Mech. Anal.* **200** (2011), 1075–1097.

[31] A. J. Majda and A. L. Bertozzi, *Vorticity and incompressible flow*. Cambridge Texts Appl. Math., 27. Cambridge University Press, Cambridge, 2002.

[32] C. Marchioro and M. Pulvirenti, Vortices and localization in Euler flows. *Comm. Math. Phys.* **154** (1993), no. 1, 49–61.

[33] N. Masmoudi and W. Zhao, Nonlinear inviscid damping for a class of monotone shear flows in finite channel. 2020, arXiv:2001.08564.

[34] J. McWilliams, The emergence of isolated coherent vortices in turbulent flow. *J. Fluid Mech.* **146** (1984), 21–43.

[35] J. McWilliams, The vortices of two-dimensional turbulence. *J. Fluid Mech.* **219** (1990), 361–385.

[36] C. Mouhot and C. Villani, On Landau damping. *Acta Math.* **207** (2011), 29–201.

[37] W. Orr, The stability or instability of the steady motions of a perfect liquid and of a viscous liquid, Part I: A perfect liquid. *Proc. R. Ir. Acad., A Math. Phys. Sci.* **27** (1907), 9–68.

[38] Lord Rayleigh, On the stability, or instability, of certain fluid motions. *Proc. Lond. Math. Soc.* **S1-11** (1880), 57–72.

[39] J. L. Rodrigo, On the evolution of sharp fronts for the quasi-geostrophic equation. *Comm. Pure Appl. Math.* **58** (2005), no. 6, 821–866.

[40] P. Santangelo, R. Benzi, and B. Legras, The generation of vortices in high-resolution, two-dimensional decaying turbulence and the influence of initial conditions on the breaking of self-similarity. *Phys. Fluids A, Fluid Dyn.* **1** (1989), 1027–1034.

[41] D. Schecter, D. Dubin, A. Cass, C. Driscoll, I. Lansky, T. O'Neil, Inviscid damping of asymmetries on a two-dimensional vortex. *Phys. Fluids* **12** (2000), no. 10, 2397–2412.

[42] S. A. Stepin, The nonselfadjoint Friedrichs model in the theory of hydrodynamic stability (Russian). *Funktsional. Anal. i Prilozhen.* **29** (1995), no. 2, 22–35; translation in *Funct. Anal. Appl.* **29** (1995), no. 2, 91–101.

[43] V. Sverak, Lecture notes on fluid mechanics, http://www-users.math.umn.edu/~sverak/course-notes2011.pdf.

[44] G. Taylor, Stability of a viscous liquid contained between two rotating cylinders. *Philos. Trans. Roy. Soc. A* **223** (1923), 289–343.

[45] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping for a class of monotone shear flow in Sobolev spaces. *Comm. Pure Appl. Math.* **71** (2018), 617–687.

[46] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping and vorticity depletion for shear flows. *Ann. PDE* **5** (2019), no. 1, Paper No. 3, 101 pp.

[47] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping and enhanced dissipation for the Kolmogorov flow. *Adv. Math.* **362** (2020), 106963, 103 pp.

[48] W. Wolibner, Un theorème sur l'existence du mouvement plan d'un fluide parfait, homogene, incompressible, pendant un temps infiniment long. *Math. Z.* **37** (1933), 698–726.

[49] C. Zillinger, Linear inviscid damping for monotone shear flows in a finite periodic channel, boundary effects, blow-up and critical Sobolev regularity. *Arch. Ration. Mech. Anal.* **221** (2016), 1449–1509.

[50] C. Zillinger, Linear inviscid damping for monotone shear flows. *Trans. Amer. Math. Soc.* **369** (2017), no. 12, 8799–8855.

ALEXANDRU D. IONESCU
Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA, aionescu@math.princeton.edu

HAO JIA
Department of Mathematics, University of Minnesota, 206 Church St. S.E., Minneapolis, MN 55455, USA, jia@umn.edu

MEAN-FIELD LIMITS FOR QUANTUM SYSTEMS AND NONLINEAR GIBBS MEASURES

MATHIEU LEWIN

ABSTRACT

We consider the linear Schrödinger equation describing N quantum (bosonic) particles at equilibrium and study its behavior as N tends to infinity. We place the system in the mean-field regime, in which the particles are very tightly packed but interact weakly. In this limit we prove that they become essentially independent and identically distributed according to a nonlinear partial differential equation. Our main tool is the quantum de Finetti theorem, an abstract result about how independence can arise due to symmetry in such systems. By considerably increasing the randomness in the system, we can also obtain nonlinear Gibbs measures. Those are probability measures over an infinite-dimensional space, which play a major role in different areas of mathematics. The two- and three-dimensional cases are particularly challenging due to the necessity of using a renormalization procedure to cancel infinities.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 35Q40; Secondary 81V73

KEYWORDS

Mean-field limits, quantum de Finetti theorem, Bose–Einstein condensates, nonlinear Gibbs measures

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3800–3821
DOI 10.4171/ICM2022/53

1. INTRODUCTION

Mathematics is efficient in describing some aspects of our world [63]. Many complicated natural phenomena are well reproduced using rather simple equations. More than that, abstract results or principles can sometimes even be used to *predict* new phenomena, later confirmed by experiments. This happened many times in physics in the 20th century, in particular through *arguments based on symmetry*. Several elementary particles were discovered this way (such as the positron predicted by Dirac in 1928 and discovered later by Anderson in 1932 or, more recently, the Higgs boson). In this respect, mathematics is not just an efficient tool to model our world, it can sometimes also be used to explore it.

Quantum mechanics is certainly one of the physical theories relying the most on mathematics. This is in part due to the strong influence of Hilbert in Göttingen, where Heisenberg and Born invented the "new quantum mechanics" around 1925 [56,58]. In fact, several mathematical concepts used all the time today (such as the Hilbert space) have been invented in this context [62]. This is a culmination of Hilbert's program to axiomatize physics, his 6th problem at the International Congress of Mathematics in 1900 [31].

At about the same time that quantum mechanics was being formalized, Bose [10] and Einstein [19] predicted the existence of a new state of matter, now called a *Bose–Einstein condensate*. Their argument was again mainly based on symmetry. Under the assumption that the wave function of a set of N particles is invariant under the action of the permutation group $\mathfrak{S}_N$, they found that those particles would have to behave rather strangely when N gets large, at very low temperatures. When the temperature passes below some critical value, they start traveling through the whole system at macroscopic distances, and all adopt the exact same behavior on average. A condensate is thus a macroscopic piece of quantum matter, where quantum effects can almost be observed with the human eye. The particles respecting this symmetry are now called *bosons*; famous examples include photons and helium atoms. Even if Bose–Einstein condensation (BEC) was suspected to play a role in many experiments (e.g., for the superfluidity of liquid helium), it was only in 1995 that a condensate could finally be realized in the laboratory [4,13]. This was recognized by a Nobel prize in 2001 and is still a very active subject of research in theoretical and experimental physics.

The argument of Bose and Einstein concerned noninteracting particles, and it can be made rigorous. However, real particles interact with each other and providing a mathematical proof of condensation in this case turned out to be very difficult, even at zero temperature. This was finally achieved in a series of works by Lieb, Seiringer, and Yngvason [43–46,50] starting in 1998. These works belong to a large trend of research in analysis and mathematical physics which was stimulated by the numerous experimental discoveries starting from 1995.

In this paper I present the results obtained on the subject with my collaborators in the four articles [37,38,41,42] published in the period 2014–2021. In [37] we realized with Phan Thành Nam and Nicolas Rougerie that BEC can, to some extent, be understood through a purely abstract result, the *quantum de Finetti theorem*. A version of this theorem was proved in 1969 [33,60] within the framework of operator algebras, and it currently plays an important role in quantum information theory. Its use for the condensation of bosons had, however, been rather anecdotal. The *classical* version of de Finetti's theorem dates back to 1931 and is often called the *Hewitt–Savage theorem* [14,30]. It plays a central role in probability and statistics. Loosely speaking, the latter says that a sequence of infinitely many *exchangeable random variables* is essentially automatically independent and identically distributed (i.i.d.). More precisely, its law must be the convex combination of i.i.d. random variables. Similarly, we will see that the emergent macroscopic i.i.d. behavior of the bosons in a condensate is a consequence of their indistinguishability, which implies a certain symmetry under permutations.

This point of view allowed us to push forward the mathematical analysis of condensates. In particular, in [38,41] we started to look at a new situation where the randomness between the condensed particles is considerably increased, due to the temperature. This corresponded to looking at how the condensate is forming, just before the phase transition. We showed that the condensed bosons are then described by *nonlinear Gibbs measures*. These probability measures in infinite dimension play a major role in several areas of mathematics. They, for instance, appear in the study of rough stochastic partial differential equations and of deterministic equations with random initial data (as promoted by Bourgain [11,12] and now studied by many authors). Our new program has generated some interest and important achievements followed, in particular by Fröhlich, Knowles, Schlein, and Sohinger [22–24].

We shall restrict here our attention to a particular regime, called the *mean-field limit*. The system is assumed to be very dense (hence the particles meet very often) but the particles interact only a little. The many-particle interaction then gets replaced by an *effective nonlinear interaction*, seen by all the particles in the system, which leads to a nonlinear partial differential equation. This is not the most common regime in experiments [4,13]. The system is often rather dilute such that the particles instead meet rarely. The Lieb–Seiringer–Yngvason analysis in this case is more involved and requires more assumptions [43–46,50].

In the next section we introduce the Schrödinger model for N bosons. In Section 3 we review our main results on Bose–Einstein condensation in the mean-field limit from [37,42]. We then turn in Section 4 to a different mean-field regime where nonlinear Gibbs measures appear [38,41]. Due to space limitations, we will avoid entering too much into the technical details. We will also not be able to cite all the existing literature. In addition to [37,38,41,42], we refer to a previous proceedings [36] for more references (in particular on the physics side), and to [57] for a recent and detailed review of known results.

2. THE N-PARTICLE QUANTUM MODEL

We consider a system composed of N identical particles evolving in $\mathbb{R}^d$. Physically $d \in \{1,2,3\}$, but for the moment any $d \geqslant 1$ is allowed. We assume that interactions take place by pairs and are described with an even potential $w : \mathbb{R}^d \to \mathbb{R}$. We ignore more complicated events involving three or more particles at a time. We also submit our system to an external potential $V : \mathbb{R}^d \to \mathbb{R}$, which is typically used to ensure that the particles do not escape.

In classical mechanics, our particles would be described by N vectors $\{(x_j, p_j)\}_{j=1}^N$ in $\mathbb{R}^d \times \mathbb{R}^d$, where x_j is the position of the jth particle and $p_j = mv_j$ is its momentum (mass times velocity). The time evolution is a Hamiltonian system based on the energy

$$\mathcal{H}_{\mathrm{cl}}(x_1,p_1,\dots,x_N,p_N)=\sum_{j=1}^N\frac{|p_j|^2}{2m}+\sum_{j=1}^N V(x_j)+\lambda\sum_{1\leqslant j<k\leqslant N}w(x_j-x_k),$$

with the usual symplectic form on the phase space. The three terms are respectively the kinetic energy, the potential energy, and the interaction energy. We have inserted a *coupling constant* λ which we will later use to tune the strength of the interaction between the particles. The usual Hamilton equations lead to Newton's equations that the acceleration is proportional to the forces felt by each particle (which depend on the positions of all the others). Stationary states correspond to critical points of $\mathcal{H}_{\mathrm{cl}}$. Those always have all the p_j equal to 0 (the particles do not move!). Equilibrium states are those where $\mathcal{H}_{\mathrm{cl}}$ is a local minimum. Of interest are also measures on the phase space $(\mathbb{R}^d\times\mathbb{R}^d)^N$ which are globally invariant under the Hamiltonian flow. This is the case of any function of the conserved Hamiltonian $\mathcal{H}_{\mathrm{cl}}$ but an important example is given by the *Gibbs measures*

$$\begin{aligned}&\mathcal{P}(x_1,p_1,\dots,x_N,p_N)=Z^{-1}e^{-T^{-1}\mathcal{H}_{\mathrm{cl}}(x_1,p_1,\dots,x_N,p_N)},\\&Z:=\int_{\mathbb{R}^{2dN}}e^{-T^{-1}\mathcal{H}_{\mathrm{cl}}}\mathrm{d}x_1\cdots\mathrm{d}p_N,\end{aligned}\tag{2.1}$$

where T is a temperature used to model the amount of randomness in the system. In the limit $T\to 0^+$, the probability measure $\mathcal{P}$ concentrates on the minimum of $\mathcal{H}_{\mathrm{cl}}$.

For microscopic particles such as atoms, the classical model is not sufficiently precise and one has to switch to quantum mechanics. The basic principle is to give up the idea that one can know the exact positions and momenta of the particles. Instead, quantum mechanics provides us with two probability measures on $(\mathbb{R}^d)^N$ corresponding to the possible positions and momenta, respectively. These two probability measures are not independent, on account of Heisenberg's uncertainty principle which states that positions and velocities cannot be known simultaneously to an arbitrary precision. This principle is mathematically expressed using the Fourier transform. Namely, our N quantum particles are represented by a square-integrable function $\Psi\in L^2(\mathbb{R}^{dN},\mathbb{C})$ called the *wave function*, normalized in the manner $\int_{\mathbb{R}^{dN}}|\Psi|^2=1$, and it is postulated that

- $|\Psi(x_1,\dots,x_N)|^2$ is the probability density that the particles are at $x_1,\dots,x_N\in\mathbb{R}^d$;
- $|\widehat{\Psi}(p_1,\dots,p_N)|^2$ is the probability density that they have the momenta $p_1,\dots,p_N$.[1]

Integrating the above two probability densities against the classical energy and using that $\widehat{-i\partial_{x_j}\Psi}=p_j\widehat{\Psi}$, we find that the quantum energy can be expressed in terms of Ψ as

$$\mathcal{E}(\Psi)=\frac{1}{2m}\int_{\mathbb{R}^{dN}}|\nabla\Psi|^2+\sum_{j=1}^N\int_{\mathbb{R}^{dN}}V(x_j)|\Psi|^2+\lambda\sum_{1\leqslant j<k\leqslant N}\int_{\mathbb{R}^{dN}}w(x_j-x_k)|\Psi|^2.\tag{2.2}$$

[1] Here $\widehat{\Psi}(p_1,\dots,p_N)=(2\pi)^{-\frac{dN}{2}}\int_{\mathbb{R}^{dN}}\Psi(x_1,\dots,x_N)e^{-i\sum_{j=1}^N x_j\cdot p_j}\mathrm{d}x_1\cdots\mathrm{d}x_N$.

This is the quadratic form associated with the operator

$$\boxed{H_{N,\lambda} := \sum_{j=1}^{N} \frac{-\Delta_{x_j}}{2m} + \sum_{j=1}^{N} V(x_j) + \lambda \sum_{1\leqslant j<k\leqslant N} w(x_j - x_k)} \tag{2.3}$$

which is our main object of interest. We will work in a system of units so that $2m = 1$.

Since our N particles are all the same, the two probability densities $|\Psi|^2$ and $|\widehat{\Psi}|^2$ must be *symmetric functions*, that is, invariant if we permute their variables. Some additional constraints are thus needed on Ψ. Only two possible choices can preserve the linear character of quantum mechanics, namely Ψ must itself be either symmetric or antisymmetric. This corresponds to the two types of quantum particles existing in Nature, respectively called *bosons* and *fermions*. In this paper we exclusively consider the bosonic case, and hence restrict $H_{N,\lambda}$ to the subspace of symmetric square-integrable functions, denoted by

$$\begin{aligned} &L^2_s\big((\mathbb{R}^d)^N, \mathbb{C}\big) \\ &\quad = \big\{\Psi \in L^2\big((\mathbb{R}^d)^N, \mathbb{C}\big) \ : \ \Psi(x_{\sigma(1)}, \ldots, x_{\sigma(N)}) = \Psi(x_1, \ldots, x_N), \ \forall \sigma \in \mathfrak{S}_N\big\}. \end{aligned}$$

We emphasize that each x_j is in $\mathbb{R}^d$. We are not permuting the coordinates of a given particle.

The quantum model is again a Hamiltonian system, in infinite dimension. Equilibrium states are critical points of the quantum energy $\mathcal{E}$ in (2.2), on the unit sphere of $L^2_s((\mathbb{R}^d)^N, \mathbb{C})$. Those are exactly the symmetric eigenfunctions of the Hamiltonian $H_{N,\lambda}$, which solve Schrödinger's equation

$$H_{N,\lambda}\Psi = E\Psi. \tag{2.4}$$

We will be particularly interested in what is called the *ground state* (the equilibrium state of lowest possible energy), that is, the first eigenfunction. The corresponding energy is

$$\boxed{E(N,\lambda) := \min \sigma(H_{N,\lambda}) = \inf_{\int |\Psi|^2 = 1} \mathcal{E}(\Psi)}$$

where $\sigma(H)$ denotes the spectrum of an operator H. Other states of interest are *quantum Gibbs states*, which are given by a formula similar to the classical case (2.1) by

$$\Gamma_{T,N,\lambda} := Z^{-1}_{T,N,\lambda} e^{-T^{-1} H_{N,\lambda}}, \quad Z_{T,N,\lambda} = \mathrm{Tr}(e^{-T^{-1} H_{N,\lambda}}), \tag{2.5}$$

with the trace taken only over the symmetric subspace $L^2_s((\mathbb{R}^d)^N, \mathbb{C})$. Those are compact operators which involve the whole spectrum of the quantum operator $H_{N,\lambda}$. The corresponding free energy of the system is then given by

$$F(T,N,\lambda) := -T \log Z_{T,N,\lambda} \tag{2.6}$$

and it converges to $E(N,\lambda)$ in the limit $T \to 0^+$. We postpone the presentation of the precise assumptions on the potentials V and w which ensure that this is all well defined.

Finding the equilibrium states (2.4) or the Gibbs state (2.5) requires diagonalizing the operator $H_{N,\lambda}$. Due to the high dimensionality of the problem, this is impossible in most physical situations, even numerically at a sufficiently high precision. It is therefore important to rely on simpler approximations that are both precise enough and suitable to numerical investigation. One of the most famous is the *mean-field model*, which consists in assuming that the particles are independent but evolve in an effective, self-consistent, potential which replaces the many-particle interaction. The *linear* many-body Schrödinger equation (2.4) on $\mathbb{R}^{dN}$ is then replaced by a more tractable *nonlinear* equation in $\mathbb{R}^d$. Introduced by Curie and Weiss to describe phase transitions in the classical Ising model, the mean-field method is now extremely popular in many areas of physics, and has even spread to other fields like biology and social sciences. We explain in the next section how the N-particle quantum system in fact converges to such a nonlinear problem in a specific limit.

3. MEAN-FIELD LIMIT TO THE GROSS–PITAEVSKII EQUATION

Gross–Pitaevskii theory. In a fully condensed system, the N bosons are by definition i.i.d. and the corresponding wave function is factorized, that is,

$$\Psi(x_1,\dots,x_N) = u^{\otimes N}(x_1,\dots,x_N) := u(x_1)\cdots u(x_N), \tag{3.1}$$

for some normalized $u \in L^2(\mathbb{R}^d,\mathbb{C})$. After some computation, one finds that the energy of such a state equals $\mathcal{E}(\Psi) = N\mathcal{E}_{\mathrm{GP}}(u)$ where $\mathcal{E}_{\mathrm{GP}}$ is the *Gross–Pitaevskii (GP) energy* [28,54],

$$\begin{aligned}\mathcal{E}_{\mathrm{GP}}(u) = {} & \int_{\mathbb{R}^d} \left|\nabla u(x)\right|^2 \mathrm{d}x + \int_{\mathbb{R}^d} V(x)\left|u(x)\right|^2 \mathrm{d}x \\ & + \frac{\lambda(N-1)}{2} \iint_{\mathbb{R}^d\times\mathbb{R}^d} w(x-y)\left|u(x)\right|^2\left|u(y)\right|^2 \mathrm{d}x\,\mathrm{d}y.\end{aligned} \tag{3.2}$$

It is often also called "Hartree" when w is a smooth function. When w is proportional to a Dirac delta, one often uses the acronym NLS for "nonlinear Schrödinger." Historically designed to describe quantized vortices in superfluid helium (in which it applies to only a small fraction of the particles), the Gross–Pitaevskii model is now the main tool to study Bose–Einstein condensates. If we minimize over all normalized u, we obtain the smallest possible energy per particle of a fully condensed system

$$\boxed{e_{\mathrm{GP}} := \inf_{\int_{\mathbb{R}^d}|u|^2=1} \mathcal{E}_{\mathrm{GP}}(u).} \tag{3.3}$$

An associated minimizer u_0, when it exists, solves the nonlinear eigenvalue equation

$$\big(-\Delta + V(x) + \lambda(N-1)|u_0|^2 * w(x)\big)u_0(x) = \varepsilon_0 u_0(x), \tag{3.4}$$

where ε_0 is a Lagrange multiplier associated with the normalization constraint in $L^2(\mathbb{R}^d)$. The nonlinearity is only through the "mean-field potential" $|u_0|^2 * w$. Equation (3.4) has been used with impressive success to describe Bose–Einstein condensates. A famous example is the vortices appearing in rotating gases, see Figure 1. Note that $\mathcal{E}_{\mathrm{GP}}$ also provides a Hamiltonian system, whose dynamics is given by the nonlinear Schrödinger equation

$$i\partial_t u = \big(-\Delta + V + \lambda(N-1)|u|^2 * w\big)u. \tag{3.5}$$

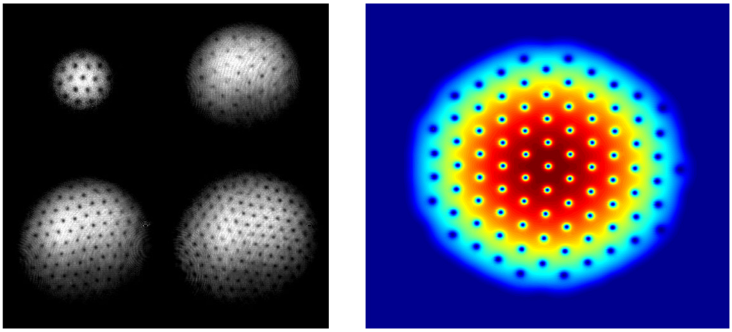


FIGURE 1

(Left) Experimental pictures of the density of fast rotating Bose–Einstein condensates from [1] (© AAAS with permission). (Right) Numerical calculation of $|u_0|^2$ for the Gross–Pitaevskii solution u_0 of (3.4) with additional terms describing the rotation, using GPELab [5] (© Antoine & Duboscq with permission). The dots are small vortices appearing under the effect of rotation, which seem to be placed on a triangular lattice [2].

Mean-field limit. The proof of Bose–Einstein condensation requires understanding how independence arises in an interacting system. The interactions will have to be weak and there are (at least) two ways this could happen. The first is when they are *rare*, which is the *dilute regime* appropriate for many experiments. Another situation is when the interactions are *small in amplitude*, that is, λ is small. In order to have them play a role, many collisions are then needed. This corresponds to a *high density regime* where the particles meet very often but interact only a little bit each time. The latter is our mean-field regime. Surprisingly, very similar theorems are expected in the two opposite regimes.

The mean-field regime corresponds to taking $N \to \infty$, with the potential V used to confine most particles to a finite region of space. At the same time, we take $\lambda \to 0$. From the formula (3.2) of the GP energy and the GP equation (3.4), we see that the interesting regime is $\lambda \sim 1/N$. This makes the quantum Hamiltonian $H_{N,\lambda}$ essentially of order N. To simplify some expressions including (3.2) and (3.4), we simply choose

$$\boxed{\lambda = \frac{1}{N-1}}$$

and denote $H_N := H_{N,\lambda}$, $E(N) = E(N,\lambda)$, etc. Then $e_{\rm GP}$ is independent of N. Using $E(N) \leqslant \mathcal{E}(u^{\otimes N}) = N\mathcal{E}_{\rm GP}(u)$ and minimizing over u, we obtain the simple upper bound

$$E(N) \leqslant Ne_{\rm GP}. \tag{3.6}$$

We need technical assumptions on the potentials V and w to make everything meaningful. In the whole paper we distinguish two situations:

- (*confined case*) w and $V_- = \max(0, -V)$ are in $L^p(\mathbb{R}^d, \mathbb{R}) + L^\infty(\mathbb{R}^d, \mathbb{R})$ with $p = 1$ if $d = 1$, $p > 1$ if $d = 2$ and $p = d/2$ if $d \geqslant 3$, V_+ is in $L^1_{\rm loc}(\mathbb{R}^d, \mathbb{R})$ and diverges at infinity;

- (*unconfined or locally-confined case*) V and w are in $L^p(\mathbb{R}^d, \mathbb{R}) + L^\infty(\mathbb{R}^d, \mathbb{R})$ with p as above, and tend to 0 at infinity.

These conditions are not optimal and can be weakened in several ways. The most important is that we make no assumption on the sign of w or its Fourier transform $\hat{w}$. The interaction can be repulsive, attractive, or both.

Theorem 1 (Convergence of energy [37]). *Under the previous assumptions, we have*

$$\lim_{N\to\infty} \frac{E(N)}{N} = e_{\mathrm{GP}}. \tag{3.7}$$

In the confined case, if, in addition, $\int_{\mathbb{R}^d} e^{-T^{-1}V(x)}\mathrm{d}x < \infty$ for all $T > 0$, then we have the same limit for $F(T, N)$ in (2.6), for all $T > 0$.

Similar results have been shown in many particular situations, but Theorem 1 is, to my knowledge, the first generic result. Important previous works in the same spirit include [6, 49] for unconfined systems, and [20, 55, 61] in the confined case. The limit for $F(T, N)$ says that the temperature plays no role at the considered scale. In Section 4 we look at the case where $T \to \infty$ and get a different limit. A simple proof of Theorem 1, different from that of [37], is provided in the appendix. The argument is inspired of [34, 49] and was first written in the proceedings [36] but, unfortunately, never published. It is also described in [57, CHAP. 2].

The reader should think that our model is expressed at the *macroscopic scale* where condensation happens. But interactions typically take place at the microscopic scale. After changing units, in $d = 3$ the more physical dilute limit corresponds to replacing $w(x)$ by $N^3 w(Nx)$ in our model. That the interaction becomes N-dependent generates many difficulties. Under further assumptions on w, the same limit (3.7) was proved in [43, 44, 46], with w replaced by $8\pi a\delta$ in $\mathcal{E}_{\mathrm{GP}}$, with $a > 0$ the scattering length of w and δ the Dirac delta. The positive temperature case is handled in [16, 17].

Our next goal is to prove that the system is really condensed, that is, the bosons are essentially independent. Unfortunately, one cannot expect that Ψ_N will be close in norm to a factorized state such as (3.1). Changing one of the u's in the tensor product (that is, exciting one particle out of the condensate) would only affect the total energy to an $O(1)$ hence not change the main result. The proper way to express Bose–Einstein condensation is, following Penrose and Onsager [53], through the corresponding *k-particle density matrices*. Those are the quantum equivalent of marginals in probability theory, which appear for events involving only k particles at a time. They are defined for all $k \geqslant 1$ through their integral kernel by

$$\Gamma^{(k)}_{\Psi_N}(x_1, \dots, x_k, y_1, \dots, y_k) := \frac{N!}{(N-k)!} \int_{\mathbb{R}^{d(N-k)}} \Psi_N(x_1, \dots, x_k, Z)\overline{\Psi(y_1, \dots, y_k, Z)}\, dZ.$$

This is a compact operator with trace equal to $\frac{N!}{(N-k)!}$, hence an operator norm bounded by

$$\left\|\Gamma^{(k)}_{\Psi_N}\right\| \leqslant \frac{N!}{(N-k)!} \underset{N\to\infty}{\sim} N^k. \tag{3.8}$$

For large bosonic systems in normal conditions (such as a gas or a solid), $\Gamma^{(k)}_{\Psi_N}$ will stay of order one. Penrose and Onsager [53] suggested that a signature of Bose–Einstein condensation is when some eigenvalues are of order N^k. In fact, factorized states such as (3.1) saturate the bound in (3.8) since their k-particle density matrices are all of rank one:

$$\Gamma^{(k)}_{u^{\otimes N}} = \frac{N!}{(N-k)!} |u^{\otimes k}\rangle\langle u^{\otimes k}|.$$

Here we use the bra-ket notation for the operator $|f\rangle\langle g|u := \langle g, u\rangle f$. The following says that, in the mean-field limit, condensation happens precisely on the set of minimizers of the Gross–Pitaevskii energy.

Theorem 2 (Convergence of states [37]). *Under the previous assumptions on V and w, let Ψ_N be any sequence such that $\langle \Psi_N, H_N\Psi_N\rangle = E(N) + o(N)$. In the* confined case*, there exist a subsequence $N_j \to \infty$ and a probability measure μ on $\mathcal{M} = \{\text{minimizers for } e_{\mathrm{GP}}\}$, invariant under multiplication by phase factors, such that*

$$\lim_{N_j\to\infty} \frac{\Gamma^{(k)}_{\Psi_{N_j}}}{(N_j)^k} = \int_{\mathcal{M}} |u^{\otimes k}\rangle\langle u^{\otimes k}| \,\mathrm{d}\mu(u), \quad \forall k \geqslant 1, \tag{3.9}$$

strongly in trace norm. In the unconfined or locally-confined case*, the result is the same except that $\mathcal{M} = \{\text{weak limits of minimizing sequences for } e_{\mathrm{GP}}\}$ and the limit in* (3.9) a priori *only holds weakly-$*$ in the trace class.*

The simplest case is when $\mathcal{M} = \{e^{i\theta}u_0,\ \theta \in [0, 2\pi)\}$, that is, the GP minimizer is unique modulo phase. Then there will always be complete Bose–Einstein condensation on u_0 and no need to extract subsequences. One should probably think of μ as a probability over experiments, where only one GP minimizer u_0 is usually observed at a time. Note that it is possible to construct sequences Ψ_N converging to any chosen probability μ on $\mathcal{M}$. The exact ground state of H_N might converge to a definite μ, but this is not addressed in the theorem. In the unconfined or locally-confined case, the result only gives condensation for the particles that stay in a neighborhood of the origin (due to the weak limits in the statement). All the information about the particles escaping to infinity is lost.

The main tool for proving Theorems 1 and 2 is the *quantum de Finetti theorem*. The following is our version of this result from [37], which involves weak limits and is thus stronger than the historical theorem from [33, 60].

Theorem 3 (Quantum de Finetti [33, 37, 60]). *Let Ψ_N be any sequence of normalized symmetric wave functions in $L^2_s((\mathbb{R}^d)^N, \mathbb{C})$. Assume that the k-particle density matrices satisfy*

$$\frac{\Gamma^{(k)}_{\Psi_N}}{N^k} \underset{*}{\rightharpoonup} \Upsilon^{(k)}, \quad \forall k \geqslant 1, \tag{3.10}$$

weakly-$*$ in the trace class. Then there exists a probability measure μ on the unit ball $B = \{u \in L^2(\mathbb{R}^d),\ \|u\|_{L^2} \leqslant 1\}$, invariant under multiplication by phase factors, such that

$$\Upsilon^{(k)} = \int_B |u^{\otimes k}\rangle\langle u^{\otimes k}| \,\mathrm{d}\mu(u), \quad \textit{for all } k \geqslant 1.$$

Convergence holds in trace norm in (3.10) *for one (hence all) $k \geqslant 1$ if and only if μ is supported on the unit sphere $S = \{u \in L^2(\mathbb{R}^d),\ \|u\|_{L^2} = 1\}$.*

The result is, in fact, valid in any separable Hilbert space, but we used $L^2(\mathbb{R}^d)$ to avoid introducing any new notation. A similar theorem appeared earlier in [3] but it was formulated differently. Note that since the k-particle density matrices are bounded in trace norm after dividing by N^{-k}, the limit (3.10) always holds for a subsequence, by the Banach–Alaoglu theorem. Theorem 3 says that whatever converges at the scale N^k has to come from factorized states, that is, condensates. This abstract result is only a consequence of the symmetry of bosonic states and it is valid for *any* sequence Ψ_N, irrespective of the physical context. This justifies and goes much further than the theory of Penrose and Onsager [53].

The above quantum de Finetti theorem makes the proof of Theorems 1 and 2 very simple for confined systems. The main observation is that the energy can be written in terms of the two-particle density matrix as follows:

$$\frac{\mathcal{E}(\Psi_N)}{N} = \frac{\operatorname{Tr}(H_2 \Gamma^{(2)}_{\Psi_N})}{2N(N-1)}.$$

After extracting a subsequence, we can assume that $N_j^{-k}\Gamma^{(k)}_{\Psi_{N_j}} \rightharpoonup_* \Upsilon^{(k)}$ weakly-$*$, for some $\Upsilon^{(k)}$ and all $k \geqslant 1$. For confined systems, H_2 has a compact resolvent and the energy bounds can be used to show that the limit holds in trace norm. Using Fatou's lemma for the trace and the quantum de Finetti Theorem 3 for $\Gamma^{(2)}$, we infer that

$$\begin{aligned}\liminf_{N\to\infty} \frac{E(N)}{N} &= \liminf_{N_j\to\infty} \frac{\operatorname{Tr}(H_2\Gamma^{(2)}_{\Psi_{N_j}})}{2N_j(N_j-1)} \geqslant \frac{1}{2}\operatorname{Tr}\big(H_2\Upsilon^{(2)}\big) = \int_S \frac{\langle u^{\otimes 2}, H_2 u^{\otimes 2}\rangle}{2}\, \mathrm{d}\mu(u) \\ &= \int_S \mathcal{E}_{\mathrm{GP}}(u)\,\mathrm{d}\mu(u) \geqslant e_{\mathrm{GP}} \int_S \mathrm{d}\mu(u) = e_{\mathrm{GP}}. \qquad (3.11)\end{aligned}$$

The upper bound (3.6) implies that there is equality everywhere, hence $E(N)/N \to e_{\mathrm{GP}}$ and μ is supported on the set of minimizers for e_{GP}. This concludes the proof of Theorems 1 and 2 for confined systems. An argument of the same kind appeared before in [20, 55, 61].

For unconfined systems, some particles can escape to infinity and the argument is much more complicated. The weak limit in (3.10) might not provide sufficient information. In [37] we treated separately the particles staying in a neighborhood of 0 (for which the quantum de Finetti theorem is valid) and those that escape. All the possible cases of K particles escaping and $N-K$ staying, with K of the order of N have to be considered. These different events are handled using a technique introduced in [35] together with the concentration–compactness method. This is, in fact, also the idea of the proof of Theorem 3.

The Bogoliubov correction. The convergence of the density matrices says very little about the behavior of the wave function Ψ_N itself. This problem requires determining the next order in the energy expansion, called the Bogoliubov correction. We quickly present here the results obtained in this direction in [42]. Our idea was to concentrate on the excitations outside of the condensate. We noticed that, given a reference normalized function u_0 in $L^2(\mathbb{R}^d)$ – for instance, a GP minimizer – any N-particle wave function can be uniquely decomposed in the form

$$\Psi_N = \varphi_0 u_0^{\otimes N} + \varphi_1 \otimes_s u_0^{\otimes N-1} + \cdots + \varphi_{N-1} \otimes_s u_0 + \varphi_N \qquad (3.12)$$

where the φ_j are completely orthogonal to u_0, that is, belong to $(\{u_0\}^\perp)^{\otimes j}$. Here $\otimes_s$ is a notation for the symmetrized tensor product, whose precise definition can be found in [42].

The map $\Psi_N \mapsto \varphi_0 \oplus \varphi_1 \oplus \cdots \oplus \varphi_{N-1} \oplus \varphi_N$ is a unitary operator from the N-particle space $L^2_s((\mathbb{R}^d)^N, \mathbb{C})$ to the *truncated bosonic Fock space*

$$\mathcal{F}_+^{\leq N} = \mathbb{C} \oplus \bigoplus_{n=1}^{N} \bigotimes_s^n \{u_0\}^\perp$$

which later became known as the *excitation map*. In the mean-field regime the φ_j's will converge to a limit in the full Fock space $\mathcal{F}_+ := \mathcal{F}_+^{\leq\infty}$ and describe the excitations.

We have seen that the leading order of the energy is given by the Gross–Pitaevskii minimization. Bogoliubov predicted in [9] that the next order can be expressed using the Hessian of $\mathcal{E}_{\rm GP}$ at the GP minimizer u_0, a bit like in a Taylor expansion (the gradient of $\mathcal{E}_{\rm GP}$ does not appear since u_0 is a critical point of $\mathcal{E}_{\rm GP}$ on the unit sphere). More precisely, the Hessian has to be *second-quantized* on the Fock space $\mathcal{F}_+$, which provides the so-called *Bogoliubov Hamiltonian*, defined using creation and annihilation operators by

$$\begin{aligned}\mathbb{H}_0 = &\int a^\dagger(x)\big(-\Delta + V + |u_0|^2 * w - \varepsilon_0\big)a(x)\,\mathrm{d}x \\ &+ \iint u_0(x)\overline{u_0(y)}w(x-y)a^\dagger(x)a(y)\,\mathrm{d}x\,\mathrm{d}y \\ &+ \frac{1}{2}\iint w(x-y)\big(u_0(x)u_0(y)a^\dagger(x)a^\dagger(y) + \overline{u_0(x)u_0(y)}a(x)a(y)\big)\mathrm{d}x\,\mathrm{d}y. \end{aligned} \tag{3.13}$$

It would take us too far to explain this formula in detail and we refer to [42]. The form of the spectrum of the operator $\mathbb{H}_0$ is important to explain the superfluidity of cold Bose gases. This spectrum occurs in the mean-field limit, as specified in the following result.

Theorem 4 (Validity of Bogoliubov's theory [42]). *We work in the confined case and assume that $e_{\rm GP}$ admits a* unique and nondegenerate *minimizer u_0 (modulo phase), which satisfies*

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} w(x-y)^2\big|u_0(x)\big|^2\big|u_0(y)\big|^2\,\mathrm{d}x\,\mathrm{d}y < \infty.$$

Then, for every fixed j, the jth eigenvalue (counted with multiplicity) satisfies

$$\lim_{N\to\infty}\big(\lambda_j(H_N) - Ne_{\rm GP}\big) = \lambda_j(\mathbb{H}_0). \tag{3.14}$$

The first eigenvalue $\lambda_1(\mathbb{H}_0)$ is always simple, with corresponding normalized ground state denoted by $\Phi = \{\varphi_n\}_{n\geq 0} \in \mathcal{F}_+$ (defined up to a phase factor). The lowest eigenfunction Ψ_N of H_N is also simple and, with a correct choice of phase, we have

$$\lim_{N\to\infty}\left\|\Psi_N - \sum_{n=0}^{N}\varphi_n \otimes_s (u_0)^{\otimes N-n}\right\| = 0. \tag{3.15}$$

A similar convergence holds for the higher eigenfunctions, up to subsequences in case of degeneracy.

The limit (3.15) provides the exact behavior of Ψ_N, which involves the condensate u_0 and all its excitations φ_n. That a second-quantized model arises for the excitations is well explained using the excitation map associated with the decomposition (3.12). Our result was stimulated by [25,27,48,59]. Many other works followed. The similar result in the much more complicated dilute regime has been open for a long time and was only solved very recently in several groundbreaking works [7,8,21,51]. At a fixed temperature $T > 0$, Bogoliubov theory also predicts the $O(1)$ correction to $Ne_{\rm GP}$ in the expansion of $F(T, N, \lambda)$ in (2.6) [37].

4. DERIVATION OF NONLINEAR GIBBS MEASURES

We have seen in Theorem 2 that the condensed particles can be represented by a probability measure μ concentrated on the set $\mathcal{M}$ of minimizers of the Gross–Pitaevskii energy. This naturally raises the question of whether one can get other kinds of measures μ in a mean-field limit. Introducing a fixed temperature T as in Theorem 1 will not change anything at that scale. In [38] we proposed that taking $T \to \infty$ at a proper speed (depending on N) should lead to a *nonlinear Gibbs measure* μ and proved this in dimension $d = 1$. The much more complicated dimensions $d \in \{2, 3\}$ were only solved later in [41] and, simultaneously, in [24] with a completely different method.

Note that since we are working at the macroscopic scale, the parameter T is not the real thermodynamic temperature of the system. After reexpressing everything in microscopic units, our limit rather corresponds to looking just above the critical temperature, right before the condensation has started to appear [41, APP. B]. Thus the nonlinear Gibbs measures are describing the way that the Bose–Einstein condensate forms.

The nonlinear Gibbs measure is formally given by

$$\boxed{\mathrm{d}\mu(u) = z^{-1} e^{-\mathcal{E}_{\rm GP}(u) - \kappa \int_{\mathbb{R}^d} |u|^2}\, \mathrm{d}u,} \tag{4.1}$$

where z is a normalization factor used to make μ a probability, and where we have perturbed the GP energy by a multiple of the mass $\int_{\mathbb{R}^d} |u|^2$, for convenience. This is the same as changing V into $V + \kappa$. The constant $-\kappa$ has the physical interpretation of a chemical potential and we have $\kappa = -\varepsilon_0$ in the GP equation (3.4). From a Hamiltonian system point of view, we are considering the linear combination of two conserved quantities (energy and mass), which are both constants along the nonlinear GP flow (3.5). We could also insert a temperature in (4.1), but we have taken it equal to one to avoid introducing too many parameters.

The formula (4.1) is completely formal. There is nothing such as $\mathrm{d}u$ for functions u in infinite dimension. Things are a little bit easier if we look at the noninteracting problem, that is, take $w = 0$. In the confined case, we call $(-\Delta + V)v_j = \lambda_j v_j$ the eigenfunctions and eigenvalues of $-\Delta + V$. We then choose the constant κ so that $-\Delta + V + \kappa > 0$, that is, $\kappa > -\lambda_1(-\Delta + V)$. The formal probability measure

$$\mathrm{d}\mu_0(u) = z_0^{-1} e^{-\langle u, (-\Delta + V + \kappa) u\rangle}\, \mathrm{d}u \tag{4.2}$$

is a Gaussian measure in infinite dimension. It is by definition the unique probability measure whose cylindrical projection to the finite-dimensional space $\mathrm{span}(v_1, \dots, v_J)$ equals the normalized Gaussian on $\mathbb{C}^J$, that is,

$$\mathrm{d}\mu_{0,J}(u) := \prod_{j=1}^{J} \frac{(\lambda_j + \kappa) e^{-(\lambda_j + \kappa)|u_j|^2}}{\pi} \mathrm{d}u_j,$$

where $u_j := \langle v_j, u\rangle$, for any $J \geqslant 1$. Under appropriate growth assumptions on V, this provides a well-defined probability measure. Note, however, that we always have $\langle u, (-\Delta + V)u\rangle = +\infty$ μ_0-almost surely, which is why (4.2) is purely formal.

In dimension $d = 1$, the Gaussian measure μ_0 concentrates on functions in $L^2(\mathbb{R})$, and we can then define μ using μ_0 as reference in the form

$$\mathrm{d}\mu(u) = z^{-1} e^{-\mathcal{I}(u)}\,\mathrm{d}\mu_0(u), \tag{4.3}$$

where

$$\mathcal{I}(u) := \frac{1}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \left|u(x)\right|^2 \left|u(y)\right|^2 w(x-y)\mathrm{d}x\mathrm{d}y, \quad z := \int e^{-\mathcal{I}(u)}\,\mathrm{d}\mu_0(u). \tag{4.4}$$

We will always assume that $\mathcal{I}(u) \geqslant 0$, which, for instance, follows if $w \geqslant 0$ or $\hat{w} \geqslant 0$ (defocusing case). If $\mathcal{I}(u)$ is not infinite μ_0-almost surely, we conclude that $0 < z < \infty$ and hence μ is a well-defined probability measure, absolutely continuous with respect to μ_0. The situation is much more complicated in dimensions $d \geqslant 2$, since μ_0 always concentrates on distributions. Then $|u(x)|^2$ does not make any sense and thus $\mathcal{I}(u)$ is not defined. It is necessary to remove infinities in $\mathcal{I}$ by a *renormalization procedure*.

The one-dimensional case. We first discuss the mean-field limit in dimension $d = 1$, following [38]. The Gibbs measure μ in (4.3) lives on the whole space $L^2(\mathbb{R}^d)$. It is not restricted to the unit sphere as was the case in Theorem 2. To obtain μ, we need to work grand-canonically, that is, average over all possible numbers of particles in a kind of Laplace transform. The corresponding quantum state takes the form, in Fock space,

$$\Gamma_{\kappa,\lambda,T} := Z_{\kappa,\lambda,T}^{-1} \bigoplus_{n\geqslant 0} e^{-T^{-1}(H_{n,\lambda}+\kappa n)}, \quad Z_{\kappa,\lambda,T} = 1 + \sum_{n\geqslant 1} e^{-T^{-1}\kappa n}\,\mathrm{Tr}(e^{-T^{-1}H_{n,\lambda}}),$$

and its density matrices are given by

$$\Gamma_{\kappa,\lambda,T}^{(k)}(X,Y) := Z_{\kappa,\lambda,T}^{-1} \sum_{n\geqslant k} \frac{n!}{(n-k)!} e^{-T^{-1}\kappa n} \int_{(\mathbb{R}^d)^{n-k}} e^{-T^{-1}H_{n,\lambda}}(X,Z;Y,Z)\,\mathrm{d}Z, \tag{4.5}$$

with $X = (x_1, \dots, x_k)$ and $Y = (y_1, \dots, y_k)$. In statistical mechanics, it is frequent to work in the grand-canonical setting, which has a much simpler algebra. It is often easy to subsequently infer a result without any average over N, but we have not yet investigated this question.

Theorem 5 (Derivation of nonlinear Gibbs measures in dimension $d = 1$ [38]). *Let $d = 1$. We work in the confined case and assume, in addition, that $\mathrm{Tr}(-\Delta + V + \kappa)^{-1} < \infty$ for some (hence all) $\kappa > -\lambda_1(-\Delta + V)$. Let $w = w_1 + w_2$ with w_1 a finite positive Borel measure on $\mathbb{R}$ and $0 \leqslant w_2 \in L^\infty(\mathbb{R})$. For any $\kappa > -\lambda_1(-\Delta + V)$ and any $k \geqslant 1$, we have the convergence*

$$\lim_{\substack{\lambda\to 0^+ \\ \lambda T\to 1}} \lambda^k \Gamma_{\kappa,\lambda,T}^{(k)} = \int \left|u^{\otimes k}\right\rangle\!\left\langle u^{\otimes k}\right| \mathrm{d}\mu(u) \tag{4.6}$$

in trace norm, where μ is the nonlinear Gibbs measure defined in (4.3).

Note the assumption that w is nonnegative, which implies $\mathcal{I}(u) \geqslant 0$. Since the number of particles has been averaged over, hence is not at our disposal anymore, the limit (4.6) involves the parameter λ with $T \sim \lambda^{-1} \to \infty$. In fact, the limit (4.6) also says that the average number of particles in the Gibbs state is of order λ^{-1}:

$$Z_{\kappa,\lambda,T}^{-1} \sum_{n \geqslant 1} n e^{-T^{-1}\kappa n} \operatorname{Tr}(e^{-T^{-1}H_{n,\lambda}}) = \operatorname{Tr}\big(\Gamma_{\kappa,\lambda,T}^{(1)}\big) \sim \lambda^{-1} \int \|u\|_{L^2(\mathbb{R}^d)}^2 \mathrm{d}\mu(u).$$

The same limit as (4.6) is expected for the N-particle Gibbs state in (2.5), when $T \sim N$ and μ is replaced by its restriction to the unit sphere. The assumptions of the theorem have been weakened in [39].

Renormalization in two and three dimensions. In physics, renormalization is not just about removing undesired infinities. The removal of the bad terms must be justified by *only changing physical parameters in the system* [15, 18]. This is unfortunately often neglected in mathematical works on the subject. Here we will see that the theory can be made finite by *only adjusting the constant κ*. For simplicity, we explain the construction at the level of the GP energy, by formally manipulating infinite quantities. This will better motivate the final result in the quantum case.

Let V_0 be any potential (to be specified later) and μ_0 be the associated Gaussian measure as in (4.2) with V replaced by V_0 and $\kappa > -\lambda_1(-\Delta + V_0)$. We can renormalize the undefined $|u(x)|^2$ using *Wick ordering* [26], which formally amounts to replacing $|u(x)|^2$ by

$$: \big|u(x)\big|^2 :_{\mu_0} = \big|u(x)\big|^2 - \int \big|u(x)\big|^2 \mathrm{d}\mu_0(u) = \big|u(x)\big|^2 - (-\Delta + V_0 + \kappa)^{-1}(x,x). \quad (4.7)$$

We hope here that the divergence of $|u(x)|^2$ is essentially independent of u, so that subtracting the average can remove it for μ_0-almost every u. Of course, the counter term is also infinite. In dimensions $d \geqslant 2$, the kernel $(-\Delta + V_0 + \kappa)^{-1}(x,y)$ of the resolvent of $-\Delta + V_0 + \kappa$, called the *Green function*, diverges when $x \to y$, at a speed depending on d. In the lower dimensions $d \in \{2,3\}$, the limit

$$\lim_{x \to y} \big((-\Delta + V_0 + \kappa)^{-1} - (-\Delta + \kappa)^{-1}\big)(x,y) \quad (4.8)$$

exists for all $\kappa > \max(0, -\lambda_1(-\Delta + V_0))$, under suitable assumptions on V_0, hence the divergence is the same as that of $(-\Delta + \kappa)^{-1}(x,y)$. Since $(-\Delta + \kappa)^{-1}$ is a translation-invariant operator, its integral kernel depends on $x - y$. It is known to diverge like $\log|x - y|^{-1}$ in dimension $d = 2$ and $|x - y|^{-1}$ in dimension $d = 3$. In dimensions $d \geqslant 4$, $(-\Delta + V_0 + \kappa)^{-1}(x,y)$ diverges like $|x - y|^{2-d}$ but there are lower divergences which remain after subtracting $(-\Delta + \kappa)^{-1}(x,y)$. With the Wick ordering (4.7), we can formally define the renormalized interaction energy by

$$\mathcal{I}_{\mathrm{r}}(u) := \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} : \big|u(x)\big|^2 :_{\mu_0} : \big|u(y)\big|^2 :_{\mu_0} w(x-y)\, \mathrm{d}x\, \mathrm{d}y. \quad (4.9)$$

The proper mathematical definition requires to first project u onto the finite-dimensional space spanned by the J first eigenfunctions of $-\Delta + V_0$, subtract the average against μ_0 and then take the limit $J \to \infty$, see [41]. This limit exists μ_0-almost surely in dimensions

$d \in \{2,3\}$, under suitable assumptions on V_0 and w described below. In addition, $\mathcal{I}_{\rm r}(u) \geqslant 0$ for $\hat{w} \geqslant 0$ and this allows us to define a renormalized Gibbs measure by

$$\mathrm{d}\mu_{\rm r}(u) = z_r^{-1} e^{-\mathcal{I}_{\rm r}(u)}\, \mathrm{d}\mu_0(u), \quad z_{\rm r} := \int e^{-\mathcal{I}_{\rm r}(u)}\, \mathrm{d}\mu_0(u). \tag{4.10}$$

Our initial goal was to construct and derive the (formal) measure μ in (4.1). If we just pick $V_0 = V$, then the new measure $\mu_{\rm r}$ seems very different from μ. It contains undesired additional terms in the interaction. More precisely, $\mu_{\rm r}$ involves a modified GP energy which, after expanding $\mathcal{I}_{\rm r}$, can formally be expressed as

$$\langle u, (-\Delta + V_0 + \kappa)u\rangle + \mathcal{I}_{\rm r}(u) = \langle u, (-\Delta + V_0 - W_{\kappa,V_0} + \kappa - \alpha)u\rangle + \mathcal{I}(u) + \beta, \tag{4.11}$$

where we have introduced the two infinite constants

$$\begin{aligned} \alpha &= \int_{\mathbb{R}^d} w(x-y)(-\Delta+\kappa)^{-1}(y,y)\,\mathrm{d}y = (-\Delta+\kappa)^{-1}(0,0)\int_{\mathbb{R}^d} w = +\infty, \\ \beta &= \frac{1}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} (-\Delta+V_0+\kappa)^{-1}(x,x)(-\Delta+V_0+\kappa)^{-1}(y,y)w(x-y)\mathrm{d}x\,\mathrm{d}y = +\infty, \end{aligned} \tag{4.12}$$

as well as the finite potential

$$W_{\kappa,V_0}(x) = \int_{\mathbb{R}^d} w(x-y)\big((-\Delta+V_0+\kappa)^{-1} - (-\Delta+\kappa)^{-1}\big)(y,y)\,\mathrm{d}y.$$

The computation (4.11) suggests to search for a potential V_0 solving the nonlinear equation

$$V_0 - W_{\kappa,V_0} = V, \tag{4.13}$$

called the *counter-term problem* in [22]. Then we have the formal equality

$$\langle u, (-\Delta + V_0 + \kappa)u\rangle + \mathcal{I}_{\rm r}(u) = \mathcal{E}_{\rm GP}(u) + (\kappa - \alpha)\int_{\mathbb{R}^d} |u|^2 + \beta. \tag{4.14}$$

Adding the infinite constant β has no effect since it is then removed when we divide by z in (4.1). Our renormalized measure $\mu_{\rm r}$ thus formally coincides with the desired μ in (4.1), but with κ shifted by the infinite constant α. This shows that it is, in principle, possible to only rely on κ, if we choose a reference potential V_0 solving the nonlinear equation (4.13). A similar situation was encountered before in [29,47]. For an interpretation in terms of quasifree states, see [41]. When $\kappa \to \infty$, the nonlinear potential W_{κ,V_0} tends to 0 and the following could be proved using a Banach fixed point.

Theorem 6 (Counter-term problem [22,41]). *Let $d \in \{2,3\}$. Assume that V satisfies*

$$\frac{1+|x|^s}{C} \leqslant V(x) \leqslant C\big(1+|x|^s\big), \quad \textit{for some } C > 0 \textit{ and } s > \frac{2d}{d-4}, \tag{4.15}$$

and that w is an even function in $L^1(\mathbb{R}^d, (1+|x|^{2s})\,\mathrm{d}x)$ such that $\hat{w}$ is nonnegative and belongs to $L^1(\mathbb{R}^d, (1+|k|^2)\,\mathrm{d}k)$. Then there exists $\bar{\kappa}$ such that equation (4.13) admits a unique solution V_0 satisfying $V/2 \leqslant V_0 \leqslant 3V/2$, for all $\kappa > \bar{\kappa}$.

On the quantum side, there are no infinities and everything is perfectly well defined. However, we need to take a divergent sequence of constants κ in the mean-field limit, in order to account for the above renormalization of the chemical potential.

Theorem 7 (Derivation of nonlinear Gibbs measures in dimension $d \in \{2,3\}$ [24, 41]). *Let $d \in \{2,3\}$ and V, w as in Theorem 6. For any $\kappa > \bar{\kappa}$, define*

$$\begin{aligned}\kappa_\lambda &:= \kappa - \frac{\int_{\mathbb{R}^d} w}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\lambda\,\mathrm{d}k}{e^{\lambda(|k|^2+\kappa)}-1} \\ &= \begin{cases} \kappa - \frac{\log(\kappa\lambda)^{-1}}{4\pi} \int_{\mathbb{R}^d} w + o(1)_{\lambda\to 0} & \text{for } d=2, \\ \kappa - \Big(\frac{\zeta(3/2)}{8\pi^{\frac{3}{2}}\sqrt{\lambda}} - \frac{\sqrt{\kappa}}{4\pi}\Big) \int_{\mathbb{R}^d} w + o(1)_{\lambda\to 0} & \text{for } d=3. \end{cases}\end{aligned} \tag{4.16}$$

The density matrices in (4.5) satisfy

$$\lim_{\substack{\lambda\to 0^+ \\ \lambda T\to 1}} \lambda^k \Gamma^{(k)}_{\kappa_\lambda,\lambda,T} = \int |u^{\otimes k}\rangle\langle u^{\otimes k}|\,\mathrm{d}\mu_{\mathrm{r}}(u) \tag{4.17}$$

in Hilbert–Schmidt norm, where μ_{r} is the nonlinear Gibbs measure (4.10) with μ_0 defined using the solution V_0 of the nonlinear equation (4.13) in place of V.

The case $d = 2$ was announced earlier in [40]. We emphasize that the quantum problem in (4.5) does not contain any *ad hoc* counter term. Only the constant κ_λ is taken to $-\infty$ as in (4.16) in order to properly renormalize the interaction. The integral

$$\frac{\lambda}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{dk}{e^{\lambda(|k|^2+\kappa)}-1} = \lambda(e^{\lambda(-\Delta+\kappa)}-1)^{-1}(0,0)$$

appearing in (4.16) is a kind of bosonic regularization of the Green function. Note that $\lambda(e^{\lambda(-\Delta+\kappa)}-1)^{-1} \to (-\Delta+\kappa)^{-1}$ in the sense of operators, so that we formally obtain the desired infinite shift α in (4.12) in the limit. We also remark that (4.16) is universal. It only depends on $\int_{\mathbb{R}^d} w$ and is otherwise completely independent of V and of the specific form of w. The same result holds if the $o(1)$ are dropped on the right side of (4.16).

Theorem 7 was simultaneously proved in [24], but with an approach completely different from [41]. Our proof of Theorem 7 is *variational*, like for Theorems 1 and 2. We use that the Gibbs quantum state and the measure μ are the unique minimizers of the *Gibbs variational principle*, and our goal is to prove the convergence of the quantum problem to the classical one. The link is via the quantum de Finetti Theorem 3. Passing to the limit is very delicate and requires a fine understanding of the way that singularities appear in the measure μ_{r} when $\lambda \to 0^+$. To this end, we proved new *quantum correlation inequalities* to control the localization to low momenta and reduce the problem to finite dimensions. But it would take us too far to describe this in detail here and we refer the reader to [41].

Conclusion. Bose–Einstein condensates offer a source of interesting and difficult mathematical problems. The quantum de Finetti theorem provides both a new physical interpretation of condensation and a practical mathematical tool to prove it. It also naturally led us to consider nonlinear Gibbs measures, which appear at the phase transition and describe how the condensate forms. These measures play an important role in many different mathematical and physical situations.

APPENDIX: AN ELEMENTARY PROOF OF THEOREM 1

Let us start with the case where w is continuous and has a positive Fourier transform $\hat{w} \geqslant 0$. The argument is based on the following two lemmas.

Lemma 7.1 (Hoffmann–Ostenhof inequality [32]). *For every symmetric* $\Psi_N \in H^1((\mathbb{R}^d)^N, \mathbb{C})$

$$\int_{\mathbb{R}^{dN}} |\nabla \Psi_N|^2 \geqslant N \int_{\mathbb{R}^d} |\nabla \sqrt{\rho_{\Psi_N}}|^2 \tag{4.18}$$

with the one-particle density $\rho_{\Psi_N}(x) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} |\Psi_N(x, x_2, \dots, x_N)|^2 \, \mathrm{d}x_2 \cdots \mathrm{d}x_N$.

Proof. Compute $\nabla \sqrt{\rho_{\Psi_N}}$ and use the Cauchy–Schwarz inequality. ■

Lemma 7.2 (Onsager inequality [52]). *If* $\hat{w} \geqslant 0$ *is in* $L^1(\mathbb{R}^d)$*, then, for all* $\eta \in L^1(\mathbb{R}^d)$*,*

$$\sum_{1 \leqslant j < k \leqslant N} w(x_j - x_k) \geqslant \sum_{j=1}^N \eta * w(x_j) - \frac{1}{2} \iint_{\mathbb{R}^{2d}} w(x-y)\eta(x)\eta(y) \, \mathrm{d}x \, \mathrm{d}y - \frac{N}{2} w(0). \tag{4.19}$$

Proof. Expand $\iint_{\mathbb{R}^{2d}} w(x-y) f(x) f(y) \, \mathrm{d}x \, \mathrm{d}y = (2\pi)^{d/2} \int_{\mathbb{R}^d} \hat{w}(k) |\hat{f}(k)|^2 \, \mathrm{d}k \geqslant 0$ with $f = \sum_{j=1}^N \delta_{x_j} - \eta$. ■

With this we can prove (3.7). The potential energy can be expressed as

$$\sum_{j=1}^N \int_{\mathbb{R}^{dN}} V(x_j) |\Psi_N|^2 = N \int_{\mathbb{R}^d} V(x) \rho_{\Psi_N}(x) \, \mathrm{d}x.$$

Taking $\eta = N \rho_{\Psi_N}$ in (4.19) and using (4.18) provides the lower bound

$$\mathcal{E}(\Psi_N) \geqslant N \mathcal{E}_{\mathrm{GP}}(\sqrt{\rho_{\Psi_N}}) - \frac{N w(0)}{2(N-1)} \geqslant N e_{\mathrm{GP}} - \frac{N w(0)}{2(N-1)}. \tag{4.20}$$

Minimizing over Ψ_N and recalling the upper bound (3.6), we obtain

$$e_{\mathrm{GP}} - \frac{w(0)}{2(N-1)} \leqslant \frac{E(N)}{N} \leqslant e_{\mathrm{GP}}, \quad \text{for } \hat{w} \geqslant 0, \tag{4.21}$$

which clearly concludes the proof of Theorem 1, provided that $0 \leqslant \hat{w} \in L^1(\mathbb{R}^d)$. If $\hat{w}$ is nonnegative but not integrable, the proof is done by approximation.

We next turn to the case of an arbitrary w. The idea, inspired by [34, 49], is to use auxiliary classical particles repelling each other, in order to model the attractive part of the interaction. For simplicity, we consider $2N$ particles which we split in two groups of N. The positions of the N first will be denoted by $x_1, \dots, x_N$ whereas those of the others will be denoted by $y_1 = x_{N+1}, \dots, y_N = x_{2N}$. Of course, the separation is completely artificial and in reality the $2N$ particles are indistinguishable. We pick a $2N$-particle state Ψ_{2N} and use its bosonic symmetry in the $2N$ variables to rewrite

$$\frac{1}{2N} \int_{\mathbb{R}^{2dN}} |\nabla \Psi_{2N}|^2 = \frac{1}{N} \left\langle \Psi_{2N}, \sum_{j=1}^N (-\Delta)_{x_j} \Psi_{2N} \right\rangle.$$

In a similar fashion, we decompose $w = w_1 - w_2$ where $\widehat{w_1} = (\hat{w})_+ \geqslant 0$ and $\widehat{w_2} = (\hat{w})_- \geqslant 0$ and write the repulsive part using only the x_j's as

$$\frac{1}{2N(2N-1)}\left\langle \Psi_{2N}, \sum_{1\leqslant j<k\leqslant 2N} w_1(x_j - x_k)\Psi_{2N}\right\rangle = \frac{1}{N(N-1)}\left\langle \Psi_{2N}, \sum_{1\leqslant j<k\leqslant N} w_1(x_j - x_k)\Psi_{2N}\right\rangle.$$

On the other hand, we express the attractive part as the difference of two terms, involving respectively only the y_ℓ's and both species:

$$\begin{aligned}-\frac{1}{2N(2N-1)}&\left\langle \Psi_{2N}, \sum_{1\leqslant j<k\leqslant 2N} w_2(x_j - x_k)\Psi_{2N}\right\rangle \\ &= \frac{1}{N(N-1)}\left\langle \Psi_{2N}, \sum_{1\leqslant \ell<m\leqslant N} w_2(y_\ell - y_m)\Psi_{2N}\right\rangle \\ &\quad - \frac{1}{N^2}\left\langle \Psi_{2N}, \sum_{j=1}^{N}\sum_{\ell=1}^{N} w_2(x_j - y_\ell)\Psi_{2N}\right\rangle.\end{aligned}$$

This means that $\langle \Psi_{2N}, H_{2N}\Psi_{2N}\rangle/2N = \langle \Psi_{2N}, \tilde{H}_N\Psi_{2N}\rangle/N$ with

$$\begin{aligned}\tilde{H}_N = \sum_{j=1}^{N}(-\Delta)_{x_j} + V(x_j) + \frac{1}{N-1}\sum_{1\leqslant j<k\leqslant 2N} w_1(x_j - x_k) \\ + \frac{1}{N-1}\sum_{1\leqslant \ell<m\leqslant N} w_2(y_\ell - y_m) - \frac{1}{N}\sum_{j=1}^{N}\sum_{\ell=1}^{N} w_2(x_j - y_\ell).\end{aligned}$$

This Hamiltonian describes a system of N quantum particles repelling through the potential $w_1/(N-1)$ and N classical particles repelling through $w_2/(N-1)$, with an attraction $-w_2/N$ between the two species. In order to bound $\tilde{H}_N$ from below, we first fix the positions $y_1, \dots, y_N$ of the particles in the second group and consider $\tilde{H}_N$ as an operator acting only over the x_j's. Let Φ_N be any bosonic N-particle state in the N first variables. Using (4.18) and (4.19) for the repulsive potential w_1 as in the previous proof, we obtain

$$\begin{aligned}\frac{\langle \Phi_N, \tilde{H}_N\Phi_N\rangle}{N} \geqslant \int_{\mathbb{R}^d} |\nabla\sqrt{\rho_{\Phi_N}}|^2 + V\rho_{\Psi_N} + \frac{1}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \rho_{\Phi_N}(x)\rho_{\Phi_N}(y)w_1(x-y)\,\mathrm{d}x\,\mathrm{d}y \\ - \frac{w_1(0)}{2(N-1)} + \frac{1}{N(N-1)}\sum_{1\leqslant \ell<m\leqslant N} w_2(y_\ell - y_m) \\ - \frac{1}{N}\sum_{\ell=1}^{N} \rho_{\Phi_N} * w_2(y_\ell).\end{aligned}$$

Next we use again (4.19) for w_2 with $\eta = (N-1)\rho_{\Phi_N}$ and obtain

$$\begin{aligned}\sum_{1\leqslant \ell<m\leqslant N} w_2(y_\ell - y_m) - (N-1)\sum_{\ell=1}^{N} \rho_{\Phi_N} * w_2(y_\ell) \\ \geqslant -\frac{(N-1)^2}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \rho_{\Phi_N}(x)\rho_{\Phi_N}(y)w_2(x-y)\,dx\,dy - \frac{Nw_2(0)}{2}.\end{aligned}$$

Therefore, we have shown that

$$\frac{\langle \Phi_N, \tilde{H}_N \Phi_N \rangle}{N} \geqslant \mathcal{E}_{\rm GP}(\sqrt{\rho_{\Phi_N}}) - \frac{w_1(0)+w_2(0)}{2(N-1)} \geqslant e_{\rm GP} - \frac{w_1(0)+w_2(0)}{2(N-1)}.$$

Since the right-hand side is independent of the y_ℓ's, we have proved the operator bound

$$\frac{\tilde{H}_N}{N} \geqslant e_{\rm GP} - \frac{w_1(0)+w_2(0)}{2(N-1)}.$$

Minimizing over Ψ_{2N} gives

$$e_{\rm GP} - \frac{w_1(0)+w_2(0)}{2(N-1)} \leqslant \frac{E(2N)}{2N} \leqslant e_{\rm GP}.$$

Note that $w_1(0)+w_2(0) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} |\hat{w}|$. We have considered an even number of particles for simplicity, but the proof works the same if we use two groups of N and $N+1$ particles. Another possibility is to use that $N \mapsto E(N)/N$ is nondecreasing, which gives, for $N \geqslant 4$,

$$\boxed{e_{\rm GP} - \frac{(2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} |\hat{w}|}{N-3} \leqslant \frac{E(N)}{N} \leqslant e_{\rm GP}.} \tag{4.22}$$

Nonintegrable $\hat{w}$ can be handled using an approximation argument. ■

Note that the two error bounds in (4.21) and (4.22) are of the optimal order $O(N^{-1})$, due to Theorem 4. If the GP minimizer exists and is unique, the convergence of the density matrices can be proved using a perturbation argument described in [57, CHAP. 2].

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement MDFT No 725528).

REFERENCES

[1] J. R. Abo-Shaeer, C. Raman, J. M. Vogels, and W. Ketterle, Observation of vortex lattices in Bose–Einstein condensates. *Science* **292** (2001), no. 5516, 476–479.

[2] A. Aftalion, *Vortices in Bose–Einstein condensates*. Progr. Nonlinear Differential Equations Appl. 67, Springer, 2006.

[3] Z. Ammari and F. Nier, Mean field limit for bosons and infinite dimensional phase-space analysis. *Ann. Henri Poincaré* **9** (2008), 1503–1574.

[4] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269** (1995), no. 5221, 198–201.

[5] X. Antoine and R. Duboscq, GPELab, a Matlab toolbox to solve Gross–Pitaevskii equations I: Computation of stationary solutions. *Comput. Phys. Commun.* **185** (2014), no. 11, 2969–2991.

[6] R. Benguria and E. H. Lieb, Proof of the stability of highly negative ions in the absence of the Pauli principle. *Phys. Rev. Lett.* **50** (1983), 1771–1774.

[7] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, Bogoliubov theory in the Gross–Pitaevskii limit. *Acta Math.* **222** (2019), no. 2, 219–335.

[8] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, The excitation spectrum of Bose gases interacting through singular potentials. *J. Eur. Math. Soc. (JEMS)* **22** (2020), no. 7, 2331–2403.

[9] N. N. Bogoliubov, About the theory of superfluidity. *Izv. Akad. Nauk SSSR* **11** (1947), 77.

[10] S. Bose, Plancks Gesetz und Lichtquantenhypothese. *Z. Phys.* **26** (1924), no. 1, 178–181.

[11] J. Bourgain, Periodic nonlinear Schrödinger equation and invariant measures. *Comm. Math. Phys.* **166** (1994), no. 1, 1–26.

[12] J. Bourgain, Invariant measures for the Gross–Pitaevskii equation. *J. Math. Pures Appl.* **76** (1997), no. 8, 649–02.

[13] K. B. Davis, M. O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, Bose–Einstein Condensation in a Gas of Sodium Atoms. *Phys. Rev. Lett.* **75** (1995), 3969–3973.

[14] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio. *Atti Accad. Naz. Lincei, Mem. Cl. Sci. Fis. Mat. Nat. (Ser. VI)* **IV** (1930), no. 5, 86–133.

[15] B. Delamotte, A hint of renormalization. *Am. J. Phys.* **72** (2004), 170.

[16] A. Deuchert and R. Seiringer, Gross–Pitaevskii limit of a homogeneous Bose gas at positive temperature. *Arch. Ration. Mech. Anal.* **236** (2020), no. 3, 1217–1271.

[17] A. Deuchert, R. Seiringer, and J. Yngvason, Bose–Einstein condensation in a dilute, trapped gas at positive temperature. *Comm. Math. Phys.* **368** (2019), no. 2, 723–776.

[18] F. J. Dyson, The S matrix in quantum electrodynamics. *Phys. Rev. (2)* **75** (1949), 1736–1755.

[19] A. Einstein, *Quantentheorie des einatomigen idealen Gases*. pp. 261–267, Sitzber. Kgl. Preuss. Akad. Wiss., Verlag der Akademie der Wissenschaften, 1924.

[20] M. Fannes, H. Spohn, and A. Verbeure, Equilibrium states for mean field models. *J. Math. Phys.* **21** (1980), no. 2, 355–358.

[21] S. Fournais and J. P. Solovej, The energy of dilute Bose gases. *Ann. of Math. (2)* **192** (2020), no. 3, 893–976.

[22] J. Fröhlich, A. Knowles, B. Schlein, and V. Sohinger, Gibbs measures of nonlinear Schrödinger equations as limits of many-body quantum states in dimensions $d \leqslant 3$. *Comm. Math. Phys.* **356** (2017), no. 3, 883–980.

[23] J. Fröhlich, A. Knowles, B. Schlein, and V. Sohinger, A microscopic derivation of time-dependent correlation functions of the 1D cubic nonlinear Schrödinger equation. *Adv. Math.* **353** (2019), 67–115.

[24] J. Fröhlich, A. Knowles, B. Schlein, and V. Sohinger, The mean-field limit of quantum Bose gases at positive temperature. *J. Amer. Math. Soc.* (2021), online first, arXiv:2001.01546.

[25] A. Giuliani and R. Seiringer, The ground state energy of the weakly interacting Bose gas at high density. *J. Stat. Phys.* **135** (2009), no. 5–6, 915–934.

[26] J. Glimm and A. Jaffe, *Quantum physics: A functional integral point of view*. Springer, 1987.

[27] P. Grech and R. Seiringer, The excitation spectrum for weakly interacting bosons in a trap. *Comm. Math. Phys.* **322** (2013), no. 2, 559–591.

[28] E. Gross, Classical theory of boson wave fields. *Ann. Phys.* **4** (1958), no. 1, 57–74.

[29] C. Hainzl, M. Lewin, and J. P. Solovej, The mean-field approximation in quantum electrodynamics: the no-photon case. *Comm. Pure Appl. Math.* **60** (2007), no. 4, 546–596.

[30] E. Hewitt and L. J. Savage, Symmetric measures on Cartesian products. *Trans. Amer. Math. Soc.* **80** (1955), 470–501.

[31] D. Hilbert, Mathematical problems. *Bull. Amer. Math. Soc.* **8** (1902), 437–479.

[32] M. Hoffmann-Ostenhof and T. Hoffmann-Ostenhof, Schrödinger inequalities and asymptotic behavior of the electron density of atoms and molecules. *Phys. Rev. A* **16** (1977), no. 5, 1782–1785.

[33] R. L. Hudson and G. R. Moody, Locally normal symmetric states and an analogue of de Finetti's theorem. *Z. Wahrsch. Verw. Gebiete* **33** (1975/76), no. 4, 343–351.

[34] J.-M. Lévy-Leblond, Nonsaturation of gravitational forces. *J. Math. Phys.* **10** (1969), 806–812.

[35] M. Lewin, Geometric methods for nonlinear many-body quantum systems. *J. Funct. Anal.* **260** (2011), 3535–3595.

[36] M. Lewin, Mean-field limit of Bose systems: rigorous results. In *Proceedings of the International Congress of Mathematical Physics, Santiago de Chile*. 2015, arXiv:1510.04407.

[37] M. Lewin, P. T. Nam, and N. Rougerie, Derivation of Hartree's theory for generic mean-field Bose systems. *Adv. Math.* **254** (2014), 570–621.

[38] M. Lewin, P. T. Nam, and N. Rougerie, Derivation of nonlinear Gibbs measures from many-body quantum mechanics. *J. Éc. Polytech. Math.* **2** (2015), 65–115.

[39] M. Lewin, P. T. Nam, and N. Rougerie, Gibbs measures based on 1D (an)harmonic oscillators as mean-field limits. *J. Math. Phys.* **59** (2018), 041901.

[40] M. Lewin, P. T. Nam, and N. Rougerie, Derivation of renormalized Gibbs measures from equilibrium many-body quantum Bose gases. *J. Math. Phys.* **60** (2019), no. 6, 061901.

[41] M. Lewin, P. T. Nam, and N. Rougerie, Classical field theory limit of many-body quantum Gibbs states in 2D and 3D. *Invent. Math.* **224** (2021), no. 2, 315–444.

[42] M. Lewin, P. T. Nam, S. Serfaty, and J. P. Solovej, Bogoliubov spectrum of interacting Bose gases. *Comm. Pure Appl. Math.* **68** (2015), no. 3, 413–471.

[43] E. H. Lieb and R. Seiringer, Proof of Bose–Einstein condensation for dilute trapped gases. *Phys. Rev. Lett.* **88** (2002), no. 17, 170409.

[44] E. H. Lieb and R. Seiringer, Derivation of the Gross–Pitaevskii equation for rotating Bose gases. *Comm. Math. Phys.* **264** (2006), no. 2, 505–537.

[45] E. H. Lieb, R. Seiringer, J. P. Solovej, and J. Yngvason, *The mathematics of the Bose gas and its condensation*. Oberwolfach Semin., Birkhäuser, 2005.

[46] E. H. Lieb, R. Seiringer, and J. Yngvason, Bosons in a trap: A rigorous derivation of the Gross–Pitaevskii energy functional. *Phys. Rev. A* **61** (2000), no. 4, 043602.

[47] E. H. Lieb and H. Siedentop, Renormalization of the regularized relativistic electron–positron field. *Comm. Math. Phys.* **213** (2000), no. 3, 673–683.

[48] E. H. Lieb and J. P. Solovej, Ground state energy of the one-component charged Bose gas. *Comm. Math. Phys.* **217** (2001), no. 1, 127–163.

[49] E. H. Lieb and H.-T. Yau, The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics. *Comm. Math. Phys.* **112** (1987), no. 1, 147–174.

[50] E. H. Lieb and J. Yngvason, Ground state energy of the low density Bose gas. *Phys. Rev. Lett.* **80** (1998), no. 12, 2504–2507.

[51] P. T. Nam and A. Triay, Bogoliubov excitation spectrum of trapped Bose gases in the Gross–Pitaevskii regime. 2021, arXiv:2106.11949.

[52] L. Onsager, Electrostatic interaction of molecules. *J. Phys. Chem.* **43** (1939), no. 2, 189–196.

[53] O. Penrose and L. Onsager, Bose–Einstein condensation and liquid helium. *Phys. Rev.* **104** (1956), 576–584.

[54] L. P. Pitaevskii, Vortex lines in an imperfect Bose gas. *Ž. èksp. Teor. Fiz.* **40** (1961), no. 40, 646–651.

[55] G. A. Raggio and R. F. Werner, Quantum statistical mechanics of general mean field systems. *Helv. Phys. Acta* **62** (1989), no. 8, 980–1003.

[56] C. Reid, *Hilbert*. IX, Springer, Berlin–Heidelberg–New York, 1970.

[57] N. Rougerie, Scaling limits of bosonic ground states, from many-body to nonlinear Schrödinger. *EMS Surv. Math. Sci.* **7** (2021), no. 2, 253–408.

[58] A. Schirrmacher, *Establishing quantum physics in Göttingen. David Hilbert, Max Born, and Peter Debye in context, 1900–1926*. Springer, Cham, 2019.

[59] R. Seiringer, The excitation spectrum for weakly interacting bosons. *Comm. Math. Phys.* **306** (2011), no. 2, 565–578.

[60] E. Størmer, Symmetric states of infinite tensor products of C^*-algebras. *J. Funct. Anal.* **3** (1969), 48–68.

[61] M. van den Berg, J. T. Lewis, and J. V. Pulè, The large deviation principle and some models of an interacting boson gas. *Comm. Math. Phys.* **118** (1988), no. 1, 61–85.

[62] J. von Neumann, *Mathematical foundations of quantum mechanics*. Princeton University Press, Princeton, NJ, 1932.

[63] E. P. Wigner, The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure Appl. Math.* **13** (1960), no. 1, 1–14.

MATHIEU LEWIN

CNRS & CEREMADE, Université Paris-Dauphine, PSL University, 75016 Paris, France, mathieu.lewin@math.cnrs.fr

GLOBAL DYNAMICS AROUND AND AWAY FROM SOLITONS

KENJI NAKANISHI

ABSTRACT

This article reviews some results, as well as open questions, on global behavior of general solutions for nonlinear dispersive equations, with an emphasis on transitions of solutions around solitons with respect to time evolution and initial perturbation.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 37K40; Secondary 35B30, 35B40, 35L71, 35Q55, 37K45

KEYWORDS

Nonlinear dispersive equations, global dynamics, solitons, scattering, blow-up, invariant manifolds

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3822–3840
DOI 10.4171/ICM2022/48

1. INTRODUCTION

Nonlinear dispersive equations describe space-time evolution of waves in various physical phenomena, which are governed mainly by dispersion and nonlinear interactions of waves. A representative example is the nonlinear Schrödinger equation (NLS)

$$i\dot{u} - \Delta u = \lambda |u|^{p-1}u, \quad u(t,x) : \mathbb{R}^{1+d} \to \mathbb{C}, \tag{1.1}$$

where $d \in \mathbb{N}$, $p > 1$, and $\lambda \in \mathbb{R}$ are constants. Depending on the balance or competition between the dispersion and interaction, which differs equation by equation, as well as the initial data, the solutions of each equation exhibit a wide range of behavior in space-time. The three major types of solutions are

- scattering solutions which are dominated by dispersion—spreading waves with decaying amplitude;
- blow-up solutions which are dominated by nonlinearity—focusing waves with diverging amplitude;
- solitons for which dispersion and nonlinearity are in balance to keep a fixed shape of the wave.

Most of the equations are in the Hamiltonian form. For example, NLS may be written as

$$\dot{u} = iE_S'(u), \quad E_S(u) := \int_{\mathbb{R}^d} \frac{|\nabla u|^2}{2} - \frac{\lambda |u|^{p+1}}{p+1} dx, \tag{1.2}$$

where $E_S'(u)$ denotes the Fréchet derivative. The Hamiltonian or energy E_S is well defined on $H^1(\mathbb{R}^d)$ if the nonlinear part is controlled by the Sobolev inequality, namely $d \le 2$ or $p + 1 \le \frac{2d}{d-2} =: 2^\star$. Then it is natural to consider solutions in the energy space $H^1(\mathbb{R}^d)$, where the energy $E_S(u)$ is conserved.

Nonlinear dispersive equations have been intensively studied since the late 20th century, so that we have by now a fair amount of knowledge on the fundamental questions from the PDE viewpoint, such as the unique existence of local solutions with wide range of regularity, of solutions with typical behavior, as well as their qualitative and quantitative properties, including asymptotic profiles.

In this century, there has been more progress in the study on *large solutions for long time*, in which the dispersion and nonlinearity have stronger and more complicated interplay, generating more diverse solutions. It is, however, in most cases too difficult to look at all general solutions and their long-time behavior, as there are so many possibilities while our method of analysis is still quite limited. Then the solitons are the natural first target to attack among all the solutions, as they are expected to indicate the balance or the threshold of dominance between the dispersion and nonlinearity. The *soliton resolution conjecture* has been the major slogan to promote this direction of study, which roughly asserts that: *Generic global solutions are asymptotic to a superposition of solitons getting away from each other and a dispersive decaying wave as $t \to \infty$.* In the case of NLS (for appropriate p), the asymptotic formula should take the form

$$u(t) - \sum_{n=1}^{N} e^{i\theta_n(t)} \varphi_n\big(x - c_n(t)\big) - v(t) \to 0, \tag{1.3}$$

in the energy space $H^1(\mathbb{R}^d)$ as $t \to \infty$, for some soliton profiles $\varphi_n \in H^1(\mathbb{R}^d)$ with some $\theta_n : \mathbb{R} \to \mathbb{R}$ and $c_n : \mathbb{R} \to \mathbb{R}^d$ satisfying $|c_m(t) - c_n(t)| \to \infty$ for $m \neq n$, and some dispersive component $v(t,x)$ solving the free equation $i\dot{v} = \Delta v$.

On the one hand, the conjecture is a natural extension from the case of completely integrable equations (e.g., $d = 1$ and $p = 3$ for NLS), where solitons are very stable and rigid: they are unchanged both by initial perturbation and collisions, up to a change of the parameters. The genericity condition in the conjecture is to eliminate some exceptional solutions, such as breathers, which appear already in the integrable case.

On the other hand, most of the nonlinear dispersive equations are not integrable, where most of solitons are unstable with respect to initial perturbations. Although this instability makes it more difficult to capture and maintain the solitons in reality and numerics, it does not diminish the importance of solitons in the study of global dynamics, especially regarding the role of a threshold. In fact, in the space of solutions or initial data, typically in the energy space, instability means that the soliton is a limit point of other types of solutions, while stability means that there is no nearby solution with much different behavior. Hence unstable solitons are naturally expected to play more distinct roles in classifying the other solutions. Even if the solitons are unstable, the threshold between different types of solutions should be clearly observed both in numerics and experiments, as one looks at a collection of solutions rather than the individual ones. Such structures among solutions may well be stable and robust with respect to perturbations of the equation, even if the behavior of each solution is changed.

Therefore, in studying the global dynamics, it is not sufficient to know merely that a soliton is unstable, we should investigate in which directions the instability appears, and in what types of behavior of solutions. In other words, we should look at *all solutions in a neighborhood of solitons*. Note that stability is an answer to this question, but instability (negation of stability) may not be a complete answer by itself. Determining the stability is, of course, the most important starting point, which has a vast amount of literature, but there has been more recent progress in getting to the next stage.

Instability means that some solutions starting nearby a soliton become eventually very different or far from the soliton in the solution space. While those solutions are still near the soliton, their behavior may be well approximated by the linearized operator, which is well described in terms of its spectrum. However, after the solutions go far away from the soliton, which is often the case, then the linearized operator tells little about their behavior. To see the essential features of those solutions, and thus the threshold nature of the unstable soliton, it is necessary to look at *those solutions after they get far from the soliton.* The recent research is getting also into this stage of study.

Also in practice, the solutions at $t = \infty$ do not have so much meaning, but the asymptotic descriptions as $t \to \infty$ should be regarded as an approximation for what happens in finite time. However, if the solitons are unstable, the asymptotic decomposition into them is useless by itself for a finite-time approximation, since unstable solitons may keep disappearing and appearing along the evolution. Hence we should look at the behavior of solutions *not only as $t \to \infty$, but also for all intermediate $t \in \mathbb{R}$*. The oscillatory scenario is an obstruction also in studying the asymptotic behavior, but the investigation for all time is even more demanding. Nevertheless, the recent research is getting also into this stage.

In short, to investigate the global dynamics of nonlinear dispersive equations, it is desired to describe the solutions *for all time and for all initial data* in a neighborhood of unstable solitons. The main interest is on *transitions of behavior both in time evolution and for initial perturbations*. The purpose of this article is to review a few results in this direction, as well as open questions.

2. GROUND STATES AS THE DYNAMICAL THRESHOLD

Among all the solitons, the most important ones are those with the least energy, namely the ground states, as its energy is the necessary amount to produce the balance between the dispersion and nonlinearity. This article is mostly focused on the ground states and their variants, even though some of them will be called excited states. For a concrete explanation, we take the nonlinear Klein–Gordon equation (NLKG)

$$\ddot{u} - \Delta u + mu = |u|^{p-1}u, \quad u(t,x) : \mathbb{R}^{1+d} \to \mathbb{R}, \tag{2.1}$$

where $d \in \mathbb{N}$, $p > 1$, and $m \geq 0$ are constants. It is the Hamiltonian flow with the energy

$$E_K\big(\vec{u}(t)\big) := \int_{\mathbb{R}^d} \frac{|\dot{u}|^2 + |\nabla u|^2 + m|u|^2}{2} - \frac{|u|^{p+1}}{p+1} dx, \tag{2.2}$$

similar to NLS in (1.2); in the energy space

$$\vec{u}(t) := \big(u(t,x), \dot{u}(t,x)\big) \in \mathcal{H} := H^1(\mathbb{R}^d) \times L^2(\mathbb{R}^d). \tag{2.3}$$

In the case $m = 0$ of the nonlinear wave equation (NLW), H^1 should be replaced with the homogeneous Sobolev space $\dot{H}^1$. The ground state $Q \in H^2(\mathbb{R}^d)$ is a nontrivial stationary solution of

$$-\Delta Q + mQ = |Q|^{p-1}Q \tag{2.4}$$

with the least energy. Its study has a long history for the stationary equation and the evolution equations, including the NLS case and the heat equation. By the existence result of Strauss [56] and the uniqueness result of Kwong [37], the entire set of the ground states is $\{\pm Q(x-c)\}_{c\in\mathbb{R}^d}$ for a unique radial positive function $Q(x) = Q(|x|) > 0$. In the massless case (NLW), the Pohozaev identity [54] implies that the ground state exists if and only if $d \geq 3$ and the nonlinear power is $p + 1 = \frac{2d}{d-2} =: 2^\star$, namely the energy-critical exponent, and the ground state Q is the explicit Aubin–Talenti function [2,57], maximizing the Sobolev inequality for $\dot{H}^1(\mathbb{R}^d) \subset L^{2^\star}(\mathbb{R}^d)$.

2.1. Below the ground states

The instability of Q follows from its min–max property:

$$E_K(\vec{Q}) = \min_{\varphi \in H^1(\mathbb{R}^d)\setminus\{0\}} \max_{\lambda>0} E_K(\lambda\vec{\varphi})$$
$$= \min\bigl\{E_K(\vec{\varphi}) \mid \varphi \in H^1(\mathbb{R}^d) \setminus \{0\}, K(\varphi) = 0\bigr\}, \tag{2.5}$$

where $\vec{Q} := (Q, 0)$, and the Nehari functional [50] is defined by

$$K(\varphi) := \frac{d}{d\lambda}\Big|_{\lambda=1} E_K(\lambda\vec{\varphi}) = \int_{\mathbb{R}^d} |\nabla u|^2 + m|u|^2 - |u|^{p+1} dx. \tag{2.6}$$

A similar characterization is given by using the dilation $\varphi(\lambda x)$, leading to Derrick's theorem [11]. Another option is the L^2-invariant scaling $\lambda^{d/2}\varphi(\lambda x)$, which yields the virial functional (see [28] for their relations including the dynamics). Thus the energy space below the ground state is split into two open sets,

$$\mathcal{H}_< := \bigl\{\varphi \in \mathcal{H} \mid E_K(\varphi) < E_K(\vec{Q})\bigr\} = \mathcal{H}_<^+ \cup \mathcal{H}_<^-,$$
$$\mathcal{H}_<^+ := \bigl\{\varphi \in \mathcal{H}_< \mid K(\varphi_1) \geq 0\bigr\}, \quad \mathcal{H}_<^- := \bigl\{\varphi \in \mathcal{H}_< \mid K(\varphi_1) < 0\bigr\}. \tag{2.7}$$

It is easy to see that $\mathcal{H}_<^+$ is bounded and $\mathcal{H}_<^-$ is unbounded. Since $E_K(\vec{u})$ is conserved and $\mathcal{H}_<^\pm$ are separated from each other, both regions $\mathcal{H}_<^\pm$ are invariant with respect to the NLKG flow. Then all the solutions in $\mathcal{H}_<^+$ are global in time as soon as the Cauchy problem is locally well posed in $\mathcal{H}$ with a uniform lower bound on the existence time (which is the case for $p + 1 < 2^\star$ by Ginibre–Velo [25]).

Payne–Sattinger [52] proved (in the bounded domain case) that all solutions in $\mathcal{H}_<^-$ blow up in finite time for NLKG, as well as for the heat equation. Thus all the solutions with the energy below the ground state are split into the cases of global existence and blow-up, as two disjoint open sets in $\mathcal{H}$, which are distinguished by the initial data explicitly by sign $K(u(0))$. The openness means that both properties are stable, and the ground states are the joint boundary of the two regions

$$\bigl\{\pm Q(x-c)\bigr\}_{c\in\mathbb{R}^d} = \overline{\mathcal{H}_<^+} \cap \overline{\mathcal{H}_<^-}. \tag{2.8}$$

More recently, Kenig–Merle [31, 32] proved this type of dichotomy in the energy-critical case $p + 1 = 2^\star$ ($d \geq 3$) for NLW, as well as for NLS (in the radial case), proving moreover that all the solutions in $\mathcal{H}_<^+$ are scattering as $t \to \pm\infty$, namely

$$\lim_{t\to\pm\infty} \bigl\|\vec{u}(t) - \vec{v}^{\pm}(t)\bigr\|_{\mathcal{H}} = 0 \tag{2.9}$$

for some $v^\pm$ solving the free equation $\ddot{v} - \Delta v = 0$. Their method has been applied to many other equations, including NLKG [28, 33] and NLS [12, 13, 21, 27, 34, 35] with the energy-(sub)critical and mass-(super)critical power, namely for $2 + \frac{4}{d} \leq p + 1 \leq 2^\star$.

2.2. Above the threshold

Although the dichotomy into scattering and blow-up is a very simple, explicit, and complete classification of the global dynamics, there seems to be no intrinsic reason in the equations to restrict the solutions below the ground states $E_K(\vec{u}) < E_K(\vec{Q})$, as those ground states are not local extrema, but rather saddle points for the energy. It also seems impossible to impose such a strict condition (inequality) both in numerical experiments and in physical ones. It is therefore more natural to impose a condition of the form

$$E_K(\vec{u}) < E_K(\vec{Q}) + \varepsilon \tag{2.10}$$

for some small $\varepsilon > 0$, which includes in particular a full neighborhood of the ground states.

As soon as the energy is above the ground state, however, the topological separation is lost between $K(u) > 0$ and $K(u) < 0$, or between the scattering and blow-up, which enables a transition between different types of behavior. One may expect that this transition could make the global dynamics very complicated and even chaotic, as it can possibly happen for many times. Nakanishi–Schlag [47, 49] showed that it is not the case, and the complication remains minimal for small $\varepsilon > 0$, at least for NLKG with $d = p = 3$, which has been extended to NLS and NLW in [36, 48]. That is because the transition is allowed only for one time for each solution from the scattering region to the blow-up region (or vice versa), taking place only in a small neighborhood of the ground states, and described well by the linearized equation around the ground states. The behavior of solutions away from the ground states is essentially the same as in the case below the ground state, in the sense that both the scattering and blow-up are characterized by monotonicity of the virial identity. Thus all the solutions with $E_K(\vec{u}) < E_K(\vec{Q}) + \varepsilon$ are classified into $9 = 3 \times 3$ sets of global behavior, depending whether it is scattering, blowing-up, or asymptotic to the ground states in $t > 0$ and $t < 0$. In the simple case of NLKG with $p = d = 3$ under the radial symmetry, the classification reads as follows. For any $\varphi \in \mathcal{H}$ and $X \subset \mathcal{H}$, let $\varphi^\dagger := (\varphi_1, -\varphi_2)$ and $X^\dagger := \{\varphi^\dagger \mid \varphi \in X\}$ denote the time inversion.

Theorem 2.1 ([47]). *Let $p = d = 3$, $m > 0$, and*

$$\mathcal{H}_{\varepsilon,r} := \{\varphi(x) = \varphi(|x|) \in \mathcal{H} \mid E_K(\varphi) < E_K(\vec{Q}) + \varepsilon\} \tag{2.11}$$

for $\varepsilon > 0$. If $\varepsilon > 0$ is small enough, then there is a C^1-manifold $\mathcal{M} \subset \mathcal{H}_{\varepsilon,r}$ of codimension 1 with the following properties: $\mathcal{H}_{\varepsilon,r} \setminus (\mathcal{M} \cup -\mathcal{M})$ is a union of two domains $\mathcal{S}$ and $\mathcal{B}$. Let u be any solution of (2.1) with $\vec{u}(0) \in \mathcal{H}_{\varepsilon,r}$. If $\vec{u}(0) \in \mathcal{S}$, then u is scattering as $t \to \infty$. If $\vec{u}(0) \in \mathcal{B}$, then u blows up in finite time for $t > 0$. If $\vec{u}(0) \in \mathcal{M}$, then $u - Q$ is scattering as $t \to \infty$. Moreover, $\mathcal{M}$ and $\mathcal{M}^\dagger$ intersect transversely, while $\mathcal{M}^\dagger \cap (-\mathcal{M}) = \varnothing$.

The transversal intersection of $\mathcal{M} \cap \mathcal{M}^\dagger$ implies that all the $9 = 3 \times 3$ combinations of behavior in $t > 0$ and $t < 0$ are nonempty. The above result clarifies the important role of the center-stable manifold $\mathcal{M}$ and the center-unstable manifold $\mathcal{M}^\dagger$ of the ground states $\pm Q$, which had been constructed by Bates–Jones [3], while the scattering on the manifolds to the ground states had been established by Schlag [55] and Beceanu [4] for NLS. The key ingredient for the above classification is the fact that the transition cannot happen more than once, which is called the one-pass theorem. It may be regarded as a small perturbation from the threshold dynamics on $E_K(\vec{u}) = E_K(\vec{Q})$, in particular the nonexistence of a homoclinic orbit for the ground states, which had been established by Duyckaerts–Merle [19] for the energy-critical NLS, and extended to other cases [18, 20]. More precisely, the one-pass theorem prohibits solutions from reentry into a small neighborhood of the ground states after escaping from there. If we distinguish between the positive ground states Q and the negative $-Q$, then the number of classification sets is $14 = 4 \times 4 - 2$, as follows:

$$\begin{gathered} \mathcal{S}^\dagger \cap \mathcal{S}, \quad \mathcal{B}^\dagger \cap \mathcal{B}, \quad \pm(\mathcal{M}^\dagger \cap \mathcal{M}), \quad (\pm\mathcal{M}^\dagger) \cap \mathcal{S}, \quad (\pm\mathcal{M}^\dagger) \cap \mathcal{B}, \\ \mathcal{S}^\dagger \cap (\pm\mathcal{M}), \quad \mathcal{B}^\dagger \cap (\pm\mathcal{M}), \quad \mathcal{S}^\dagger \cap \mathcal{B}, \quad \mathcal{B}^\dagger \cap \mathcal{S}, \end{gathered} \tag{2.12}$$

The subtraction of -2 from 4×4 is due to the absence of $(\pm\mathcal{M}^\dagger) \cap (\mp\mathcal{M})$, namely connecting orbits between Q and $-Q$, which is also precluded by the one-pass theorem.

2.3. Higher energy

The next question is what if the energy is much bigger than the ground state, namely $E_K(\vec{u}) > E_K(\vec{Q}) + \varepsilon$. Actually, the more general statement of the above result in the nonradial case [49], taking account of the Lorentz invariance and conserved momentum $P(\vec{u}) := \int_{\mathbb{R}^d} \dot{u}\nabla u dx$, is in a bigger region

$$\left[E_K(\vec{u})^2 - \left|P(\vec{u})\right|^2\right]^{1/2} < E_K(\vec{Q}) + \varepsilon \quad \left([z]^\alpha := |z|^{\alpha-1}z\right), \tag{2.13}$$

which includes the ground state solitons of any traveling speed (slower than the light), but the classification is essentially the same as above. The main interest is how and where the dynamics change essentially.

There are at least two obvious candidates for the next energy level. One is the other stationary solutions, namely the excited state, and the other is multisolitons. Note that the excited states have at least twice the ground state energy $E_K(\vec{u}) > 2E_K(\vec{Q})$, because they have to be sign-changing due to the uniqueness of positive solutions by Gigas–Ni–Nirenberg [24] and Kwong [37], then both the positive and negative parts must have more energy than $E_K(\vec{Q})$ due to the characterization (2.5). On the other hand, if a solution u is asymptotic to a sum of $N \in \mathbb{N}$ ground states moving away from each other, as in the soliton resolution conjecture, then $E_K(\vec{u}) \ge NE_K(\vec{Q})$, where the equality $E_K(\vec{u}) = NE_K(\vec{Q})$ happens only if the asymptotic speeds of the solitons are all zero. Asymptotic multisolitons were constructed by Martel–Merle [39] for NLS with positive speeds in the stable case, which has been extended to the unstable case [6], as well as to NLKG [9], and with zero speed for NLS [51] and NLKG [1]. Therefore, in view of the soliton resolution conjecture, it is natural to expect that the classification in Theorem 2.1 should extend up to $E_K(\vec{u}) < 2E_K(\vec{Q})$, at least concerning the asymptotic behavior.

If one looks at the full-time dynamics, however, there is another candidate for an essential change of the dynamics. It is a heteroclinic orbit connecting the two distinct ground states Q and $-Q$ in a weak sense: more precisely, a solution u satisfying

$$\lim_{t\to\pm\infty} \left\|\vec{u}(t) \pm \vec{Q} - \vec{v}_\pm(t)\right\|_{\mathcal{H}} = 0 \tag{2.14}$$

for some free solutions $v_\pm$, which may be called *heteroscattering*. The one-pass theorem precludes such solutions for $E_K(\vec{u}) < E_K(\vec{Q}) + \varepsilon$, which is $\mathcal{M}^\dagger \cap (-\mathcal{M}) = \varnothing$, but it is not difficult to construct such a solution with the energy close to $2E_K(\vec{Q})$, by superposing in space-time two heteroscattering solutions, from Q to 0, and from 0 to $-Q$, respectively.

A simple numerical experiment indicates that such solutions appear at a much lower energy level than $2E_K(\vec{Q})$. Then it seems natural to conjecture that there is a threshold energy $E^* \in (E_K(\vec{Q}), 2E_K(\vec{Q}))$ such that for $E_K(\vec{u}) < E^*$ the 14-set classification (2.12) is valid, while for $E_K(\vec{u}) > E^*$ there are solutions satisfying (2.14), increasing the number of solution sets to $4 \times 4 = 16$ at least. Related questions are if there is heteroscattering between Q and $-Q$ with the minimal energy E^*, and what the complete classification of dynamics is for $E_K(\vec{u}) < E^* + \varepsilon$, or for $E_K(\vec{u}) < 2E_K(\vec{Q})$.

A remarkably successful method to go to higher energy is the channel-of-energy by Duyckaerts–Kenig–Merle [16], which settled the soliton resolution conjecture for the energy-critical NLW in the radial 3D case, including the blow-up solutions with a bounded energy norm. It has recently been extended to the higher odd dimensions [17], as well as to the 4D case [15] and to the wave maps under rotational symmetry. Without the rotational symmetry, there are also similar results [14] along time sequences. It seems, however, that this method depends heavily on the special property of the wave equation, that is, the single speed of propagation, while dispersive equations in general have wide ranges of group velocity. It is a challenging and important problem to extend the method or find a similar one for the more dispersive equations such as NLKG and NLS.

3. TRANSITION BETWEEN SOLITONS

Since solitons are the key junctions of global dynamics for nonlinear dispersive equations, it is an important problem to understand the behavior of the solutions migrating from a neighborhood of one soliton to another. In fact, when the equation has both stable and unstable solitons, it is generally expected that solutions starting near the unstable ones will get away from them and eventually approach some of the stable ones. However, the conservation laws prohibit the solutions to approach the latter solitons in the energy norm, unless the two solitons happen to be close to each other in the conserved quantities. In general, the approach should be only in the weak or local topology, where the excessive energy is radiated away in a dispersive wave component.

This type of transition from one soliton to another should happen also between unstable solitons. Trying to include such a behavior into a classification as above seems still a bit too ambitious, as the complete classification for NLKG or NLS is yet much below all the excited state solitons. However, we can make a model problem by adding some spatial inhomogeneity, which is an easy way to create standing waves. Specifically, the nonlinear Schrödinger equation with a potential (NLSP)

$$i\dot{u} - \Delta u + Vu = |u|^2 u, \quad u(t,x) : \mathbb{R}^{1+3} \to \mathbb{R}, \tag{3.1}$$

is a good model to consider a classification of solutions including two different solitons, both stable and unstable ones. The standing waves for NLSP are solutions in the form $u(t,x) = e^{-it\omega}\varphi(x)$ for some $\omega \in \mathbb{R}$, for which the equation is reduced to

$$-\Delta u + \omega u + Vu = |u|^2 u. \tag{3.2}$$

More precisely, let $V : \mathbb{R}^3 \to \mathbb{R}$ be “nice” enough, e.g., a radial Schwartz function, such that the linear Schrödinger operator $-\Delta + V$ has only one eigenvalue, denoted by $e_0 < 0$. Let $\phi_0 \in H^2(\mathbb{R}^3)$ be the corresponding eigenfunction or the ground state of $-\Delta + V$, normalized in $L^2(\mathbb{R}^3)$. Let E_V be the Hamiltonian of NLSP, defined by

$$E_V(u) := E_S(u) + \int_{\mathbb{R}^3} \frac{V|u|^2}{2} dx. \tag{3.3}$$

Then one may construct two different types of standing waves for NLSP. One family is generated from the linear ground state ϕ_0 by bifurcation, which is small and stable with negative energy in the asymptotic form (see [26])

$$\begin{aligned} &\Phi[z] = z\big[\phi_0 + O\big(|z|^2\big)\big] \quad \text{in } H^1(\mathbb{R}^3), \quad \omega[z] = e_0 + O\big(|z|^2\big), \\ &E_V\big(\Phi[z]\big) = e_0|z|^2/2 + O\big(|z|^4\big), \quad M\big(\Phi[z]\big) = |z|^2/2 + O\big(|z|^4\big), \end{aligned} \tag{3.4}$$

with a small parameter $z = (\Phi[z]|\phi_0) \in \mathbb{C}$, where

$$M(u) := \int_{\mathbb{R}^d} \frac{|u|^2}{2} dx \tag{3.5}$$

denotes the conserved mass. The other family is generated from the scaling limit of the ground state Q of NLS ($V = 0$), which is large and unstable with positive energy, in the asymptotic form (see [45])

$$\begin{aligned} &\Psi[\zeta] = \zeta Q(|\zeta|x) + O\big(|\zeta|^{-3/2}\big) \quad \text{in } H^1(\mathbb{R}^3), \quad \omega[\zeta] = |\zeta|^2, \\ &E_V\big(\Psi[\zeta]\big) = |\zeta|E_S(Q) + O\big(|\zeta|^{-1}\big), \quad M\big(\Psi[\zeta]\big) = |\zeta|^{-1}M(Q) + O\big(|\zeta|^{-3}\big), \end{aligned} \tag{3.6}$$

with a large parameter $\zeta \in \mathbb{C}$. Since both the families converge to 0 in $L^2(\mathbb{R}^3)$ in the limits $z \to 0$ and $\zeta \to \infty$, respectively, the asymptotic regimes are contained in the small mass region $M(u) \ll 1$. We can prove that for each fixed $M(u) \ll 1$, there is a unique $|z| \ll 1$ such that $\Phi[z]$ are the least energy standing waves, namely the ground states for the prescribed mass $M(u)$, and also there is a unique $|\zeta| \gg 1$ such that $\Psi[\zeta]$ are the second least energy ones, or the first excited states.

Actually, both of them are the ground state solutions of (3.2) for the corresponding $\omega > 0$, which may be obtained by the min–max variational argument. From the dynamical viewpoint, however, it seems more appropriate to compare them in terms of the energy and mass, without fixing parameter ω, which is not intrinsic in the equation.

Gustafson–Nakanishi–Tsai [26] proved the scattering to the ground states Φ for small solutions in $H^1(\mathbb{R}^3)$, that is,

$$\lim_{t \to \pm\infty} \big\|u(t) - \Phi\big[z(t)\big] - v_\pm(t)\big\|_{H^1} = 0, \tag{3.7}$$

for some free solutions $v_\pm$ and some function $z : \mathbb{R} \to \mathbb{C}$ with convergent $|z(t)|$ as $t \to \pm\infty$. The exceptional case $|z(t)| \to 0$ is also included. This result has been extended by Nakanishi [45, 46] to the energy slightly above the first excited solitons Ψ under the radial symmetry restriction (which was not imposed in [26]), with a classification of the global dynamics similar to Theorem 2.1, or more closely to the NLS case in [48].

More precisely, let $\mathcal{E}_V(\mu) = E_V(\Psi[\zeta])$ be the first excited energy for the mass $M(\Psi[\zeta]) = \mu$. Then for sufficiently small $\varepsilon > 0$, all the radial solutions with

$$M(u) < \varepsilon, \quad E_V(u) \le \mathcal{E}_V\big(M(u)\big) + \varepsilon/M(u), \tag{3.8}$$

are classified into $9 = 3 \times 3$ sets characterized by their behavior in $t > 0$ and $t < 0$, either scattering to the ground states Φ as in (3.7), blowing-up, or staying around the excited states Ψ. Moreover, the solutions in the last case make the center-stable manifold of Ψ for $t > 0$, and the center-unstable manifold for $t < 0$.

Note that the restriction $M(u) < \varepsilon$ may be removed if $V = 0$, trivially by using the scaling invariance. Then the above result is reduced to [48], except that the scattering to the excited states is not established in [45]. The same problem with the defocusing nonlinearity

$$i\dot{u} - \Delta u + Vu = -|u|^2 u, \tag{3.9}$$

was also studied in [46], for which all solutions in $H^1(\mathbb{R}^3)$ with small mass scatter to the ground states Φ as $t \to \pm\infty$, while there is no other standing wave in $H^1(\mathbb{R}^3)$.

3.1. Threshold dynamics

As mentioned above, the scattering to the first excited states Ψ remains to be proven on the center-stable manifold. This is mainly because of the lack of complete information on the spectrum of the linearized operator. It is a highly nontrivial problem even without the potential, which was solved by Marzuola–Simpson [43] by a computer-assisted proof.

In the nonradial case, however, a notable difference appears from the case without the potential, where Beceanu [4] proved the scattering to the solitons generated by the Galilei and translation invariance from the ground state Q. Both invariances are destroyed by the potential, and thus the only remaining soliton is fixed at the origin, provided that the potential has a simple shape, e.g., $V(x) = ae^{-|x|^2}$ with some constant $a < 0$. Then the natural conjecture on the dynamics on the center-stable manifold of Ψ is the following:

(1) For $E_V(u) < \mathcal{E}_0(M(u))$, all solutions on the manifold scatter to Ψ.

(2) For $E_V(u) = \mathcal{E}_0(M(u))$, there are solutions with the asymptotic behavior

$$\lim_{t\to\infty} \|u(t) - e^{-i\theta(t)} Q_\omega\big(x - c(t)\big)\|_{H^1} = 0 \tag{3.10}$$

for some $\theta : \mathbb{R} \to \mathbb{R}$ and $c : \mathbb{R} \to \mathbb{R}^3$ satisfying $\dot{\theta} \to \omega$, $|c| \to \infty$, and $\dot{c} \to 0$ as $t \to \infty$, where Q_ω is the ground state of (3.2) with $V = 0$ for some $\omega > 0$ satisfying $M(Q_\omega) = M(u)$. The other solutions on the manifold scatter to Ψ.

(3) For $E_V(u) > \mathcal{E}_0(M(u))$, there are also scatterings into a sum of the Galilei transforms of some Q_ω and the ground states Φ.

In short, the solutions on the center-stable manifold scatter either to the excited states Ψ trapped by the potential at $x = 0$, or to the ground state solitons without potential escaping to $|x| \to \infty$. The threshold between the two cases is the solitons escaping to $|x| \to \infty$ but with the zero asymptotic speed, for which the minimal energy is as in the case (2).

The scattering to Ψ for the solutions initially away from $x = 0$ requires the attractive force of the potential, which may be derived by the Newtonian approximation, but it is valid only on a finite-time interval. Extending it to $t \to \infty$ requires the dissipative effect by the radiation of dispersive waves. Such a scattering result was established by Gang–Sigal [23] in the case of stable solitons for initial data close to the origin. The scattering described above in the case of (3) is also complicated as it contains three different components. Such a scattering result was established by Cuccagna–Maeda [10], also in the stable case for initial data that are already escaping. Classifying all the solutions on the manifold may well require more ideas than the combination of those results.

3.2. Higher mass

Another problem is to extend the classification to $M(u) > \varepsilon$. This sounds plausible at least in the simple defocusing case, where Φ may be extended to all mass as the unique energy minimizers. However, the argument in [45, 46] does not simply extend, because it relies heavily on the smallness of Φ, as well as on $M(u)$, to control all the interactions with the ground states, especially during the concentration–compactness argument for the dispersive component. In the focusing case, the problem does not seem easy even for the smaller potentials, e.g., $V(x) = -ae^{-|x|^2}$ with $0 < a \ll 1$ such that $-\Delta + V > 0$. In this case, there are no small solitons and so the ground states are the perturbations of Q_ω for all mass. Hence it is natural to expect that the same results as in Kenig–Merle [31] (or Holmer–Roudenko [27] for the cubic NLS), and in Nakanishi–Schlag [48] should hold without the small mass condition. It may an option to rely on the stability of the threshold structure with respect to the change of equation (here by the parameter a), including the case of bigger a.

4. TRANSITION BETWEEN MULTISOLITONS

In view of the soliton resolution conjecture, it is an important and necessary step in the study of global dynamics to understand the behavior of solutions migrating between neighborhoods of multisolitons, where the neighborhood may be in the weaker sense as in the previous section. Obviously, this is an even harder problem, so it seems natural to seek for simpler models which admit similar dynamics. The nonlinear Klein–Gordon equation with the damping term

$$\ddot{u} + 2\alpha\dot{u} - \Delta u + u = |u|^{p-1}u, \quad u(t,x) : \mathbb{R}^{1+d} \to \mathbb{R}, \tag{4.1}$$

for some constants $\alpha > 0$, $p > 1$, turns out to be a good model. In fact, Burq–Raugel–Schlag [5] proved the soliton resolution conjecture for all radial solutions in the energy space for the energy-subcritical power $p + 1 < 2^\star$. In this case, solutions asymptotic to solitons are those exponentially converging to some (radial) stationary solutions. The major difference from the conservative NLKG comes from the energy decay

$$\partial_t E_K(\vec{u}) = -2\alpha\|\dot{u}\|_{L^2}^2, \tag{4.2}$$

which makes the analysis much simpler, both in the linear and nonlinear parts. The stable and unstable manifolds had been constructed much earlier by Keller [30] around general stationary solutions. The soliton resolution along time sequences had been established by Feireisl [22] without the radial restriction (but for smaller p), as a consequence of the concentration–compactness due to Lions [38] for the stationary problem. The soliton resolution in the general case takes, as $t \to \infty$, the form of

$$\vec{u}(t) = \sum_{n=1}^{N} \vec{\varphi}_n\big(x - c_n(t)\big) + o(1) \quad \text{in } \mathcal{H}, \tag{4.3}$$

where φ_n are some stationary solutions and $c_n : [0,\infty) \to \mathbb{R}^d$ are some functions satisfying $|c_m - c_n| \to \infty$ for each $m \neq n$. The existence of such solutions with polygonal symmetry was also proven by Feireisl [22]. This allows us to discuss the dynamics around, away, and between multisolitons, as a model case for the more difficult conservative case (NLKG).

More recently, Côte–Martel–Yuan–Zhao [8] characterized the set of asymptotic 2-solitons consisting of the ground state Q of NLKG, namely

$$\vec{u}(t) = \vec{Q}\big(x - c_1(t)\big) - \vec{Q}\big(x - c_2(t)\big) + o(1) \quad \text{in } \mathcal{H} \quad (t \to \infty) \tag{4.4}$$

as a manifold with codimension 2 in the energy space $\mathcal{H}$, together with the asymptotic formula for $c_n(t)$, as well as nonexistence of similar solutions with the same sign on $\vec{Q}$.

Moreover, Côte–Martel–Yuan [7] proved the soliton resolution conjecture in the 1D case without any restriction in the energy space. That is, for any initial data in $\mathcal{H}$, the solution either blows up in finite time, or is asymptotic to a form

$$\vec{u}(t) = \pm \sum_{n=1}^{N} (-1)^n \vec{Q}\big(x - c_n(t)\big) + o(1), \tag{4.5}$$

in $\mathcal{H}$, for some $N \in \mathbb{Z}$ with $c_n - c_{n-1} \to \infty$ as $t \to \infty$. The existence of such solutions for every N is also proven in [7]. To the best of the author's knowledge, this is the first and only result so far of soliton resolution in the entire energy space with no restriction for the full limit $t \to \infty$ that contains moving solitons, provided that the damping is acceptable for the conjecture.

Then it is natural to ask the questions raised in the first section, namely the global dynamics in the full neighborhood of such solutions for all $t \geq 0$. In particular, it is a good place to investigate the migration between different numbers of multisolitons. Ishizuka–Nakanishi [29] considered the simplest case, namely a neighborhood of 2-solitons and transition to 1-solitons, and established a classification into 5 sets of different behavior. To state the precise result, some notation is needed. Let

$$L := -\Delta + 1 - pQ^{p-1} \tag{4.6}$$

be the linearized operator for the static NLKG around the ground state Q, and let $\rho \in H^2(\mathbb{R}^d)$ be its normalized ground state with $L\rho = -\nu^2\rho$ for some constant $\nu > 0$. Define operators acting on $\mathcal{H}$ in the matrix form

$$J := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathcal{L}_\alpha := \begin{pmatrix} L & 2\alpha \\ 0 & 1 \end{pmatrix}. \tag{4.7}$$

Then the linearization of (4.1) around $\vec{Q}$ is written as $\partial_t \vec{u} = J\mathcal{L}_\alpha \vec{u}$. The damped linearized operator $J\mathcal{L}_\alpha$ has eigenfunctions of the form

$$\nu^\pm := -\alpha \pm \sqrt{\nu^2 + \alpha^2}, \quad Y^\pm := (1, \nu^\pm)\rho \implies J\mathcal{L}_\alpha Y^\pm = \nu^\pm Y^\pm. \tag{4.8}$$

For any $z = (z_1, z_2) \in (\mathbb{R}^d)^2$, let $\mathcal{H}_\perp(z) \subset \mathcal{H}$ be the energy subspace defined by

$$\mathcal{H}_\perp(z) := \{\varphi \in \mathcal{H} \mid \langle J\varphi | Y^-(x - z_k)\rangle = 0 \ (k = 1, 2)\}, \tag{4.9}$$

where $\langle \cdot | \cdot \rangle$ denotes the inner product of $(L^2(\mathbb{R}^d))^2$. Then it is easy to see that for $|z_1 - z_2| \gg 1$ (depending on $\alpha > 0$), the energy space is decomposed into a direct sum

$$\mathcal{H} = \mathbb{R}Y^+(x - z_1) \oplus \mathbb{R}Y^+(x - z_2) \oplus \mathcal{H}_\perp(z). \tag{4.10}$$

Let $\mathcal{H}_\perp(z; \delta) := \{\varphi \in \mathcal{H}_\perp(z) \mid \|\varphi\|_{\mathcal{H}} < \delta\}$ be the open ball in the subspace. Then

Theorem 4.1 ([29]). *For any $d \in \mathbb{N}$, $\alpha > 0$ and $p \in (2, 2^\star - 1)$, there is a small $\delta > 0$ such that for any $z \in (\mathbb{R}^d)^2$ satisfying $|z_1 - z_2| > 1/\delta$, there are two Lipschitz functions $G_1, G_2 : (-\delta, \delta) \times \mathcal{H}_\perp(z; \delta) \to (-\delta, \delta)$ with the following properties. For any $h_1, h_2 \in (-\delta, \delta)$ and any $\varphi \in \mathcal{H}_\perp(z; \delta)$, let u be the solution of (4.1) with the initial data*

$$\vec{u}(0) = \sum_{n=1,2} (-1)^n [\vec{Q} + h_n Y^+](x - z_n) + \varphi. \tag{4.11}$$

Then its global behavior is classified by the initial data as follows. Let $n^* := 3 - n$.

(1) *If $h_n < G_n(h_{n^*}, \varphi)$ for both $n = 1, 2$, then u is global with $\|\vec{u}(t)\|_{\mathcal{H}} \to 0$ as $t \to \infty$; we have the global decaying case.*

(2) *If $h_n = G_n(h_{n^*}, \varphi)$ and $h_{n^*} < G_{n^*}(h_n, \varphi)$ for one of $n = 1, 2$, then u is global with $\vec{u}(t) \to (-1)^n \vec{Q}(x - z_\infty)$ in $\mathcal{H}$, for some $z_\infty \in \mathbb{R}^d$, as $t \to \infty$; this is the 1-soliton case with $(-1)^n Q$.*

(3) *If $h_n = G_n(h_{n^*}, \varphi)$ for both $n = 1, 2$, then u is global with*

$$\vec{u}(t) + \vec{Q}(x - z_1(t)) - \vec{Q}(x - z_2(t)) \to 0$$

in $\mathcal{H}$, for some $z_n : [0, \infty) \to \mathbb{R}^d$ satisfying $|z_1(t) - z_2(t)| \to \infty$, as $t \to \infty$; this is the 2-soliton case.

(4) *Otherwise, u blows up in finite time.*

The 2-soliton case (3) may be characterized as $h = G_0(\varphi)$ by another Lipschitz function $G_0 : \mathcal{H}_\perp(z; \delta) \to (-\delta, \delta)^2$.

Moreover, we obtain a full-time description for all those solutions. In particular, in the 1-soliton case (2), the soliton component starting from $(-1)^{n^*}\vec{Q}(x - z_{n^*})$ decays due to the instability, while the other component from $(-1)^n \vec{Q}(x - z_n)$ remains for all time, moving in space and eventually converging to $(-1)^n \vec{Q}(x - z_\infty)$.

The above classification of dynamics is for all initial data in a small neighborhood of any superposition of $\pm\vec{Q}$ with sufficient distance from each other. For each sign of $\pm\vec{Q}$, there is a Lipschitz manifold of codimension 1 consisting of solutions convergent to $\pm\vec{Q}$, translated in space. The two manifolds are joined together at their boundary by the manifold of codimension 2 consisting of solutions asymptotic to 2-solitons, moving away from each other. The connected union of those three manifolds separates the rest of the neighborhood into the open set of global decaying solutions and the open set of blow-up solutions.

The 2-soliton case of (3) was already established by Côte–Martel–Yuan–Zhao [8]. The above theorem extends the dynamics description to the full neighborhood. Note that the manifolds of 1-solitons in (2) are far from those constructed by Keller [30], or by any general method to construct local invariant manifolds, because the manifolds in the above theorem are in a neighborhood of 2-solitons. In other words, it describes the transition from the 2-solitons to the 1-solitons with respect to initial perturbations. In the proof, we also need to describe the transition in time for each initial data on the 1-soliton manifolds. The transition time tends to infinity as the initial data approaches the 2-soliton manifold, so the global dynamics is not at all uniform or continuous within the small neighborhood.

The structure or relation of those three manifolds is in the simplest form as one may expect, by a small perturbation in the energy space from the superposition of two ground states, each having one unstable direction. However, proving this is not so simple as it may appear, because we need to control the two unstable modes with the same eigenvalue, namely $Y^{\pm}(x - z_n(t))$ with $n = 1, 2$. The difficulty comes from the fact that the solitons are getting away from each other, but very slowly, namely $|z_1(t) - z_2(t)| \sim |\log t|$, and the soliton interactions are of order $O(1/t)$ and not integrable in time. In fact, this changes the growth order of the unstable modes from the linearized approximation, making the unstable dynamics far from the superposition of the 1-soliton case. The coupling of the two unstable modes could be even more complicated because the interaction can possibly change the direction of instability, too. It may be illustrated by a simple ODE model with a small parameter $\varepsilon \in \mathbb{R}$, namely

$$\frac{d}{dt}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \nu^+ & \varepsilon^2 e^{-t} \\ \varepsilon^2 e^{-t} & \nu^+ + \frac{\varepsilon}{1+t} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \tag{4.12}$$

which mimics the linearized interaction of the two unstable modes $h_n(t)Y^+(x - c_n(t))$. It is easy to check for the above ODE that

$$\begin{aligned} &\varepsilon > 0, \quad \big(h_1(0), h_2(0)\big) = (1, 0) \implies \lim_{t\to\infty} h_2(t)/h_1(t) = \infty, \\ &\varepsilon < 0, \quad \big(h_1(0), h_2(0)\big) = (0, 1) \implies \lim_{t\to\infty} h_1(t)/h_2(t) = \infty, \end{aligned} \tag{4.13}$$

so the direction of $h(t) \in \mathbb{R}^2$ is completely changed by the interaction. If such a transfer were to happen for the 2-soliton interactions, then the structure of the neighborhood could be more complicated than the above result.

Fortunately, it is not the case because we can prove that nonintegrable interactions are essentially in the scalar part of the above matrix, and the remainder, namely the nonscalar part of the matrix, is uniformly integrable and small. This follows from the reflection symmetry of the equation and the 2-solitons, together with a detailed description of the behavior of solutions in the full neighborhood and all time.

4.1. 3-solitons and soliton merger

It is natural to expect a similar structure for more than 2 solitons ($N \geq 3$ in (4.5)), namely the joint boundary of manifolds with less solitons. However, to prove or disprove such a result seems to be fundamentally more difficult, as the full-time dynamics in the full neighborhood should include a new and more dramatic phenomenon, which may be called *soliton merger*. The distinction between $N \leq 2$ and $N \geq 3$ stems from the fact that the soliton interaction is attractive for the same sign and repulsive for the opposite sign. It is essential for the proof of the above result; 2-solitons with opposite signs are repelling each other as long as both of them exist.

If we start from a small neighborhood of the 3-soliton in the form of (4.5), then the situation is different. Even though the solitons initially have alternating signs, if the middle soliton is destroyed by the instability and the other two survive, then the remaining 2-solitons have the same sign and so start attracting each other. The result of Côte–Martel–Yuan [7] implies that they cannot remain to be 2-solitons, but the solution either blows-up, decays to 0, or is asymptotic to 1-soliton. The transition in the last case from 2-solitons to 1-soliton is very different from the case in Theorem 4.1. As the simplest case, consider the initial data with the even symmetry

$$\vec{u}(0) = \vec{Q}(x+c) + hY^{+}(x+c) + \vec{Q}(x-c) + hY^{+}(x-c) \tag{4.14}$$

with a small parameter $h \in \mathbb{R}$ and a large parameter $c > 1$. It is easy to see that if $0 < h \ll 1$ and $c \gg 1$ is large enough depending on h, then the solution u blows up, and similarly if $0 > h \gg -1$ with $c \gg 1$ then the solution is globally decaying to 0. Since both types of behavior are stable (in 1D), there must be some intermediate $h \in \mathbb{R}$ for a fixed large c such that the solution u converges to $\pm Q$. For such solutions, the even symmetry implies that both the soliton components from $x = \pm c$ are destroyed, but afterward another soliton emerges at $x = 0$. Because of the energy damping, the latter component has to absorb some energy, at least half of $E(\vec{Q})$ from each of the two destroyed solitons, before they are dissipated. This may be regarded as a sort of collision, but very far from the elastic ones in the completely integrable case.

Inelastic collisions have been studied for the generalized KdV by Mizumachi [44] and Martel–Merle [40, 41], where the inelasticity is in a small radiation. For perturbation from the integrable NLS, Perelman [53] proved that the collision splits the smaller soliton into two pieces. For the energy-critical NLW in 5D, Martel–Merle [42] showed the existence of radiation after collision. The above phenomenon looks quite different also from those cases.

Describing the soliton merger and determining the manifold structure around the 3-solitons (or more) seem to be challenging problems. It does not look obvious even whether the merged soliton can take both signs $\pm Q$ or only one. Another question is whether there exists a similar solution in the conservative case such as NLKG. Those questions may be difficult also for numerical experiments because the merger requires some balance between the two dynamics of different orders, namely the exponential instability and the logarithmic movement of solitons.

FUNDING

This work was partially supported by JSPS KAKENHI Grant Number JP17H02854.

REFERENCES

[1] S. Aryan, Existence of two-solitary waves with logarithmic distance for the nonlinear Klein–Gordon equation. *Commun. Contemp. Math.* (2020), 2050091, 25 pp.

[2] T. Aubin, Problèmes isopérimétriques et espaces de Sobolev. *J. Differential Geom.* **11** (1976), no. 4, 573–598.

[3] P. W. Bates and C. K. R. T. Jones, The solutions of the nonlinear Klein–Gordon equation near a steady state. In *Advanced topics in the theory of dynamical systems (Trento, 1987)*, pp. 1–9, Notes Rep. Math. Sci. Eng. 6, Academic Press, Boston, MA, 1989.

[4] M. Beceanu, A critical center-stable manifold for Schrödinger's equation in three dimensions. *Comm. Pure Appl. Math.* **65** (2012), no. 4, 431–507.

[5] N. Burq, G. Raugel, and W. Schlag, Long time dynamics for damped Klein–Gordon equations. *Ann. Sci. Éc. Norm. Supér.* **50** (2017), no. 6, 1447–1498.

[6] R. Côte, Y. Martel, and F. Merle, Construction of multi-soliton solutions for the L^2-supercritical gKdV and NLS equations. *Rev. Mat. Iberoam.* **27** (2011), no. 1, 273–302.

[7] R. Côte, Y. Martel, and X. Yuan, Long-time asymptotics of the one-dimensional damped nonlinear Klein–Gordon equation. *Arch. Ration. Mech. Anal.* **239** (2021), no. 3, 1837–1874.

[8] R. Côte, Y. Martel, X. Yuan, and L. Zhao, Description and classification of 2-solitary waves for nonlinear damped Klein–Gordon equations. *Commun. Math. Phys.* **388** (2021), 1557–1601.

[9] R. Côte and C. Muñoz, Multi-solitons for nonlinear Klein–Gordon equations. *Forum Math. Sigma* **2** (2014), e15, 38 pp.

[10] S. Cuccagna and M. Maeda, On weak interaction between a ground state and a trapping potential. *Discrete Contin. Dyn. Syst.* **35** (2015), 3343–3376.

[11] G. H. Derrick, Comments on nonlinear wave equations as models for elementary particles. *J. Math. Phys.* **5** (1964), 1252–1254.

[12] B. Dodson, Global well-posedness and scattering for the mass critical nonlinear Schrödinger equation with mass below the mass of the ground state. *Adv. Math.* **285** (2015), 1589–1618.

[13] T. Duyckaerts, J. Holmer, and S. Roudenko, Scattering for the non-radial 3D cubic nonlinear Schrödinger equation. *Math. Res. Lett.* **15** (2008), no. 6, 1233–1250.

[14] T. Duyckaerts, H. Jia, C. Kenig, and F. Merle, Soliton resolution along a sequence of times for the focusing energy critical wave equation. *Geom. Funct. Anal.* **27** (2017), no. 4, 798–862.

[15] T. Duyckaerts, C. Kenig, Y. Martel, and F. Merle, Soliton resolution for critical co-rotational wave maps and radial cubic wave equation. 2021, arXiv:2103.01293.

[16] T. Duyckaerts, C. Kenig, and F. Merle, Classification of radial solutions of the focusing, energy-critical wave equation. *Camb. J. Math.* **1** (2013), no. 1, 75–144.

[17] T. Duyckaerts, C. Kenig, and F. Merle, Soliton resolution for the radial critical wave equation in all odd space dimensions. 2019, arXiv:1912.07664.

[18] T. Duyckaerts and F. Merle, Dynamics of threshold solutions for energy-critical wave equation. *Int. Math. Res. Pap.* **2008** (2008), rpn002, 67 pp.

[19] T. Duyckaerts and F. Merle, Dynamic of threshold solutions for energy-critical NLS. *Geom. Funct. Anal.* **18** (2009), no. 6, 1787–1840.

[20] T. Duyckaerts and S. Roudenko, Threshold solutions for the focusing 3D cubic Schrödinger equation. *Rev. Mat. Iberoam.* **26** (2010), no. 1, 1–56.

[21] D. Fang, J. Xie, and T. Cazenave, Scattering for the focusing energy-subcritical nonlinear Schrödinger equation. *Sci. China Math.* **54** (2011), no. 10, 2037–2062.

[22] E. Feireisl, Finite energy travelling waves for nonlinear damped wave equations. *Quart. Appl. Math.* **56** (1998), no. 1, 55–70.

[23] Z. Gang and I. M. Sigal, Relaxation of solitons in nonlinear Schrödinger equations with potential. *Adv. Math.* **216** (2007), no. 2, 443–490.

[24] B. Gidas, W. M. Ni, and L. Nirenberg, Symmetry and related properties via the maximum principle. *Comm. Math. Phys.* **68** (1979), no. 3, 209–243.

[25] J. Ginibre and G. Velo, The global Cauchy problem for the nonlinear Klein–Gordon equation. *Math. Z.* **189** (1985), no. 4, 487–505.

[26] S. Gustafson, K. Nakanishi, and T.-P. Tsai, Asymptotic stability and completeness in the energy space for nonlinear Schrödinger equations with small solitary waves. *Int. Math. Res. Not.* **66** (2004), 3559–3584.

[27] J. Holmer and S. Roudenko, A sharp condition for scattering of the radial 3D cubic nonlinear Schrödinger equation. *Comm. Math. Phys.* **282** (2008), no. 2, 435–467.

[28] S. Ibrahim, N. Masmoudi, and K. Nakanishi, Scattering threshold for the focusing nonlinear Klein–Gordon equation. *Anal. PDE* **4** (2011), no. 3, 405–460.

[29] K. Ishizuka and K. Nakanishi, Global dynamics around 2-solitons for the nonlinear damped Klein–Gordon equations. 2021, arXiv:2109.03737.

[30] C. Keller, Stable and unstable manifolds for the nonlinear wave equation with dissipation. *J. Differential Equations* **50** (1983), no. 3, 330–347.

[31] C. Kenig and F. Merle, Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case. *Invent. Math.* **166** (2006), no. 3, 645–675.

[32] C. Kenig and F. Merle, Global well-posedness, scattering and blow-up for the energy-critical focusing non-linear wave equation. *Acta Math.* **201** (2008), no. 2, 147–212.

[33] R. Killip, B. Stovall, and M. Visan, Scattering for the cubic Klein–Gordon equation in two space dimensions. *Trans. Amer. Math. Soc.* **364** (2012), no. 3, 1571–1631.

[34] R. Killip, T. Tao, and M. Visan, The cubic nonlinear Schrödinger equation in two dimensions with radial data. *J. Eur. Math. Soc. (JEMS)* **11** (2009), no. 6, 1203–1258.

[35] R. Killip, M. Visan, and X. Zhang, The mass-critical nonlinear Schrödinger equation with radial data in dimensions three and higher. *Anal. PDE* **1** (2008), no. 2, 229–266.

[36] J. Krieger, K. Nakanishi, and W. Schlag, Global dynamics of the nonradial energy-critical wave equation above the ground state energy. *Discrete Contin. Dyn. Syst.* **33** (2013), no. 6, 2423–2450.

[37] M. K. Kwong, Uniqueness of positive solutions of $\Delta u - u + u^p = 0$ in $\mathbb{R}^n$. *Arch. Ration. Mech. Anal.* **105** (1989), no. 3, 243–266.

[38] P.-L. Lions, On positive solutions of semilinear elliptic equations in unbounded domains. Nonlinear diffusion equations and their equilibrium states. In *Nonlinear diffusion equations and their equilibrium states, II (Berkeley, CA, 1986)*, pp. 85–122, Math. Sci. Res. Inst. Publ. 13, Springer, New York, 1988.

[39] Y. Martel and F. Merle, Multi solitary waves for nonlinear Schrödinger equations. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **23** (2006), no. 6, 849–864.

[40] Y. Martel and F. Merle, Description of two soliton collision for the quartic gKdV equation. *Ann. of Math.* **174** (2011), no. 2, 757–857.

[41] Y. Martel and F. Merle, Inelastic interaction of nearly equal solitons for the quartic gKdV equation. *Invent. Math.* **183** (2011), no. 3, 563–648.

[42] Y. Martel and F. Merle, Inelasticity of soliton collisions for the 5D energy critical wave equation. *Invent. Math.* **214** (2018), no. 3, 1267–1363.

[43] J. Marzuola and G. Simpson, Spectral analysis for matrix Hamiltonian operators. *Nonlinearity* **24** (2011), no. 2, 389–429.

[44] T. Mizumachi, Weak interaction between solitary waves of the generalized KdV equations. *SIAM J. Math. Anal.* **35** (2003), no. 4, 1042–1080.

[45] K. Nakanishi, Global dynamics above the first excited energy for the nonlinear Schrödinger equation with a potential. *Comm. Math. Phys.* **354** (2017), no. 1, 161–212.

[46] K. Nakanishi, Global dynamics below excited solitons for the nonlinear Schrödinger equation with a potential. *J. Math. Soc. Japan* **69** (2017), no. 4, 1353–1401.

[47] K. Nakanishi and W. Schlag, Global dynamics above the ground state energy for the focusing nonlinear Klein–Gordon equation. *J. Differential Equations* **250** (2011), no. 5, 2299–2333.

[48] K. Nakanishi and W. Schlag, Global dynamics above the ground state energy for the cubic NLS equation in 3D. *Calc. Var. Partial Differential Equations* **44** (2012), no. 1–2, 1–45.

[49] K. Nakanishi and W. Schlag, Global dynamics above the ground state for the nonlinear Klein–Gordon equation without a radial assumption. *Arch. Ration. Mech. Anal.* **203** (2012), no. 3, 809–851.

[50] Z. Nehari, On a class of nonlinear second-order differential equations. *Trans. Amer. Math. Soc.* **95** (1960), 101–123.

[51] T. V. Nguyễn, Existence of multi-solitary waves with logarithmic relative distances for the NLS equation. *C. R. Math. Acad. Sci. Paris* **357** (2019), no. 1, 13–58.

[52] L. E. Payne and D. H. Sattinger, Saddle points and instability of nonlinear hyperbolic equations. *Israel J. Math.* **22** (1975), no. 3–4, 273–303.

[53] G. Perelman, Two soliton collision for nonlinear Schrödinger equations in dimension 1. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **28** (2011), no. 3, 357–384.

[54] S. I. Pohožaev, On the eigenfunctions of the equation $\Delta u + \lambda f(u) = 0$. *Dokl. Akad. Nauk SSSR* **165** (1965), 36–39.

[55] W. Schlag, Stable manifolds for an orbitally unstable nonlinear Schrödinger equation. *Ann. of Math.* **169** (2009), no. 1, 139–227.

[56] W. A. Strauss, Existence of solitary waves in higher dimensions. *Comm. Math. Phys.* **55** (1977), no. 2, 149–162.

[57] G. Talenti, Best constant in Sobolev inequality. *Ann. Mat. Pura Appl.* **110** (1976), 353–372.

KENJI NAKANISHI

Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502, Japan, kenji@kurims.kyoto-u.ac.jp

VARIETY OF FRACTIONAL LAPLACIANS

ALEXANDER I. NAZAROV

ABSTRACT

This paper is a survey of recent results on comparison of various fractional Laplacians.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 47A63; Secondary 35A23, 47G20

KEYWORDS

Fractional Laplacians, nonlocal differential operators

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3842–3853
DOI 10.4171/ICM2022/34

Published by EMS Press

Fractional Laplacians (FLs for brevity) and equations with them have been actively studied in last decades throughout the world in various fields of mathematics (analysis, partial differential equations, the theory of random processes) and its applications (in physics, biology). Hundreds of articles have been written on this topic. Note that the study of such operators and equations is complicated not only by the fact of nonlocality itself, but also by the existence of several nonequivalent definitions of a fractional Laplacian.

Historically, the first FL was the fractional Laplacian of order $s > 0$ in $\mathbb{R}^n$ defined (say, on the Schwartz class $\mathcal{S}(\mathbb{R}^n)$) as

$$(-\Delta)^s u := \mathcal{F}^{-1}\big(|\xi|^{2s}\mathcal{F}u(\xi)\big),$$

where $\mathcal{F}$ is the Fourier transform

$$\mathcal{F}u(\xi) = \frac{1}{(2\pi)^{\frac{n}{2}}}\int_{\mathbb{R}^n} e^{-i\langle\xi,x\rangle}u(x)\,dx.$$

For $s \in (0,1)$, the following relation holds:

$$(-\Delta)^s u(x) = c_{n,s}\cdot \text{V.P.}\int_{\mathbb{R}^n}\frac{u(x)-u(y)}{|x-y|^{n+2s}}\,dy,$$

where

$$c_{n,s} = \frac{2^{2s}s}{\pi^{\frac{n}{2}}}\frac{\Gamma(\frac{n+2s}{2})}{\Gamma(1-s)}.$$

We recall the definitions of the classical Sobolev–Slobodetskii spaces in $\mathbb{R}^n$ (see [21, SUBSECTION 2.3.3] or [7]),

$$H^s(\mathbb{R}^n) = \big\{u \in \mathcal{S}'(\mathbb{R}^n) : \big(1+|\xi|^2\big)^{\frac{s}{2}}\mathcal{F}u(\xi) \in L_2(\mathbb{R}^n)\big\},$$

and corresponding spaces in a (say, Lipschitz and bounded) domain Ω (see [21, SUBSECTION 4.2.1] and [21, SUBSECTION 4.3.2]),

$$H^s(\Omega) = \big\{u|_\Omega : u \in H^s(\mathbb{R}^n)\big\}; \quad \tilde{H}^s(\Omega) = \big\{u \in H^s(\mathbb{R}^n) : \operatorname{supp}(u) \subset \overline{\Omega}\big\}.$$

Notice that the quadratic form of $(-\Delta)^s$ is naturally defined on $H^s(\mathbb{R}^n)$ by[1]

$$\big((-\Delta)^s u, u\big) = \int_{\mathbb{R}^n}|\xi|^{2s}\big|\mathcal{F}u(\xi)\big|^2\,d\xi, \tag{1}$$

and define the *restricted Dirichlet* FL as the positive self-adjoint operator with quadratic form (see, e.g., [1, CHAP. 10])

$$Q_s^{\text{DR}}[u] \equiv \big((-\Delta_\Omega)^s_{\text{DR}}u, u\big) := \big((-\Delta)^s u, u\big); \quad \text{Dom}(Q_s^{\text{DR}}) = \tilde{H}^s(\Omega).$$

Remark 1. For $s \in (0,1)$, the following relation evidently holds:

$$Q_s^{\text{DR}}[u] = \frac{c_{n,s}}{2}\iint_{\mathbb{R}^n\times\mathbb{R}^n}\frac{|u(x)-u(y)|^2}{|x-y|^{n+2s}}\,dx\,dy.$$

[1] As usual, we denote by $(\cdot,\cdot)$ the duality generated by the scalar product in L_2.

Notice that for $s \in (0,1)$ one can also define the *restricted Neumann* (or *regional*) FL by the quadratic form

$$Q_s^{\mathrm{NR}}[u] := \frac{c_{n,s}}{2} \iint_{\Omega\times\Omega} \frac{|u(x)-u(y)|^2}{|x-y|^{n+2s}}\,dx\,dy; \quad \mathrm{Dom}(Q_s^{\mathrm{NR}}) = H^s(\Omega).$$

For some "intermediate" fractional Laplacians of this type, see, e.g., [16] and the references therein.

Now we turn to a different type of FLs, namely, to the spectral ones. Recall that the *spectral Dirichlet and Neumann* FLs are the sth powers of conventional Dirichlet and Neumann Laplacian in the sense of spectral theory. In a Lipschitz bounded domain Ω, they can be defined as the positive self-adjoint operators with quadratic forms

$$Q_s^{\mathrm{DSp}}[u] \equiv \big((-\Delta_\Omega)^s_{\mathrm{DSp}}u, u\big) := \sum_{j=1}^{\infty} \lambda_j^s \big|(u,\varphi_j)\big|^2, \tag{2}$$

$$Q_s^{\mathrm{NSp}}[u] \equiv \big((-\Delta_\Omega)^s_{\mathrm{NSp}}u, u\big) := \sum_{j=0}^{\infty} \mu_j^s \big|(u,\psi_j)\big|^2, \tag{3}$$

where λ_j, φ_j and μ_j, ψ_j are eigenvalues and (normalized) eigenfunctions of the Dirichlet and Neumann Laplacian in Ω, respectively. Notice that $\mu_0 = 0$ and $\psi_0 \equiv \mathrm{const}$.

For $s \in (0,1)$, the domains of these quadratic forms are

$$\mathrm{Dom}(Q_s^{\mathrm{DSp}}) = \tilde{H}^s(\Omega); \quad \mathrm{Dom}(Q_s^{\mathrm{NSp}}) = H^s(\Omega).$$

For $s > 1$, the domains of spectral quadratic forms are more complicated. However, the following relations hold ([21, **THEOREM 1.17.1/1**] and [21, **THEOREM 4.3.2/1**]; see also [12, **LEMMA 1**] and [14, **LEMMA 2**]):

$$\tilde{H}^s(\Omega) = \mathrm{Dom}(Q_s^{\mathrm{DSp}}), \quad 0 < s < \frac{3}{2}; \quad \tilde{H}^s(\Omega) \subsetneq \mathrm{Dom}(Q_s^{\mathrm{DSp}}), \quad s \ge \frac{3}{2};$$
$$\tilde{H}^s(\Omega) = \mathrm{Dom}(Q_s^{\mathrm{NSp}}), \quad 0 < s < \frac{1}{2}; \quad \tilde{H}^s(\Omega) \subsetneq \mathrm{Dom}(Q_s^{\mathrm{NSp}}), \quad s \ge \frac{1}{2}.$$

It follows from the well-known Heinz inequality ([10]; see also [1, **§10.4**]) that for $u \in \tilde{H}^s(\Omega)$, $s \in (0,1)$, the following inequality holds:

$$Q_s^{\mathrm{DSp}}[u] \ge Q_s^{\mathrm{NSp}}[u]. \tag{4}$$

On the other hand, the inequality $Q_s^{\mathrm{DR}}[u] \ge Q_s^{\mathrm{NR}}[u]$ for $u \in \tilde{H}^s(\Omega)$, $s \in (0,1)$, is trivial.

Below we provide a wide generalization and sharpening of (4). To this end, we recall the basic facts on the generalized harmonic extensions related to fractional Laplacians of the order $\sigma \in (0,1)$ and of the *negative* order $-\sigma \in (-1,0)$.

It was known long ago that the square root of Laplacian is related to the harmonic extension and to the Dirichlet-to-Neumann map. In the breakthrough paper [4], the FL $(-\Delta)^\sigma$ (and therefore $(-\Delta_\Omega)^\sigma_{\mathrm{DR}}$) for any $\sigma \in (0,1)$ was related to the *generalized harmonic extension* and to the generalized Dirichlet-to-Neumann map.

Namely, let $u \in \tilde{H}^\sigma(\Omega)$. Then there exists a unique solution $w_\sigma^{\mathrm{DR}}(x,y)$ of the boundary value problem in the half-space

$$-\operatorname{div}(y^{1-2\sigma}\nabla w) = 0 \quad \text{in } \mathbb{R}^n \times \mathbb{R}_+; \quad w|_{y=0} = u,$$

with finite energy (weighted Dirichlet integral)

$$\mathcal{E}_\sigma^{\mathrm{R}}(w) = \int_0^\infty \int_{\mathbb{R}^n} y^{1-2\sigma} \left|\nabla w(x,y)\right|^2 dxdy,$$

and the relation

$$(-\Delta_\Omega)_{\mathrm{DR}}^\sigma u(x) = -C_\sigma \cdot \lim_{y\to 0^+} y^{1-2\sigma} \partial_y w_\sigma^{\mathrm{DR}}(x,y) \tag{5}$$

with

$$C_\sigma = \frac{4^\sigma \Gamma(1+\sigma)}{\Gamma(1-\sigma)}$$

holds in the sense of distributions in Ω and pointwise at every point of smoothness of u. Moreover, the function $w_\sigma^{\mathrm{DR}}(x,y)$ minimizes $\mathcal{E}_\sigma^{\mathrm{R}}$ over the set

$$\mathcal{W}_\sigma^{\mathrm{DR}}(u) = \left\{ w(x,y) : \mathcal{E}_\sigma^{\mathrm{R}}(w) < \infty,\ w|_{y=0} = u \right\},$$

and the following equality holds:

$$Q_\sigma^{\mathrm{DR}}[u] = \frac{C_\sigma}{2\sigma} \cdot \mathcal{E}_\sigma^{\mathrm{R}}(w_\sigma^{\mathrm{DR}}). \tag{6}$$

In [20] this approach was substantially generalized. In particular, for $u \in \tilde{H}^\sigma(\Omega)$ (for $u \in H^\sigma(\Omega)$) there is a unique solution of the boundary value problem in the half-cylinder

$$-\operatorname{div}(y^{1-2\sigma}\nabla w) = 0 \quad \text{in } \Omega \times \mathbb{R}_+; \quad w|_{y=0} = u,$$

satisfying, respectively, the Dirichlet or the Neumann boundary condition on the lateral surface of the half-cylinder and having finite energy

$$\mathcal{E}_\sigma^{\mathrm{Sp}}(w) = \int_0^\infty \int_\Omega y^{1-2\sigma} \left|\nabla w(x,y)\right|^2 dxdy.$$

Denote these solutions $w_\sigma^{\mathrm{DSp}}(x,y)$ and $w_\sigma^{\mathrm{NSp}}(x,y)$, respectively. The following relations hold in the sense of distributions in Ω and pointwise at every point of smoothness of u:

$$(-\Delta_\Omega)_{\mathrm{DSp}}^\sigma u(x) = -C_\sigma \cdot \lim_{y\to 0^+} y^{1-2\sigma} \partial_y w_\sigma^{\mathrm{DSp}}(x,y), \tag{7}$$

$$(-\Delta_\Omega)_{\mathrm{NSp}}^\sigma u(x) = -C_\sigma \cdot \lim_{y\to 0^+} y^{1-2\sigma} \partial_y w_\sigma^{\mathrm{NSp}}(x,y). \tag{8}$$

Moreover, these solutions minimize $\mathcal{E}_\sigma^{\mathrm{Sp}}$ over the sets

$$\mathcal{W}_{\sigma,\Omega}^{\mathrm{DSp}}(u) = \left\{ w(x,y) : \mathcal{E}_\sigma^{\mathrm{Sp}}(w) < \infty,\ w|_{y=0} = u,\ w|_{x\in\partial\Omega} = 0 \right\},$$
$$\mathcal{W}_{\sigma,\Omega}^{\mathrm{NSp}}(u) = \left\{ w(x,y) : \mathcal{E}_\sigma^{\mathrm{Sp}}(w) < \infty,\ w|_{y=0} = u \right\},$$

respectively, and the following equalities hold:

$$Q_\sigma^{\mathrm{DSp}}[u] = \frac{C_\sigma}{2\sigma} \cdot \mathcal{E}_\sigma^{\mathrm{Sp}}(w_\sigma^{\mathrm{DSp}}); \quad Q_\sigma^{\mathrm{NSp}}[u] = \frac{C_\sigma}{2\sigma} \cdot \mathcal{E}_\sigma^{\mathrm{Sp}}(w_\sigma^{\mathrm{NSp}}). \tag{9}$$

Now we set $s = -\sigma \in (-1, 0)$. The operators $(-\Delta_\Omega)^{-\sigma}_{\mathrm{DR}}$, $(-\Delta_\Omega)^{-\sigma}_{\mathrm{DSp}}$, and $(-\Delta_\Omega)^{-\sigma}_{\mathrm{NSp}}$ are defined by corresponding quadratic forms (1)–(3)[2] with domains

$$\mathrm{Dom}(Q^{\mathrm{DR}}_{-\sigma}) = \begin{cases} \tilde{H}^{-\sigma}(\Omega) & \text{if either } n \geq 2 \text{ or } \sigma < \frac{1}{2}; \\ \{u \in \tilde{H}^{-\sigma}(\Omega) : (u, \mathbf{1}) = 0\} & \text{if } n = 1 \text{ and } \sigma \geq \frac{1}{2}; \end{cases}$$

$$\mathrm{Dom}(Q^{\mathrm{DSp}}_{-\sigma}) = H^{-\sigma}(\Omega); \quad \mathrm{Dom}(Q^{\mathrm{NSp}}_{-\sigma}) = \big\{u \in \tilde{H}^{-\sigma}(\Omega) : (u, \mathbf{1}) = 0\big\}.$$

The first two equalities were proved in [14, LEMMA 1]; the third follows from [21, THEOREM 2.10.5/1]. We notice that $(-\Delta_\Omega)^{-\sigma}_{\mathrm{NSp}} u$ is defined up to an additive constant which can be naturally fixed by assumption $((-\Delta_\Omega)^{-\sigma}_{\mathrm{NSp}} u, \mathbf{1}) = 0$.

Remark 2. By [21, THEOREMS 4.3.2/1 AND 2.10.5/1], for $0 < \sigma \leq \frac{1}{2}$ we have $\tilde{H}^{-\sigma}(\Omega) \subseteq H^{-\sigma}(\Omega)$ (even $\tilde{H}^{-\sigma}(\Omega) = H^{-\sigma}(\Omega)$ if $0 < \sigma < \frac{1}{2}$) whereas in the case $\frac{1}{2} < \sigma < 1$, $H^{-\sigma}(\Omega)$ is a subspace of $\tilde{H}^{-\sigma}(\Omega)$. However, in the latter case we can consider an arbitrary $f \in \mathrm{Dom}(Q^{\mathrm{DR}}_{-\sigma})$ as a functional on $H^{\sigma}(\Omega)$, put $\tilde{f} = f|_{\tilde{H}^{-\sigma}(\Omega)} \in \mathrm{Dom}(Q^{\mathrm{DSp}}_{-\sigma})$ and define $Q^{\mathrm{DSp}}_{-\sigma}[f] := Q^{\mathrm{DSp}}_{-\sigma}[\tilde{f}]$.

Next, we connect FLs of the negative order with the generalized Neumann-to-Dirichlet map. It was done in [5] for the spectral Dirichlet FL and in [3] for the FL in $\mathbb{R}^n$ (and therefore for the restricted Dirichlet FL). Variational characterization of these operators was given in [14]. The spectral Neumann FL was considered in [17].

Let $u \in \tilde{H}^{-\sigma}(\Omega)$ (for $n = 1$ and $\sigma \geq \frac{1}{2}$ assume in addition that $(u, \mathbf{1}) = 0$). We consider the problem[3]

$$\tilde{\mathcal{E}}^{\mathrm{R}}_{-\sigma}(w) := \mathcal{E}^{\mathrm{R}}_{\sigma}(w) - 2(u, w|_{y=0}) \to \min \tag{10}$$

on the set $\mathcal{W}^{\mathrm{DR}}_{-\sigma}$, that is, the closure of smooth functions on $\mathbb{R}^n \times \bar{\mathbb{R}}_+$ with bounded support, with respect to $\mathcal{E}^{\mathrm{R}}_{\sigma}(\cdot)$.

If $n > 2\sigma$ (this is a restriction only for $n = 1$) then the minimizer is determined uniquely. Denote it by $w^{\mathrm{DR}}_{-\sigma}(x, y)$. Then (5) and (6) imply

$$(-\Delta_\Omega)^{-\sigma}_{\mathrm{DR}} u(x) = \frac{2\sigma}{C_\sigma} w^{\mathrm{DR}}_{-\sigma}(x, 0); \quad Q^{\mathrm{DR}}_{-\sigma}[u] = -\frac{2\sigma}{C_\sigma} \cdot \tilde{\mathcal{E}}^{\mathrm{R}}_{-\sigma}(w^{\mathrm{DR}}_{-\sigma}) \tag{11}$$

(the first relation holds for a.a. $x \in \Omega$).

In case $n = 1 \leq 2\sigma$, the minimizer $w^{\mathrm{DR}}_{-\sigma}(x, y)$ is defined up to an additive constant. However, by assumption $(u, \mathbf{1}) = 0$, the functional $\tilde{\mathcal{E}}^{\mathrm{R}}_{-\sigma}(w^{\mathrm{DR}}_{-\sigma})$ does not depend on the choice of the constant, and the second relation in (11) holds. The first equality in (11) also holds if we choose the constant such that $w^{\mathrm{DR}}_{-\sigma}(x, 0) \to 0$ as $|x| \to \infty$.

Notice that the function $w^{\mathrm{DR}}_{-\sigma}$ solves the Neumann problem in the half-space

$$-\operatorname{div}(y^{1-2\sigma} \nabla w) = 0 \quad \text{in } \mathbb{R}^n \times \mathbb{R}_+; \qquad \lim_{y \to 0^+} y^{1-2\sigma} \partial_y w = -u$$

(the boundary condition holds in the sense of distributions). So, we can consider $(-\Delta_\Omega)^{-\sigma}_{\mathrm{DR}}$ as the Neumann-to-Dirichlet map, and (10) gives the "dual" variational characterization of negative restricted Dirichlet FL.

[2] We emphasize that $(-\Delta_\Omega)^{-\sigma}_{\mathrm{DR}}$ is not the inverse to $(-\Delta_\Omega)^{\sigma}_{\mathrm{DR}}$.

[3] Notice that by the result of [4] the duality $(u, w|_{y=0})$ is well defined.

In a similar way we provide the "dual" variational characterization of the operators $(-\Delta_\Omega)^{-\sigma}_{\mathrm{DSp}}$ and $(-\Delta_\Omega)^{-\sigma}_{\mathrm{NSp}}$. Namely, let $u \in \tilde{H}^{-\sigma}(\Omega)$ (for the spectral Neumann operator assume in addition that $(u, \mathbf{1}) = 0$). Consider the problem

$$\tilde{\mathcal{E}}^{\mathrm{Sp}}_{-\sigma}(w) = \mathcal{E}^{\mathrm{Sp}}_{\sigma}(w) - 2(u, w|_{y=0}) \to \min$$

respectively on the sets

$$\mathcal{W}^{\mathrm{DSp}}_{-\sigma,\Omega} = \big\{w(x,y) : \mathcal{E}^{\mathrm{Sp}}_{\sigma}(w) < \infty,\ w|_{x\in\partial\Omega} = 0\big\},$$
$$\mathcal{W}^{\mathrm{NSp}}_{-\sigma,\Omega} = \big\{w(x,y) : \mathcal{E}^{\mathrm{Sp}}_{\sigma}(w) < \infty\big\}.$$

Denote corresponding minimizers $w^{\mathrm{DSp}}_{-\sigma}(x,y)$ and $w^{\mathrm{NSp}}_{-\sigma}(x,y)$, respectively[4]. Then (7)–(9) imply

$$Q^{\mathrm{DSp}}_{-\sigma}[u] = -\frac{2\sigma}{C_\sigma}\cdot\tilde{\mathcal{E}}^{\mathrm{Sp}}_{-\sigma}(w^{\mathrm{DSp}}_{-\sigma}); \quad (-\Delta_\Omega)^{-\sigma}_{\mathrm{DSp}}u(x) = \frac{2\sigma}{C_\sigma}w^{\mathrm{DSp}}_{-\sigma}(x,0); \tag{12}$$

$$Q^{\mathrm{NSp}}_{-\sigma}[u] = -\frac{2\sigma}{C_\sigma}\cdot\tilde{\mathcal{E}}^{\mathrm{Sp}}_{-\sigma}(w^{\mathrm{NSp}}_{-\sigma}); \quad (-\Delta_\Omega)^{-\sigma}_{\mathrm{NSp}}u(x) = \frac{2\sigma}{C_\sigma}w^{\mathrm{NSp}}_{-\sigma}(x,0) \tag{13}$$

(the second equalities in (12) and (13) hold for a.a. $x \in \Omega$; in the latter case, we should choose the additive constant such that $w^{\mathrm{NSp}}_{-\sigma}(x,y) \to 0$ as $y \to +\infty$).

Also the functions $w^{\mathrm{DSp}}_{-\sigma}$ and $w^{\mathrm{NSp}}_{-\sigma}$ solve the boundary value problem in the half-cylinder

$$-\operatorname{div}(y^{1-2\sigma}\nabla w) = 0 \quad \text{in } \Omega\times\mathbb{R}_+; \qquad \lim_{y\to 0^+} y^{1-2\sigma}\partial_y w = -u$$

with the Dirichlet or the Neumann boundary condition on the lateral surface $\partial\Omega\times\mathbb{R}_+$, respectively (the Neumann boundary condition on the bottom holds in the sense of distributions).

Now we are in a position to formulate the first group of our main results, namely, the comparison of various FLs in the sense of quadratic forms. These statements were proved in [12, THEOREM 2], [14, THEOREM 1], and [17, THEOREM 3] (for some partial results see also [6,9,19]).

Theorem 3. *Let $s > -1$ and $s \notin \mathbb{N}_0$. Suppose that*[5] *$u \in \tilde{H}^s(\Omega)$, $u \not\equiv 0$. Then the following relations hold:*

$$Q^{\mathrm{DSp}}_s[u] > Q^{\mathrm{DR}}_s[u] > Q^{\mathrm{NSp}}_s[u], \quad \textit{if } s \in (2k, 2k+1),\ k \in \mathbb{N}_0; \tag{14}$$

$$Q^{\mathrm{DSp}}_s[u] < Q^{\mathrm{DR}}_s[u] < Q^{\mathrm{NSp}}_s[u], \quad \textit{if } s \in (2k-1, 2k),\ k \in \mathbb{N}_0. \tag{15}$$

Proof. We prove the theorem in three steps.

1. Let $s \in (0,1)$. Notice that we can assume any function $w \in \mathcal{W}^{\mathrm{DSp}}_{s,\Omega}(u)$ to be extended by zero to $(\mathbb{R}^n \setminus \Omega)\times\mathbb{R}_+$. Then evidently

$$\mathcal{W}^{\mathrm{DSp}}_{s,\Omega}(u) \subset \mathcal{W}^{\mathrm{DR}}_{s}(u) \quad \text{and} \quad \mathcal{E}^{\mathrm{Sp}}_s = \mathcal{E}^{\mathrm{R}}_s|_{\mathcal{W}^{\mathrm{DSp}}_{s,\Omega}(u)}.$$

[4] Notice that $w^{\mathrm{NSp}}_{-\sigma}(x,y)$ is defined up to an additive constant. By assumption $(u,\mathbf{1}) = 0$, the functional $\tilde{\mathcal{E}}^{\mathrm{Sp}}_{-\sigma}(w^{\mathrm{NSp}}_{-\sigma})$ does not depend on the choice of the constant.

[5] We assume in addition that $(u,\mathbf{1}) = 0$ in two cases:

(1) for the left inequality in (15), if $n = 1$ and $s \le -\frac{1}{2}$;

(2) for the right inequality in (15), if $s < 0$.

Therefore, formulae (6) and (9) provide

$$Q_s^{\mathrm{DSp}}[u] = \frac{C_s}{2s} \cdot \min_{w \in \mathcal{W}_{s,\Omega}^{\mathrm{DSp}}(u)} \mathcal{E}_s^{\mathrm{DSp}}(w) \geq \frac{C_s}{2s} \cdot \min_{w \in \mathcal{W}_s^{\mathrm{DR}}(u)} \mathcal{E}_s^{\mathrm{DR}}(w) = Q_s^{\mathrm{DR}}[u],$$

and the first inequality in (14) follows with the "$\geq$" sign.

To complete the proof, we observe that for $u \not\equiv 0$ the corresponding extension w_s^{DSp} (extended by zero) cannot be a solution of the homogeneous equation in the whole half-space $\mathbb{R}^n \times \mathbb{R}_+$ since such a solution should be analytic in the half-space. Thus, it cannot provide $\min_{w \in \mathcal{W}_s^{\mathrm{DR}}(u)} \mathcal{E}_s^{\mathrm{DR}}(w)$.

Since $w_s^{\mathrm{DR}}|_{\Omega \times \mathbb{R}_+} \in \mathcal{W}_{s,\Omega}^{\mathrm{NSp}}(u)$, the proof of the second inequality in (14) is even simpler.

2. Now let $s = -\sigma \in (-1, 0)$. We again extend functions in $\mathcal{W}_{-\sigma,\Omega}^{\mathrm{DSp}}$ by zero and obtain

$$\mathcal{W}_{-\sigma,\Omega}^{\mathrm{DSp}} \subset \mathcal{W}_{-\sigma}^{\mathrm{DR}} \quad \text{and} \quad \tilde{\mathcal{E}}_{-\sigma}^{\mathrm{Sp}} = \tilde{\mathcal{E}}_{-\sigma}^{\mathrm{R}}|_{\mathcal{W}_{-\sigma,\Omega}^{\mathrm{DSp}}}.$$

Therefore, formulae (11) and (12) provide

$$Q_{s,\Omega}^{\mathrm{DSp}}[u] = -\frac{2\sigma}{C_\sigma} \cdot \min_{w \in \mathcal{W}_{-\sigma,\Omega}^{\mathrm{DSp}}} \tilde{\mathcal{E}}_{-\sigma}^{\mathrm{Sp}}(w) \leq -\frac{2\sigma}{C_\sigma} \cdot \min_{w \in \mathcal{W}_{-\sigma}^{\mathrm{DR}}} \tilde{\mathcal{E}}_{-\sigma}^{\mathrm{R}}(w) = Q_s^{\mathrm{DR}}[u],$$

and the left part in (15) follows with the "$\leq$" sign. To complete the proof, we repeat the argument of the first part. The proof of the right part is similar.

3. Now let $s > 1$, $s \notin \mathbb{N}$. We put $k = \lfloor \frac{s+1}{2} \rfloor$ and define for $u \in \tilde{H}^s(\Omega)$,

$$v = (-\Delta)^k u \in \tilde{H}^{s-2k}(\Omega), \quad s - 2k \in (-1, 0) \cup (0, 1).$$

Note that $v \not\equiv 0$ if $u \not\equiv 0$, and

$$(v, \mathbf{1}) = \mathcal{F}v(0) = |\xi|^{2k} \mathcal{F}u(\xi)|_{\xi=0} = 0.$$

Then we have

$$Q_{s,\Omega}^{\mathrm{DSp}}[u] = Q_{s-2k,\Omega}^{\mathrm{DSp}}[v], \quad Q_s^{\mathrm{DR}}[u] = Q_{s-2k}^{\mathrm{DR}}[u], \quad Q_s^{\mathrm{NSp}}[u] = Q_{s-2k}^{\mathrm{NSp}}[u],$$

and the conclusion follows from steps 1 and 2. ■

The second group of our results is related to the pointwise comparison of FLs. These statements were proved in [12, THEOREM 1], [14, THEOREM 3], and [17, THEOREM 4] (a partial result can be found in [8]).

Theorem 4. A. *Let $s \in (0, 1)$, and let $u \in \tilde{H}^s(\Omega)$, $u \geq 0$, $u \not\equiv 0$. Then the following relation holds in the sense of distributions:*

$$(-\Delta_\Omega)_{\mathrm{DSp}}^s u > (-\Delta_\Omega)_{\mathrm{DR}}^s u \quad \text{in } \Omega. \tag{16}$$

B. *Let $s \in (-1, 0)$ for $n \geq 2$, and let $s \in (-\frac{1}{2}, 0)$ for $n = 1$. Suppose that $u \in \tilde{H}^s(\Omega)$, $u \geq 0$ in the sense of distributions, $u \not\equiv 0$. Then the following relation holds:*

$$(-\Delta_\Omega)_{\mathrm{DSp}}^s u < (-\Delta_\Omega)_{\mathrm{DR}}^s u \quad \text{in } \Omega. \tag{17}$$

C. *Suppose that* Ω *is convex. Let* $s \in (0,1)$*, and let* $u \in \tilde{H}^s(\Omega)$*,* $u \geq 0$*,* $u \not\equiv 0$*. Then the following relation holds in the sense of distributions:*

$$(-\Delta_\Omega)^s_{\mathrm{DR}} u > (-\Delta_\Omega)^s_{\mathrm{NSp}} u \quad \textit{in } \Omega. \tag{18}$$

Proof. A. We introduce the function

$$W_s(x,y) := w_s^{\mathrm{DR}}(x,y) - w_s^{\mathrm{DSp}}(x,y).$$

Note that formulae (5) and (7) imply

$$(-\Delta_\Omega)^s_{\mathrm{DSp}} u - (-\Delta_\Omega)^s_{\mathrm{DR}} u = C_\sigma \cdot \lim_{y \to 0^+} y^{1-2s} \partial_y W_s(x,y) \tag{19}$$

in the sense of distributions.

By the strong maximum principle, the assumptions $u \geq 0$, $u \not\equiv 0$ imply that $w_s^{\mathrm{DR}} > 0$ in $\mathbb{R}^n \times \mathbb{R}_+$. Thus, $w_s^{\mathrm{DR}} > w_s^{\mathrm{DSp}}$ at $\partial\Omega \times \mathbb{R}_+$ and, again by the strong maximum principle, $W_s > 0$ in $\Omega \times \mathbb{R}_+$.

After changing of the variable $t = y^{2s}$, the function W_s satisfies the following relations:

$$\Delta_x W_s(x, t^{\frac{1}{2s}}) + 4s^2 t^{\frac{2s-1}{s}} \partial^2_{tt} W_s(x, t^{\frac{1}{2s}}) = 0 \quad \text{in } \Omega \times \mathbb{R}_+; \quad W_s|_{t=0} = 0. \tag{20}$$

The differential operator in (20) satisfies the assumptions of the boundary point lemma [11] at any point $(x,0) \in \Omega \times \{0\}$. Therefore, we have for any $x \in \Omega$,

$$\liminf_{y \to 0^+} y^{1-2s} \partial_y W_s(x,y) = 2s \liminf_{t \to 0^+} \frac{W_s(x, t^{\frac{1}{2s}})}{t} > 0.$$

This gives (16) in view of (19).

B. Put $\sigma = -s \in (0,1)$ and consider extensions $w_{-\sigma}^{\mathrm{DR}}$ and $w_{-\sigma}^{\mathrm{DSp}}$. Making the change of the variable $t = y^{2\sigma}$, we rewrite the boundary value problem for $w_{-\sigma}^{\mathrm{DR}}(x, t^{\frac{1}{2\sigma}})$ as follows:

$$\Delta_x w_{-\sigma}^{\mathrm{DR}} + 4\sigma^2 t^{\frac{2\sigma-1}{\sigma}} \partial^2_{tt} w_{-\sigma}^{\mathrm{DR}} = 0 \quad \text{in } \mathbb{R}^n \times \mathbb{R}_+; \quad \partial_t w_{-\sigma}^{\mathrm{DR}}|_{t=0} = -\frac{u}{2\sigma}. \tag{21}$$

Since $w_{-\sigma}^{\mathrm{DR}}$ vanishes at infinity, $w_{-\sigma}^{\mathrm{DR}}(x, t^{\frac{1}{2\sigma}}) > 0$ for $t > 0$ by the maximum principle.

Further, the function $w_{-\sigma}^{\mathrm{DSp}}(x, t^{\frac{1}{2\sigma}})$ satisfies the equalities in (21) for $x \in \Omega$. Since $w_{-\sigma}^{\mathrm{DSp}}|_{x \in \partial\Omega} = 0$, we infer that the function

$$\hat{W}_s(x,t) := w_{-\sigma}^{\mathrm{DR}}(x, t^{\frac{1}{2\sigma}}) - w_{-\sigma}^{\mathrm{DSp}}(x, t^{\frac{1}{2\sigma}})$$

verifies the following relations:

$$\Delta_x \hat{W}_s + 4\sigma^2 t^{\frac{2\sigma-1}{\sigma}} \partial^2_{tt} \hat{W}_s = 0 \quad \text{in } \Omega \times \mathbb{R}_+; \quad \partial_t \hat{W}_s|_{t=0} = 0; \quad \hat{W}_s|_{x \in \partial\Omega} > 0.$$

Now the boundary point lemma [11] implies $\hat{W}_s(x,0) > 0$, which gives (17) in view of (11) and (12).

C. This statement is more complicated and requires the representation formulae for w_s^{DR} and w_s^{NSp}, see [4] and [20], respectively:

$$w_s^{\mathrm{DR}}(x,y) = \mathrm{const} \cdot \int_{\mathbb{R}^n} \frac{y^{2s} u(z)\, dz}{(|x-z|^2 + y^2)^{\frac{n+2s}{2}}};$$

$$w_s^{\mathrm{NSp}}(x,y) = \sum_{j=0}^{\infty} (u,\psi_j)_{L_2(\Omega)} \cdot \mathcal{Q}_s(y\sqrt{\mu_j})\psi_j(x), \quad \mathcal{Q}_s(\tau) = \frac{2^{1-s}\tau^s}{\Gamma(s)}\mathcal{K}_s(\tau)$$

(here $\mathcal{K}_s(\tau)$ stands for the modified Bessel function of the second kind).

First of all, these formulae imply for $u \geq 0$, $u \not\equiv 0$ that

$$\lim_{y\to+\infty} w_s^{\mathrm{DR}}(x,y) = 0; \qquad \lim_{y\to+\infty} w_s^{\mathrm{NSp}}(x,y) = (u,\psi_0)_{L_2(\Omega)} \cdot \psi_0(x) > 0;$$

the second relation follows from the asymptotic behavior (see, e.g., [20, (3.7)])

$$\mathcal{K}_s(\tau) \sim \Gamma(s)2^{s-1}\tau^{-s}, \quad \text{as } \tau \to 0;$$
$$\mathcal{K}_s(\tau) \sim \left(\frac{\pi}{2\tau}\right)^{\frac{1}{2}} e^{-\tau}\big(1 + O(\tau^{-1})\big), \quad \text{as } \tau \to +\infty.$$

Next, for $x \in \partial\Omega$ we derive by convexity of Ω that

$$\partial_{\mathbf{n}} w_s^{\mathrm{DR}}(x,y) = \mathrm{const} \cdot \int_{\mathbb{R}^n} \frac{y^{2s}\langle(z-x),\mathbf{n}\rangle u(z)\,dz}{(|x-z|^2+y^2)^{\frac{n+2s+2}{2}}} < 0.$$

Thus, the difference $\tilde{W}_s(x,y) = w_s^{\mathrm{NSp}}(x,y) - w_s^{\mathrm{DR}}(x,y)$ has the following properties in the half-cylinder $\Omega \times \mathbb{R}_+$:

$$-\operatorname{div}(y^{1-2s}\nabla\tilde{W}_s) = 0; \quad \tilde{W}_s|_{y=0} = 0; \quad \tilde{W}_s|_{y=\infty} > 0; \quad \partial_{\mathbf{n}}\tilde{W}_s|_{x\in\partial\Omega} > 0.$$

By the strong maximum principle, $\tilde{W}_s > 0$ in $\Omega \times \mathbb{R}_+$. Finally, we apply again the boundary point lemma [11] to the function $\tilde{W}_s(x, t^{\frac{1}{2s}})$ and obtain for $x \in \Omega$,

$$\liminf_{y\to0^+} y^{1-2s}\partial_y\tilde{W}_s(x,y) = 2s \liminf_{t\to0^+} \frac{\tilde{W}_s(x,t^{\frac{1}{2s}})}{t} > 0.$$

This gives (18) in view of (5) and (8). ■

Notice that for nonconvex domains, the relation (18) does not hold in general. We provide a corresponding counterexample.

Example 5. Put temporarily $\Omega = \Omega_1 \cup \Omega_2$ where $\Omega_1 \cap \Omega_2 = \emptyset$. If $u \geq 0$ is a smooth function supported in Ω_1 then easily $(-\Delta_\Omega)^s_{\mathrm{NSp}}u \equiv 0$ in Ω_2. On the other hand, $w_s^{\mathrm{DR}}(x,y) > 0$ for all $x \in \mathbb{R}^n$, $y > 0$, and the boundary point lemma gives $(-\Delta_\Omega)^s_{\mathrm{DR}}u < 0$ in Ω_2. Now we join Ω_1 with Ω_2 by a small channel, and the inequality $(-\Delta_\Omega)^s_{\mathrm{DR}}u < (-\Delta_\Omega)^s_{\mathrm{NSp}}u$ in Ω_2 holds by continuity.

The last group of results in our survey is related to an obvious identity

$$(-\Delta u, u) = \int_\Omega |\nabla u|^2\,dx = \int_\Omega \big|\nabla|u|\big|^2\,dx = (-\Delta|u|,|u|), \quad u \in \tilde{H}^1(\Omega).$$

The following statement was proved in [13, THEOREM 3].[6]

[6] The proof was given for the Dirichlet operators (restricted and spectral); however, it is mentioned in [22, PROPOSITION 1] that for the spectral Neumann FL the proof runs without changes.

Theorem 6. *Let $s \in (0,1)$. Then*

A. *For any $u \in \tilde{H}^s(\Omega)$, we have $|u| \in \tilde{H}^s(\Omega)$ and*

$$Q_s^{\mathrm{DR}}[u] \geq Q_s^{\mathrm{DR}}\big[|u|\big]; \quad Q_s^{\mathrm{DSp}}[u] \geq Q_s^{\mathrm{DSp}}\big[|u|\big];$$

B. *For any $u \in H^s(\Omega)$, we have $|u| \in H^s(\Omega)$ and*

$$Q_s^{\mathrm{NR}}[u] \geq Q_s^{\mathrm{NR}}\big[|u|\big]; \quad Q_s^{\mathrm{NSp}}[u] \geq Q_s^{\mathrm{NSp}}\big[|u|\big].$$

For a sign-changing u, all inequalities are strict.

Proof. For $s \in (0,1]$, the Nemytskii operator $u \mapsto |u|$ is a continuous transform of $H^s(\mathbb{R}^n)$ into itself, see, e.g., [18, THEOREM 5.5.2/3].

There are several proofs of the inequality for Q_s^{DR}; in particular, its representation in Remark 1 provides this inequality immediately. This proof works for Q_s^{NR} as well.

We show another proof that works also for spectral quadratic forms.

Let u be sign-changing. Consider the extension w_s^{DR} and notice that $|w_s^{\mathrm{DR}}| \in \mathcal{W}_s^{\mathrm{DR}}(|u|)$. Therefore,

$$\frac{2s}{C_s} \cdot Q_s^{\mathrm{DR}}\big[|u|\big] = \min_{w \in \mathcal{W}_s^{\mathrm{DR}}(|u|)} \mathcal{E}_s^{\mathrm{R}}(w) \leq \mathcal{E}_s^{\mathrm{R}}\big(|w_s^{\mathrm{DR}}|\big) = \mathcal{E}_s^{\mathrm{R}}(w_s^{\mathrm{DR}}) = \frac{2s}{C_s} \cdot Q_s^{\mathrm{DR}}[u].$$

Moreover, w_s^{DR} is sign-changing, so $|w_s^{\mathrm{DR}}|$ cannot be a solution of the homogeneous equation by the maximum principle and thus cannot be a minimizer for the energy. ■

What happens for $s > 1$? If $s \in (1, \frac{3}{2})$ then the operator $u \mapsto |u|$ is a bounded transform of $H^s(\mathbb{R}^n)$ into itself, see, e.g., [2, SECTION 4]. To the best of our knowledge, its continuity is still an open problem. Moreover, it is easy to show that the assumption $s < \frac{3}{2}$ cannot be improved, see, e.g., [15, EXAMPLE 1].

So, the question about the behavior of quadratic forms of FLs under the transform $u \mapsto |u|$ seems reasonable for $s \in (1, \frac{3}{2})$. The following statement was proved in [15].

Theorem 7. *Let $s \in (1, \frac{3}{2})$, and let $u \in \tilde{H}^s(\Omega)$ be sign-changing. Then*

$$Q_s^{\mathrm{DR}}[u] < Q_s^{\mathrm{DR}}\big[|u|\big]. \tag{22}$$

The sketch of proof. Define $u^{\pm} = \frac{1}{2}(|u| \pm u)$ and assume for a moment that u^+ and u^- are smooth and have disjoint supports. Then

$$Q_s^{\mathrm{DR}}\big[|u|\big] - Q_s^{\mathrm{DR}}[u] = 4\big((-\Delta_\Omega)_{\mathrm{DR}}^s u^+, u^-\big) = 4\big((-\Delta_\Omega)_{\mathrm{DR}}^{s-1} u^+, (-\Delta) u^-\big).$$

By Remark 1,

$$\begin{aligned}
&\big((-\Delta_\Omega)_{\mathrm{DR}}^{s-1} u^+, (-\Delta) u^-\big) \\
&\quad = \frac{c_{n,s-1}}{2} \iint_{\mathbb{R}^n \times \mathbb{R}^n} \frac{(u^+(x) - u^+(y))(-\Delta u^-(x) + \Delta u^-(y))}{|x-y|^{n+2s-2}}\, dx\, dy \\
&\quad = c_{n,s-1} \iint_{\mathbb{R}^n \times \mathbb{R}^n} \frac{u^+(x) \Delta u^-(y)}{|x-y|^{n+2s-2}}\, dx\, dy
\end{aligned}$$

(notice that $u^+(x)u^-(x) \equiv 0$).

Since the supports of u^+ and u^- are disjoint, we can integrate by parts. Using the definition of $c_{n,s}$, we derive

$$\Delta_y \frac{c_{n,s-1}}{|x-y|^{n+2s-2}} = \frac{2s(n+2s-2)c_{n,s-1}}{|x-y|^{n+2s}} = -\frac{c_{n,s}}{|x-y|^{n+2s}}$$

and obtain

$$Q_s^{\mathrm{DR}}[|u|] - Q_s^{\mathrm{DR}}[u] = -4c_{n,s} \iint_{\mathbb{R}^n \times \mathbb{R}^n} \frac{u^+(x)u^-(y)}{|x-y|^{n+2s}}\, dxdy.$$

It remains to observe that $c_{n,s} < 0$ for $s \in (1,2)$, and (22) follows.

In the general case, the result was obtained in [15] using a quite nontrivial approximation procedure. ■

Conjecture 8. *For $s \in (1, \frac{3}{2})$, the inequalities similar to (22) should hold for spectral quadratic forms.*

ACKNOWLEDGMENTS

I am deeply grateful to my friend and coauthor Professor Roberta Musina for many years of cooperation. Indeed, without her, I could not even get started in this field.

FUNDING

This work was supported by RFBR grant 20-01-00630.

REFERENCES

[1] M. S. Birman and M. Z. Solomyak, *Spectral theory of self-adjoint operators in Hilbert space*. 2nd edn., revised and extended, Lan', St. Petersburg, 2010 (in Russian); English transl. of the 1st ed.: Math. Appl., Sov. Ser., 5, Kluwer, Dordrecht, 1987.

[2] G. Bourdaud and W. Sickel, Composition operators on function spaces with fractional order of smoothness. In *Harmonic analysis and nonlinear partial differential equations*, edited by T. Ozawa and M. Sugimoto, pp. 93–132, RIMS Kôkyûroku Bessatsu B26, Res. Inst. for Math. Sci, Kyoto, 2011.

[3] X. Cabré and Y. Sire, Nonlinear equations for fractional Laplacians. I: Regularity, maximum principles, and Hamiltonian estimates. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **31** (2014), no. 1, 23–53.

[4] L. Caffarelli and L. Silvestre, An extension problem related to the fractional Laplacian. *Comm. Partial Differential Equations* **32** (2007), no. 8, 1245–1260.

[5] A. Capella, J. Dávila, L. Dupaigne, and Y. Sire, Regularity of radial extremal solutions for some non-local semilinear equations. *Comm. Partial Differential Equations* **36** (2011), no. 8, 1353–1384.

[6] Z.-Q. Chen and R. Song, Two-sided eigenvalue estimates for subordinate processes in domains. *J. Funct. Anal.* **226** (2005), 90–113.

[7] E. Di Nezza, G. Palatucci, and E. Valdinoci, Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.* **136** (2012), no. 5, 521–573.

[8] M. M. Fall, Semilinear elliptic equations for the fractional Laplacian with Hardy potential. *Nonlinear Anal.* **193** (2020), 111311, 29 pp. arXiv:1109.5530v4, 2012.

[9] R. L. Frank and L. Geisinger, Refined semiclassical asymptotics for fractional powers of the Laplace operator. *J. Reine Angew. Math.* **712** (2016), 1–37.

[10] E. Heinz, Beiträge zur Störungstheorie der Spektralzerlegung. *Math. Ann.* **193** (1951), 415–438 (in German).

[11] L. I. Kamynin and B. N. Himčenko, Theorems of Giraud type for second order equations with a weakly degenerate non-negative characteristic part. *Sib. Math. J.* **18** (1977), 76–91.

[12] R. Musina and A. I. Nazarov, On fractional Laplacians. *Comm. Partial Differential Equations* **39** (2014), no. 9, 1780–1790.

[13] R. Musina and A. I. Nazarov, On the Sobolev and Hardy constants for the fractional Navier Laplacian. *Nonlinear Anal.* **121** (2015), 123–129.

[14] R. Musina and A. I. Nazarov, On fractional Laplacians–2. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **33** (2016), no. 6, 1667–1673.

[15] R. Musina and A. I. Nazarov, A note on truncations in fractional Sobolev spaces. *Bull. Math. Sci.* **9** (2019), no. 1, 1950001, 7 pp.

[16] R. Musina and A. I. Nazarov, Strong maximum principles for fractional Laplacians. *Proc. Roy. Soc. Edinburgh Sect. A* **149** (2019), no. 5, 1223–1240.

[17] A. I. Nazarov, On comparison of fractional Laplacians. 2021, arXiv:2108.05416.

[18] T. Runst and W. Sickel, *Sobolev spaces of fractional order, Nemytskij operators, and nonlinear partial differential equations*. De Gruyter Ser. Nonlinear Anal. Appl. 3, de Gruyter, Berlin, 1996.

[19] R. Servadei and E. Valdinoci, On the spectrum of two different fractional operators. *Proc. Roy. Soc. Edinburgh Sect. A* **144** (2014), no. 4, 831–855.

[20] P. R. Stinga and J. L. Torrea, Extension problem and Harnack's inequality for some fractional operators. *Comm. Partial Differential Equations* **35** (2010), no. 11, 2092–2122.

[21] H. Triebel, *Interpolation theory, function spaces, differential operators*. Deutscher Verlag Wissensch, Berlin, 1978.

[22] N. S. Ustinov, On solvability of a semilinear problem with spectral Neumann Laplacian and critical right-hand side. *Algebra i Analiz* **33** (2021), no. 1, 194–212 (Russian).

ALEXANDER I. NAZAROV

PDMI RAS, Fontanka 27, St. Petersburg 191023, Russia, al.il.nazarov@gmail.com

FORMATION OF SINGULARITIES IN NONLINEAR DISPERSIVE PDES

GALINA PERELMAN

ABSTRACT

This contribution addresses the problem of singularity formation in nonlinear dispersive equations. Despite significant progress made in the last 20 years, for most even simplest canonical models our understanding of the question is far from being complete. The aim of this note is to give a selection of results and open questions illustrating the present state of the problem in the context of some basic model equations, mostly of Schrödinger type, such as the semilinear Schrödinger and Schrödinger map equations, putting an emphasis on the role of solitons in the mechanisms of singularity formation.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 35Q55; Secondary 35B44, 35Q51

KEYWORDS

Nonlinear Schrödinger equations, formation of singularities, ground states

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3854–3879
DOI 10.4171/ICM2022/123

1. INTRODUCTION

Many physical processes involving nonlinear evolution of wave-like objects are modeled by semilinear Hamiltonian PDEs of dispersive type. Among the canonical examples are the nonlinear Schrödinger equation (NLS)

$$iu_t = -\Delta u + \mu |u|^{2p} u, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d,$$

and the nonlinear wave equation (NLW)

$$u_{tt} = \Delta u - \mu |u|^{2p} u, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d, \tag{1.1}$$

where $p > 0$ and $\mu \in \{-1, 1\}$. The equations are said to be focusing if $\mu = -1$ and defocusing if $\mu = 1$.

Other important examples are the wave and Schrödinger map equations. They are respectively the hyperbolic and Schrödinger analogues of the harmonic map heat flow, which is the gradient flow associated to the Dirichlet energy $\int_{\mathbb{R}^d} |\nabla u(x)|^2 dx$ for maps u from $\mathbb{R}^d$ to an embedded Riemannian manifold[1] $M \subset \mathbb{R}^n$. We will limit ourselves to the case of $M = \mathbb{S}^2 \subset \mathbb{R}^3$, where the equations take a particular simple form:

$$u_{tt} = \Delta u + u\big(|\nabla u|^2 - |u_t|^2\big) \tag{1.2}$$

for wave maps,

$$u_t = u \times \Delta u$$

for Schrödinger maps, and

$$u_t = \Delta u + u|\nabla u|^2$$

for the heat flow. Here u is a map from $\mathbb{R}_t \times \mathbb{R}_x^d$ to $\mathbb{S}^2 \subset \mathbb{R}^3$.

Of course, the first question in the theory of such equations is the local well-posedness of the corresponding Cauchy problem (including existence, uniqueness, and continuous dependence of solutions on initial data). But once the local well-posedness is understood, which is often the case at least for the simplest models, the next step is to study the qualitative behavior of solutions, and in particular to answer the following questions:

- Do all maximal solutions exist globally in time or does finite time blow-up occur? If yes, for what classes of initial data?
- If the solution blows up in finite time, can one determine when, where, and how the singularities form?
- If the solution is global, can one determine its behavior as $t \to \infty$?

Despite substantial progress made in the last 20–30 years, a complete answer to most of these questions remains an open problem even for the relatively simple models. The general belief is that in nonlinear dispersive equations the linear dispersion tends to stabilize the dynamics, leading to a kind of universality in the long-time behavior: global solutions

[1] For the Schrödinger map equation, one needs M to be a Kähler manifold.

are expected to decompose asymptotically as $t \to \infty$ into a sum of decoupled nonlinear bound states, such as solitary wave solutions, plus a radiation that disperses to zero, typically as a free linear wave. This prediction is known as the soliton resolution conjecture and is motivated by the theory of completely integrable equations such as the one-dimensional cubic NLS, KdV, and mKdV equations, for which this kind of behavior can be justified by means of the inverse scattering method at least for some classes of initial data. For the nonintegrable equations, this conjecture remains largely open. Most of the available results concern either the soliton-free dynamics typical for the defocusing nonlinearities or small data (usually below some threshold determined by the ground state of the equation), or the perturbative regimes near a single soliton or near a superposition of well-decoupled solitons. The only exceptions are wave-type models, such as the energy critical NLW equation ((1.1) with $p = \frac{2}{d-2}$, $d \geq 3$) and the energy critical wave maps into the two sphere ((1.2) with $d = 2$), for which satisfactory global results begin to emerge starting from the breakthrough work of Duyckaerts, Merle, and Kenig [25], where a resolution into solitons was established for all energy bounded radial solutions of the energy critical nonlinear wave equation in dimension 3 (see Section 5). For the Schrödinger-type models, such results are still out of reach. The only known global results in this setting correspond to a much weaker version of the soliton resolution conjecture as that proved by Tao [106] for the nonlinear Schrödinger equations with mass supercritical and energy subcritical nonlinearities in high dimensions. This weak version gives a decomposition of any global, energy bounded solution into a dispersive part that evolves according to the linear Schrödinger equation, a sum of decoupled pieces, each piece evolving (modulo the space translations) on a compact invariant set, and a remainder going to zero as $t \to \infty$ in the energy space, thus reducing the problem to a classification of solutions with a compact trajectory, the question which is largely open for the NLS equations.

In the case of finite time blow-up, even less is known. For the Schrödinger-type equations, the theory is still on the level of searching for possible blow-up mechanisms and studying their stability. Below we give a selection of corresponding results. The choice of the results is unavoidably related to those aspects of the problem which are most familiar to the author, the list of the references is by no means complete.

2. OVERVIEW OF THE WELL-POSEDNESS THEORY FOR THE NLS EQUATION

Consider the nonlinear Schrödinger equation on $\mathbb{R}^d$:

$$iu_t = -\Delta u + \mu |u|^{2p} u, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^d, \quad p > 0, \quad \mu \in \{-1, 1\}, \tag{2.1}$$

with initial condition

$$u|_{t=0} = u_0 \in H^s(\mathbb{R}^d). \tag{2.2}$$

The solutions to (2.1), (2.2) satisfy formally mass, energy, and momentum conservation laws:

$$M\big(u(t)\big) \equiv \int_{\mathbb{R}^d} \big|u(t,x)\big|^2 dx = M(u_0),$$
$$E\big(u(t)\big) \equiv \int_{\mathbb{R}^d} \bigg(\big|\nabla u(t,x)\big|^2 + \frac{\mu}{p+1}\big|u(t,x)\big|^{2p+2}\bigg) dx = E(u_0),$$
$$P\big(u(t)\big) \equiv \operatorname{Im} \int_{\mathbb{R}^d} \overline{u(t,x)} \nabla u(t,x) dx = P(u_0).$$

The NLS equation is invariant with respect to time translations, spatial translations and rotations, and phase rotations. A less evident symmetry is the invariance under Galilei transformations, $u(t,x) \mapsto e^{-i\frac{|v|^2}{4}t+i\frac{v}{2}\cdot x}u(t,x-vt)$, $v \in \mathbb{R}^d$. In the case of $p = \frac{2}{d}$, there is an additional symmetry

$$u(t,x) \mapsto \frac{1}{|t|^{\frac{d}{2}}} e^{i\frac{|x|^2}{4t}} u\bigg(-\frac{1}{t}, \frac{x}{t}\bigg), \quad t \neq 0, \tag{2.3}$$

called pseudoconformal symmetry.

The NLS equation (2.1) is also invariant with respect to the scaling, $u(t,x) \mapsto \lambda^{\frac{1}{p}} u(\lambda^2 t, \lambda x)$, $\lambda > 0$, that preserves the homogeneous Sobolev norm $\|u_0\|_{\dot{H}^{s_c}(\mathbb{R}^d)}$ with $s_c = \frac{d}{2} - \frac{1}{p}$. This defines a notion of criticality: the Cauchy problem (2.1), (2.2) is said to be subcritical if $s > s_c$, critical if $s = s_c$, and supercritical if $s < s_c$. As we will see below, the notion of criticality plays a fundamental role in the well-posedness theory of (2.1). Of a particular interest are the mass critical case $s_c = 0$ and the energy critical case $s_c = 1$ when the critical regularity coincides with one of the conservation laws.

The local well-poseness of the NLS equation is well understood (see, e.g., [10, 105] and the references therein). The Cauchy problem (2.1), (2.2) is locally well-posed in H^s for[2] $s \geq \max\{0, s_c\}$, and typically, also in $\dot{H}^{s_c}$ if $s_c \geq 0$. In the latter case the solutions arising from $\dot{H}^{s_c}$ small initial data are global and scatter both forward and backward in time (i.e., converge to a linear solution as $t \to \pm\infty$).

In the subcritical case $s > s_c$, the lifespan of the solutions admits a lower bound depending only on the H^s norm of initial data,[3] which in a standard way implies that the solution of (2.1), (2.2) is either global or its H^s norm becomes unbounded in finite time. By the mass and energy conservation, this ensures global well-posedness in H^s for $s \geq 0$ in the mass subcritical range $p < \frac{2}{d}$ independently of the sign of μ, and in H^s with $s \geq 1$ in the defocusing energy subcritical case $p < \frac{2}{d-2}$.

The critical well-posedness ($s = s_c \geq 0$) is more subtle. In this case the lifespan of solutions given by the local theory depends on the profile of the initial data, not only on its $\dot{H}^{s_c}$ norm. In the defocusing case, however, typically the uniform boundedness of the solution in $\dot{H}^{s_c}$ on its maximal interval of existence implies that the solution is global and scatters. In particular, one has global well-posedness and scattering in $\dot{H}^{s_c}$ for the defocusing

[2] In the case when p is not an integer, one has also to assume that s is compatible with the smoothness of the nonlinear term.

[3] In fact, typically one has a slightly stronger result including the persistence of regularity: if $u_0 \in H^{s'}$ with $s' > s$, then the solution stays in $H^{s'}$ as long as it exists in H^s.

energy critical ($s_c = 1$) and mass critical ($s_c = 0$) NLS equations. In the energy critical case, this was proved by Bourgain [6], Grillakis [34], Tao [104] for spherically symmetric initial data, and by Colliander, Keel, Staffilani, Takaoka, and Tao [12], Ryckman and Visan [98], and Visan [111] for general data. We also refer to the seminal paper of Kenig and Merle [48] where the powerful concentration compactness/rigidity method was introduced. The corresponding result for the mass critical NLS was proved by Killip, Tao, Visan, and Zhang [50,53,109] in the case of spherically symmetric initial data, and by Dodson for general initial data, see [17] and the references therein. In the energy supercritical case $s_c > 1$, the fact that the $\dot{H}^{s_c}$ bounds imply global existence and scattering was established by Killip and Visan [51] in dimension $d \geq 5$. We also refer to Miao, Murphy, and Zheng [84] for the case of $d = 4$. Similar results hold in the mass supercritical, energy subcritical range, see, e.g., Kenig and Merle [49], although in this case unconditional global existence and scattering in $\dot{H}^{s_c}$ is expected, see Dodson [19] for some partial results in this direction as well as for the history of the problem. In the energy supercritical case, finite time blow-up may occur. This has been recently proved by Merle, Raphaël, Rodnianski, and Szeftel [79], see Section 6.

For the focusing NLS, the picture is different. On the one hand, large initial data may lead to finite time blow-up as soon as $0 \leq s_c$, and, on the other hand, if $s_c \leq 1$, the equation admits solitary wave solutions, which shows that even for global in time solutions scattering may not occur.

The existence of finite time blow-up for the focusing NLS with $p \geq \frac{2}{d}$ follows from the virial identity [33]:

$$\frac{d^2}{dt^2} \int |x|^2 \big|u(x,t)\big|^2 dx = 8E(u) - \frac{4(dp-2)}{p+1} \int_{\mathbb{R}^d} \big|u(t,x)\big|^{2p+2}\, dx,$$

which holds for finite variance $H^1 \cap H^{s_c}$ solutions of (2.1), and shows that if $E(u) < 0$, then the solution breaks down in finite time. Note that in the mass critical case, any blow-up solution is trivially bounded in the critical Sobolev space. For the energy critical NLS, one also might have blow-up solutions with bounded $\dot{H}^1$ norm. In the case $0 < s_c < 1$, the situation is different: solutions that stay bounded in $\dot{H}^{s_c}$ are expected to be global. In the radial case this property was proved by Merle and Raphaël [76].

As mentioned above, the focusing NLS with $s_c \leq 1$ admits a family of solitary wave solutions. In the energy subcritical range $0 < p < \frac{2}{d-2}$, they have the form

$$u(t,x) = e^{i(\omega - \frac{v^2}{4})t + i\frac{v}{2}\cdot x} Q_\omega(x - vt),$$

where $v \in \mathbb{R}^d$, $\omega > 0$, and the profile Q_ω solves the elliptic equation

$$-\Delta Q_\omega + \omega Q_\omega - |Q_\omega|^{2p} Q_\omega = 0,$$

which after the rescaling $Q_\omega(y) = \omega^{\frac{1}{2p}} Q(\omega^{\frac{1}{2}} x)$ takes the form

$$-\Delta Q + Q - |Q|^{2p} Q = 0. \tag{2.4}$$

For any $0 < p < \frac{2}{d-2}$, there is a unique positive radially symmetric H^1 solution to this equation, called ground state (see, e.g., [10,105] and the references therein); the ground state solution is smooth and exponentially decaying. In the one-dimensional case, the ground state is explicit, namely $Q(x) = \frac{(p+1)^{\frac{1}{2p}}}{\cosh^{\frac{1}{p}}(px)}$. If $d \geq 2$, equation (2.4) has also radial H^1 solutions which change sign (called exited states).

In the energy critical case, solitary wave solutions are given by

$$u(t,x) = e^{-i\frac{v^2}{4}t + i\frac{v}{2}\cdot x} W(x - vt),$$

where $v \in \mathbb{R}^d$ and W is a stationary solution, satisfying

$$\Delta W + |W|^{\frac{4}{d-2}} W = 0, \quad W \in \dot{H}^1(\mathbb{R}^d). \tag{2.5}$$

The radial solutions of this elliptic equation are completely classified: they are of the form $W_{\alpha,\lambda}(x) = e^{i\alpha}\lambda^{\frac{d-2}{2}} W(\lambda x)$, where

$$W(x) = \left(1 + \frac{|x|^2}{d(d-2)}\right)^{-\frac{d-2}{2}}. \tag{2.6}$$

3. MASS CRITICAL FOCUSING NLS

In this section we consider the mass critical NLS

$$\begin{cases} iu_t = -\Delta u - |u|^{\frac{4}{d}} u, & (t,x) \in \mathbb{R} \times \mathbb{R}^d, \\ u|_{t=0} = u_0. \end{cases} \tag{3.1}$$

3.1. Global existence and scattering below the ground state

For $p \geq \frac{2}{d}$, the local well-posedness theory ensures global existence and scattering for initial data with small $\dot{H}^{s_c}$ norm. Typically, in the range $0 \leq s_c \leq 1$, this smallness can be related to the ground state of the problem. In the case of the mass critical NLS, one has:

Theorem 3.1 (Global existence and scattering below the ground state). *For any $u_0 \in L^2(\mathbb{R}^d)$ with $\|u_0\|_{L^2} < \|Q\|_{L^2}$, the solution to (3.1) is global and scatters forward and backward in time (that is, there exist $u_-, u_+ \in L^2$ such that $\|u(t) - e^{i\Delta t}u_\pm\|_{L^2} \to 0$ as $t \to \pm\infty$).*

This result has a long history. In the case of H^1 solutions, the global existence follows from the variational characterization of the mass critical ground state proved by M. Weinstein in [112]:

$$\|f\|_{L^{2+\frac{4}{d}}}^{2+\frac{4}{d}} \leq \frac{d+2}{d}\left(\frac{\|f\|_{L^2}^{\frac{4}{d}}}{\|Q\|_{L^2}^{\frac{4}{d}}}\right)\|\nabla f\|_{L^2}^2, \quad \forall f \in H^1, \tag{3.2}$$

the equality being achieved if and only if $f(x) = zQ(\lambda(x-a))$ for some $z \in \mathbb{C}$, $\lambda > 0$ and $a \in \mathbb{R}^d$. Inequality (3.2) shows that the H^1 norm of the solutions is controlled by their mass and energy as soon as $M(u) < M(Q)$. Global existence and scattering for L^2 data with finite invariance is also classical, see, e.g., [10]. The general L^2 result is much more difficult and has been proved only recently, see Killip, Tao, Visan, Zhang [50, 53] and Dodson [16].

Applying the pseudoconformal transformation to the soliton $e^{it}Q$ gives an explicit blow-up solution

$$S(t) = \frac{1}{t^{\frac{d}{2}}} e^{i\frac{|x|^2}{4t} - i\frac{1}{t}} Q\left(\frac{x}{t}\right) \tag{3.3}$$

that has the same mass as the ground state Q. Thus the bound $M(u) < M(Q)$ is optimal not only for scattering but also for global existence.

The minimal mass dynamics is also well understood. In [72] Merle proved that up to the symmetries of the equation, $S(t)$ is the only H^1 minimal mass finite time blow-up solution, see also Hmidi, Keraani [40] for a simplified proof. By the pseudoconformal invariance, this result also implies that any global minimal mass nonscattering H^1 solution with finite variance is a ground state solitary wave. The L^2 case was studied by Dodson [20,21] who proved:

Theorem 3.2 (Threshold dynamics, Dodson [20, 21]). *Let $1 \leq d \leq 15$ and consider $u_0 \in L^2(\mathbb{R}^d)$ with $\|u_0\|_{L^2} = \|Q\|_{L^2}$. Then either the solution of (3.1) is global and scatters as $t \to \pm\infty$ or it coincides with $e^{it}Q$ up to the symmetries of the equation (including the pseudoconformal symmetry).*

We next turn to the case $M(u) > M(Q)$. In this case the virial identity ensures the existence of a large set of initial data leading to finite time blow-up both forward and backward in time, but gives no information on the structure of the singularity, and for general large mass data little is known in this direction. Essentially, only two general results are available. First, one has the following lower bound on the blow-up rate of H^s solutions, which is a direct consequence of the scaling invariance of the problem: if u_0 is in H^s with $s > 0$, such that the corresponding solution u blows up in finite time $T > 0$, then

$$\|u(t)\|_{\dot{H}^s} \geq \frac{C(u_0)}{(T-t)^{\frac{s}{2}}}, \quad \forall t \in [0, T[.$$

Second, it is known that any blow-up solution concentrates at the blow-up time at least the mass of the ground state. This is a consequence of Theorem 3.1. We refer to [50] and to the references therein for the precise statements and for the history of the L^2 concentration results. For masses slightly above the critical mass, more can be done. We discuss the corresponding results in the next subsection.

3.2. Near ground state blow-up dynamics

In this subsection, we focus on the H^1 blow-up solutions with mass slightly above the critical mass:

$$u_0 \in H^1(\mathbb{R}^d), \quad \|Q\|_{L^2} < \|u_0\|_{L^2} \leq \|Q\|_{L^2} + \alpha, \quad 0 < \alpha \ll 1. \tag{3.4}$$

The mass and energy conservation, together with the variational characterization of the ground state (3.2), ensures that near the blow-up time these solutions behave as a modulated ground state, admitting a decomposition of the following form:

$$u(t,x) = \lambda^{\frac{d}{2}}(t) e^{i\mu(t)} \big(Q(z) + r(t,z)\big), \quad z = \lambda(t)\big(x - q(t)\big),$$

with $\lambda(t) \sim \|\nabla u(t)\|_{L^2}$, $\|r(t)\|_{H^1} \ll 1$. Although giving no information on the blow-up rate $\lambda(t)$ and on the blow-up location $q(t)$, this variational result is conceptually important, showing that the blow-up profiles arising from initial data (3.4) are close to the ground state, and thus providing a starting point for their perturbative analysis. Such analysis was initiated in [90] where we considered the one-dimensional mass critical NLS with even initial data of the form $u_0 = Q + \eta_0$, $\|\eta_0\|_{H^1} + \|x\eta_0\|_{L^2} \ll 1$, and showed that for an open set of initial perturbations η_0 the corresponding solution u blows up in finite time $T > 0$ with the following asymptotic behavior as $t \to T$:

$$u(t,x) = e^{i\mu(t)}\lambda^{\frac{1}{2}}(t)\big(Q\big(\lambda(t)x\big) + r\big(t,\lambda(t)x\big)\big), \quad \big\|r(t)\big\|_{H^1} \ll 1, \quad \big\|r(t)\big\|_{L^\infty} = o(1),$$
$$\lambda(t) = \left(\frac{\ln|\ln(T-t)|}{2\pi(T-t)}\right)^{1/2}\big(1+o(1)\big). \tag{3.5}$$

The existence of a stable blow-up regime with the log-log blow up rate (3.5) was predicted by numerical computations and formal arguments in a number of works, see, e.g., Landman, LeMesurier, Papanicolaou, Sulem, and Sulem [62,64], Smirnov and Fraiman [99], Sulem and Sulem [102] and the references therein. In the H^1 setting the log-log blow-up regime was studied in details by Merle and Raphaël. Assuming some coercivity property of the linearization around Q, they proved the following (see [66,73–75,93], and the references therein).

Theorem 3.3 (Merle, Raphaël). (i) *Any solution arising from initial data* (3.4) *and blowing up in finite time T admits a representation of the form*

$$u(t) = e^{i\mu(t)}\lambda^{\frac{d}{2}}(t)Q\big(\lambda(t)\big(\cdot - q(t)\big)\big) + u^* + o_{L^2}(1), \quad t \to T,$$

with $\lim_{t\to T}\lambda(t) = +\infty$, $\lim_{t\to T} q(t) = q^* \in \mathbb{R}^d$, *and* $u^* \in L^2(\mathbb{R}^d)$. *Furthermore, one of the following alternatives holds:*

- *either the blow-up rate $\lambda(t)$ satisfies the log-log law* (3.5) *and then the limiting profile u^* does not belong to*[4] H^1,
- *or*

$$\lambda(t) \geq \frac{c(u_0)}{T-t} \tag{3.6}$$

and then $u^* \in H^1$.

(ii) *The set of initial data satisfying* (3.4) *and such that the corresponding solution blows up in finite time in the log-log regime* (3.5) *is open in H^1 and contains the initial data* (3.4) *with* $E(u_0) \leq 0$.

The coercivity property required in Theorem 3.3 was proved in dimension 1 in [73] and checked numerically for $2 \leq d \leq 4$ in [30], and for $5 \leq d \leq 10$ in [113].

[4] More precisely, in this case one has $\int_{|x-q(t)|\leq R} |u^*(x)|^2 dx \sim \frac{C}{(\ln|\ln R|)^2}$, as $R \to 0$.

The log-log blow-up regime of Theorem 3.3 is known to remain stable under H^s perturbations of initial data for all $s > 0$, see Colliander and Raphaël [13]. An interesting open question is wether this stability persists in the L^2 setting.

The set of initial data (3.4) such that the corresponding solution blows up satisfying (3.6) is nonempty. A large class of solutions with the pseudoconformal blow-up rate $\|\nabla u(t)\|_{L^2} \sim \frac{1}{T-t}$ was constructed by Bourgain and Wang [7] in dimensions 1 and 2, starting from perturbations of the minimal blow-up solution (3.3) by smooth rapidly decaying limiting profiles u^* vanishing at zero to a sufficiently large order, see also Krieger and Schlag [58]. In [82] Merle, Raphaël, and Szeftel proved that the Bourgain–Wang solutions with slightly supercritical mass are unstable, belonging to the boundary of the H^1 open sets of global solutions that scatter both forward and backward in time, and solutions that blow up in finite time in the log-log regime. It is not known whether blow-up solutions with a blow-up rate strictly greater than the pseudoconformal rate exist in the regime (3.4). For larger masses, an example of such solutions was constructed by Martel and Raphaël [68] in the 2D case. The solutions of [68] are obtained by considering the interaction of K solitary waves concentrated at the vertices of a K-sided regular polygon and showing that this leads to a $KM(Q)$-mass solution blowing up at infinity with the rate $\|\nabla u(t)\|_{L^2} \sim \ln t$ as $t \to +\infty$, and thus, after applying the pseudoconformal transformation (2.3), to a solution that blows up as $t \to 0$ with the rate $\|\nabla u(t)\|_{L^2} \sim \frac{|\ln t|}{t}$.

We also refer to [66, 67] and references therein for the results on the near soliton blow-up dynamics for the mass-critical gKdV equation where the picture is more complete.

The regimes discussed above correspond to a single point finite time blow-up. The examples of multipoint blow-up solutions can be also constructed using as building blocks either the explicit blow-up solution (3.3) (see Merle [71]), or the log-log blow-up solutions of Theorem 3.3 (see Fan [28]). The general conjecture is that for any finite time blow-up solution, the singular set is given by a finite number of points, each point concentrating at least the mass of the ground state, see, e.g., [74].

4. MASS SUPERCRITICAL, ENERGY SUBCRITICAL NLS

In this section we discuss briefly the known blow-up regimes for the focusing NLS in the range $\frac{2}{d} < p < \frac{2}{d-2}$.

4.1. Self-similar blow-up

Numerical simulations and formal arguments (see, e.g., Sulem and Sulem [102] and the references therein) strongly suggest the existence of stable blow-up solutions of the following self-similar form:

$$u(t,x) \approx \frac{1}{(2b(T-t))^{\frac{1}{2p}}} e^{-i\frac{1}{2b}\ln(T-t)} V\left(\frac{x}{(2b(T-t))^{\frac{1}{2}}}\right), \quad b > 0. \tag{4.1}$$

Substituting this ansatz into the NLS equation leads to the following elliptic equation for the profile $V(y)$:

$$-\Delta V + V - ib\left(\frac{1}{p} + y \cdot \nabla\right)V - |V|^{2p}V = 0, \quad y \in \mathbb{R}^d. \tag{4.2}$$

It is expected that for a discrete set of values of b, this equation admits nontrivial zero-energy radial solutions, although these solutions fail to belong to $\dot{H}^{s_c}$ (in accordance with the growth of the $\dot{H}^{s_c}$ norm proved in [75]) due to a slow decay at infinity, $V(y) \sim \frac{C}{|y|^{\frac{1}{p}+\frac{i}{b}}}$, as $|y| \to \infty$. Thus, to obtain, say, H^1 solutions, one has to view (4.1) as a local approximation near the blow-up point ($x = 0$), and then to extend it to the region $|x| \gg \sqrt{T-t}$ by a well-localized time-independent profile, smooth away from the origin and behaving as $\frac{C}{|x|^{\frac{1}{p}+\frac{i}{b}}}$ near the origin.

Rigorous results justifying the above self-similar blow-up scenario are currently available only in the case $0 < s_c \ll 1$ where bifurcation-type arguments starting from the mass critical case can be used, see Merle, Raphaël, and Szeftel [81], Bahri, Martel, and Raphaël [2].

4.2. Standing sphere and contracting sphere blow-up solutions

In addition to the self-similar blow-up (4.1), which is expected to be generic, two other blow-up regimes are known for the mass supercritical NLS. The first is given by the so-called standing sphere blow-up solutions discovered by Raphaël [94] in the context of the two-dimensional quintic NLS and later on generalized to the quintic NLS in higher dimensions $d \geq 3$ by Raphaël and Szeftel [96]. Standing sphere blow-up solutions are radial, stable in their symmetry class solutions that blow up in finite time on a fixed sphere in the 1D log-log regime. The heuristic behind these solutions is that in the radial setting the quintic NLS takes the following form:

$$iu_t = -\partial_r^2 u - \frac{d-1}{r}\partial_r u - |u|^4 u, \quad r = |x|,$$

where for solutions concentrated near a fixed sphere $r = r_0 > 0$ the second term on the right-hand side can be viewed as a lower order term. Thus, one can expect the dynamics to be governed by the one-dimensional quintic NLS,

$$iu_t = -\partial_r^2 u - |u|^4 u,$$

for which one has a stable log-log blow-up regime. The above idea of reduction to a lower dimensional mass critical NLS was adapted to the 3D cylindrically symmetric cubic NLS by Holmer and Roudenko [43] and Zwiers [114], yielding the existence of finite time blow-up solutions concentrating on a fixed circle in the 2D log-log regime. These are the only known examples of blow-up solutions with a nontrivial blow-up set.

Another blow-up scenario occurs in the range $d \geq 2$, $\frac{2}{d} < p < 5$. In this case there exist radial solutions, called contracting sphere blow-up solutions, that blow up in finite time by concentration of the corresponding 1D ground state on a sphere of radius[5] $\sim t^{\frac{\alpha}{1+\alpha}}$ at the

[5] With blow-up time set to $t = 0$.

rate $\sim t^{\frac{1}{1+\alpha}}$ with $\alpha = \frac{2-p}{p(d-1)}$:

$$u(t,x) \approx e^{i\theta(t)+iv(t)r/2}\lambda(t)Q\big(\lambda(t)\big(r-q(t)\big)\big), \quad r = |x|,$$

where

$$q(t) \sim t^{\frac{\alpha}{\alpha+1}}, \quad \lambda(t) \sim t^{-\frac{1}{\alpha+1}}, \quad v(t) \sim t^{-\frac{1}{\alpha+1}}, \quad \theta \sim t^{\frac{\alpha-1}{\alpha+1}},$$

and $Q(y) = \frac{(p+1)^{\frac{1}{2p}}}{\cosh^{\frac{1}{p}}(py)}$. The contracting sphere blow-up was predicted numerically and heuristically in [29,42], see also the references therein. Rigorously, the existence of contracting sphere blow-up solutions was proved for the 3D cubic NLS in [41], and in the range $d \geq 2$, $0 < s_c < 1$, $p < 5$ by Merle, Raphaël, and Szeftel in [83].

Both the standing sphere and contracting sphere blow-up are L^2-concentration mechanisms in a contrast to the self-similar collapse where no mass concentration occurs.

5. TYPE II BLOW-UP IN THE ENERGY CRITICAL MODELS

In the last 15–20 years there have been significant developments in the study of the blow-up phenomenon for the energy critical equations and, more specifically, in the study of energy bounded blow-up solutions (the so-called type II blow-up), including their constructions and in some cases their classification. Below we review some of these developments.

5.1. Blow-up for Schrödinger maps from $\mathbb{R}^2$ to $\mathbb{S}^2$

Consider the Schrödinger flow for maps from $\mathbb{R}^2$ to $\mathbb{S}^2$:

$$\begin{aligned} &u_t = u \times \Delta u, \quad x = (x_1, x_2) \in \mathbb{R}^2, \quad t \in \mathbb{R}, \\ &u|_{t=0} = u_0, \end{aligned} \tag{5.1}$$

where $u(t,x) = (u_1(t,x), u_2(t,x), u_3(t,x)) \in \mathbb{S}^2 \subset \mathbb{R}^3$. Equation (5.1) is a special case of the Landau–Lifshitz equation

$$u_t = a_1 u \times \Delta u + a_2\big(\Delta u + |\nabla u|^2 u\big), \quad a_1 \in \mathbb{R}, \quad a_2 \geq 0, \tag{5.2}$$

arising in the theory of ferromagnetism. In the case $a_1 = 0$, $a_2 = 1$, one recovers the harmonic map heat flow.

Schrödinger map equation (5.1) conserves the energy

$$\mathcal{E}(u) = \frac{1}{2}\int_{\mathbb{R}^2} dx |\nabla u|^2. \tag{5.3}$$

The problem is energy critical since both the equation (5.1) and energy (5.3) are invariant with respect to the scaling $u(t,x) \to u(\lambda^2 t, \lambda x)$, $\lambda \in \mathbb{R}_+$.

The local/global well-posedness of (5.1) has been extensively studied. Local existence for smooth initial data goes back to Sulem, Sulem, and Bardos [103], see also McGahagan [70]. The case of small data of low regularity was studied in several works. Global existence for equivariant small energy initial data was established by Chang, Shatah, and

Uhlenbeck in [11]. We recall that a map $u : \mathbb{R}^2 \to \mathbb{S}^2 \subset \mathbb{R}^3$ is called equivariant if it has the form

$$u(x) = e^{m\theta R} v(r), \quad v : \mathbb{R}_+ \to \mathbb{S}^2 \subset \mathbb{R}^3, \tag{5.4}$$

for some $m \in \mathbb{Z}$. Here (r, θ) are the polar coordinates in $\mathbb{R}^2$, $x_1 + ix_2 = e^{i\theta} r$, and R is the generator of the horizontal rotations,

$$R = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

or equivalently, $Ru = \mathbf{k} \times u$, $\mathbf{k} = (0,0,1)$. The m-equivariance is preserved by the Schrödinger flow (5.1). Global existence and scattering for general small energy initial data was proved by Bejenaru, Ionescu, Kenig, and Tataru in [3]. For large data, such a result cannot hold because of the existence of a rich family of finite energy stationary solutions, i.e., harmonic maps. The lowest energy giving rise to a nontrivial harmonic map is 4π, the corresponding harmonic map is, up to the symmetries, the stereographic projection

$$\phi_1(x) = e^{\theta R} Q(r), \quad Q = \left(\frac{2r}{r^2+1}, 0, \frac{r^2-1}{r^2+1} \right). \tag{5.5}$$

The stereographic projection is a member of a family of equivariant harmonic maps ϕ_m, $m \in \mathbb{Z}^+$:

$$\begin{aligned} &\phi_m(x) = e^{m\theta R} Q^m(r), \quad Q^m = (h_1^m, 0, h_3^m) \in \mathbb{S}^2, \\ &h_1^m(r) = \frac{2r^m}{r^{2m}+1}, \quad h_3^m(r) = \frac{r^{2m}-1}{r^{2m}+1}. \end{aligned} \tag{5.6}$$

All these maps are minimizers of the energy in their homotopy class. Namely, one has

$$\deg \phi_m = m, \quad \mathcal{E}(\phi_m) = 4\pi m.$$

The general threshold conjecture is that global existence and scattering hold for all zero homotopy data with energies $\mathcal{E}(u) < 8\pi$. The corresponding result for the wave map equation (1.2) follows from the works of Sterbenz and Tataru [100], where the wave maps from $\mathbb{R}^2$ into general compact target manifold were considered, and it was shown that any smooth finite energy solution to the wave map equation is either global and scatters in a suitable sense or concentrates a nontrivial harmonic map at its maximal time of existence; see Lawrie and Oh [63]. Under symmetry reductions, the results of this type were obtained earlier, see, e.g., Struwe [101] and the references therein. To the best of author's knowledge, for the Schrödinger map equation (5.1) in full generality, the threshold conjecture is open. In the equivariant setting, global existence and scattering for initial data with $\mathcal{E}(u_0) < 4\pi$ was proved by Bejenaru, Ionescu, Kenig, and Tataru [4].

Local existence for the Scrödinger map equation (5.1) in the energy space in the case of nontrivial homotopy equivariant initial data with energies slightly above the energy of the ground state,

$$u_0 \in \Sigma_m, \quad \mathcal{E}(u_0) \le 4\pi m + \varepsilon^2, \quad 0 < \varepsilon \ll 1, \tag{5.7}$$

where $\Sigma_m = \{u(x) = e^{m\theta R} v(r) \in \mathbb{S}^2 \subset \mathbb{R}^3 : E(u) < \infty, v(0) = -v(\infty) = -\mathbf{k}\}$, was established by Gustafson, Kang, and Tsai [36]. The conservation of energy, together with the inequality [35]

$$\operatorname{dist}^2_{\dot{H}^1}(u, \mathcal{S}_m) \lesssim E(u) - 4\pi m, \quad \forall \quad u \in \Sigma_m,$$

where $\mathcal{S}_m = \{e^{\alpha R}\phi_m(\lambda x), \alpha \in \mathbb{R}, \lambda > 0\}$, ensures that the Schrödinger maps with initial data (5.7) have the form $u(t,x) = e^{\alpha(t)R}\phi_m(\lambda(t)x) + O_{\dot{H}^1}(\varepsilon)$, which reduces the problem to understanding the behavior of the functions $\lambda(t)$ and $\alpha(t)$. It was shown by Gustafson, Kang, and Tsai [36], as well as Gustafson, Nakanishi, and Tsai [37], that if $m \geq 3$, then the initial data (5.7) lead to global solutions that, for all t, remain $\dot{H}^1$-close to the initial soliton $e^{\alpha(0)R}\phi_m(\lambda(0)x)$ and, furthermore, scatter as $t \to \infty$ to a nearby member of the family $\mathcal{S}_m$ (in fact, in [37] this is proved for the general Landau–Lifshitz equation (5.2)). The paper [37] treats also the case of $m = 2$ for the harmonic map heat flow under further restriction $v_2 = 0$, showing that global existence persists in this case while the stability may fail: as $t \to +\infty$, the solutions still converge to the family $\mathcal{S}_2$ but the evolution along this family described by the parameter $\lambda(t)$ does not necessarily converge or stay close to a particular harmonic map, more complicated asymptotics of $\lambda(t)$ occur as well. It is natural to expect similar behavior for the 2-equivariant Schrödinger maps. The case of $m = 1$ was studied by Bejenaru and Tataru in [5] where it was proved that ϕ_1 is unstable in $\dot{H}^1$ but stable within its equivariant class in some smaller space that includes H^1.

The question of existence of finite time blow-up for both the wave and Schrödinger maps from $\mathbb{R}^2$ to $\mathbb{S}^2$ has been a long standing problem. The bubbling results mentioned above show that for the wave maps the only possible scenario for singularity formation is by concentration of a nontrivial harmonic map. The first rigorous confirmations of this scenario were obtained by Krieger, Schlag, and Tataru [60], Rodnianski–Sterbenz [97], and Raphaël– Rodnianski [95]. Krieger, Schlag, and Tataru considered the equivariant wave maps (that is, the wave maps of the form (5.4) with $v(t,r) = (\sin\varphi(t,r), 0, \cos\varphi(t,r)))$ of corotation index 1 and showed that for any $\nu > \frac{1}{2}$ there exist initial data arbitrary close to ϕ_1 in $\dot{H}^1$, leading to finite time blow-up solutions of the form $u(t,x) \approx \phi_1(\lambda(t)x)$ with $\lambda(t) = (T-t)^{-1-\nu}$. The specificity of these solutions is that they are of finite Sobolev regularity depending on the blow-up rate. Namely, one has $u \in \dot{H}^{1+\nu-}$. The construction of [60] was extended to the whole range $\nu > 0$ by Gao and Krieger [32] ($\nu \leq 0$ is precluded by the concentration results of [100, 101]). Similar results were obtained for the focusing energy critical wave equation in dimension 3 by Krieger, Schlag, and Tataru [59, 61], see also [22] for the case of more exotic scales, and Donninger–Krieger [23] for the case of infinite time blow-up. We also refer to Jendrej [45] for the construction of near ground state blow up solutions for the energy critical wave equation in dimension 5.

In a contrast to the Krieger–Schlag–Tataru solutions, the blow-up regimes exhibited in [95, 97] arise from C^∞ finite energy initial data and are characterized by some specific blow-up rates. Namely, the following was proved in [95]: for any $m \geq 1$, there exists a set of C^∞ m-equivariant initial data arbitrary close to ϕ_m in the energy space such that the corresponding solution blows up in finite time T and, as $t \to T$, has the form $u(t,x) \approx$

$\phi_m(\lambda(t)x)$ with

$$\lambda(t) = \begin{cases} (T-t)^{-1} e^{\sqrt{|\ln(T-t)|}+O(1)} & \text{if } m = 1, \\ c_m (T-t)^{-1} |\ln(T-t)|^{\frac{1}{2m-2}} (1+o(1)) & \text{if } m \geq 2, \end{cases} \quad \text{as } t \to T.$$

Furthermore, it was shown that these blow-up regimes are stable under smooth equivariant perturbations of the initial data. Similar results were obtained for the focusing energy critical nonlinear wave equation in dimension $d = 4$ by Hillairet and Raphaël [39].

Compared to [95], the construction of [60] gives no information on the stability/instability of the corresponding solutions. Recently, Krieger and Miao [54] proved that for ν small, these solutions are stable under sufficiently smooth corotational initial perturbations. Furthermore, Krieger, Miao, and Schlag [55] showed that this stability persists under nonequivariant smooth perturbations that vanish near the light cone. See also Burzio and Krieger [9] for the related results for the 3D energy critical nonlinear wave equation.

While for m-equivariant Schrödinger maps with $m \geq 3$ the possibility of blow-up near ϕ_m is excluded by the stability results of Gustafson, Kang, Nakanishi, and Tsai, for $m = 1$ near ϕ_1 blow-up does occur. This was proved by Merle, Raphaël, and Rodnianski [77].

Theorem 5.1 (Merle, Raphael, Rodnianski [77]). *There exists a set of C^∞ 1-equivariant initial data with elements arbitrary close to ϕ_1 in the energy space such that the corresponding solution to the Schrödinger map equation* (5.1) *blows up in finite time T and, as $t \to T$, one has*

$$u(t) = e^{\alpha(t)R}\phi_1(\lambda(t)\cdot) + u^* + o_{\dot{H}^1}(1),$$
$$\lambda(t) = c\frac{|\ln(T-t)|^2}{T-t}(1+o(1)), \quad \alpha(t) = \alpha_0(1+o(1)),$$

with some $u^* \in \dot{H}^1 \cap \dot{H}^2$, $\alpha_0 \in \mathbb{R}$, and $c > 0$.

In contrast to the wave map result of [95], the initial data in Theorem 5.1 form a set of codimension one.

In [91] we complemented the result of [77] by showing that (5.1) admits Krieger–Schlag–Tataru-type blow-up solutions as well. Namely, we proved:

Theorem 5.2 ([91]). *For any $\nu > 1$, $\alpha_0 \in \mathbb{R}$, there exist 1-equivariant initial data arbitrary close to ϕ_1 in $\dot{H}^1 \cap \dot{H}^3$ such that the corresponding solution to the Schrödinger map equation* (5.1) *blows up in finite time T and, as $t \to T$, one has*

$$u(t) = e^{\alpha(t)R}\phi_1(\lambda(t)\cdot) + u^* + o_{\dot{H}^1\cap\dot{H}^2}(1), \tag{5.8}$$

where $\lambda(t) = (T-t)^{-1/2-\nu}$, $\alpha(t) = \alpha_0 \ln(T-t)$, and $u^* \in H^{1+2\nu-}$. Furthermore, $u^*(x) = e^{\theta R}v^*(r)$, $v^* = (v_1^*, v_2^*, v_3^*)$, is compactly supported, C^∞ away from $x = 0$ and, as $|x| \to 0$, behaves as

$$v_1^*(r) + iv_2^*(r) \sim c_{\alpha,\nu} r^{2i\alpha_0+2\nu} \ln r.$$

In fact, the solutions constructed in [91] belong to $\dot{H}^{1+2\nu-}$. Observe that the regularity of the limiting profile u^* in (5.8) is also related to the blow-up rate (as in the case of the mass critical NLS, see Section 3.2).

As in [60,61], the proof of Theorem 5.2 relies on obtaining an approximate solution to an arbitrary high order $O((T-t)^N)$, which we construct using matching asymptotic expansions. We also refer to [31] for some closely related constructions in the parabolic setting. To convert the approximate solution into an exact solution, one then solves the problem for the small remainder backward in time with zero initial data at $t = T$. Once one can solve the equation up to any order, some very rough energy estimates are enough to control the remainder, in contrast to the approach of [77] that requires more advanced mixed energy/Morawetz estimates. Of course, a drawback of this procedure is that it gives no information on the stability of the constructed solutions.

Although we do not discuss the parabolic problems in this note, let us stress that as far as slow blow-up is concerned, there are a lot of direct connections between Schrödinger-type equations and their parabolic counterpart.

5.2. Energy critical NLS

In this subsection, we consider the energy critical focusing nonlinear Schrödinger equation

$$iu_t = -\Delta u - |u|^{\frac{4}{d-2}}u, \quad x \in \mathbb{R}^d, \quad d \geq 3, \tag{5.9}$$

restricting ourselves to the case of radial solutions

$$u|_{t=0} = u_0 \in \dot{H}^1_{\text{rad}}(\mathbb{R}^d). \tag{5.10}$$

Recall that this equation admits a family of stationary states $W_{\alpha,\lambda}(x) = e^{i\alpha}\lambda^{\frac{d-2}{2}}W(\lambda x)$, $\alpha \in \mathbb{R}$, $\lambda > 0$, with W given by (2.6). We denote by $\mathcal{S}$ the two-dimensional manifold of these solutions, $\mathcal{S} = \{W_{\alpha,\lambda}, \alpha \in \mathbb{R}, \lambda > 0\}$.

The dynamics for the energies below the ground state energy was classified by Kenig and Merle [48] for radial data in dimensions 3, 4, 5, and by Killip–Visan [52] ($d \geq 5$) and Dodson [18] ($d = 4$) for general initial data in dimension $d \geq 4$. The results of [48, 52] ensure that for $u_0 \in \dot{H}^1_{\text{rad}}(\mathbb{R}^d)$ with $E(u_0) < E(W)$ one has global existence and scattering if $\|\nabla u_0\|_{L^2} < \|\nabla W\|_{L^2}$, and finite time blow-up both forward and backward in time if $\|\nabla u_0\|_{L^2} > \|\nabla W\|_{L^2}$ and $u_0 \in L^2$.

A classification of radial solutions with critical energy $E(u_0) = E(W)$ was obtained by Duyckaerts and Merle [27] in dimensions 3, 4, 5, and by Li and Zhang [65] in dimension $d \geq 6$. In this case, in addition to scattering in both directions if $\|\nabla u_0\|_{L^2} < \|\nabla W\|_{L^2}$ and finite time blow-up in both directions if $\|\nabla u_0\|_{L^2} > \|\nabla W\|_{L^2}$ and $u_0 \in L^2$, there exist solutions that converge to W in one direction and scatter or blow-up in the opposite direction. More precisely, there exist unique, up to the symmetries, solutions W^-, W^+ that converge to W in $\dot{H}^1$ as $t \to +\infty$ satisfying $\|W^-\|_{L^2} < \|\nabla W\|_{L^2}$, $\|W^+\|_{L^2} > \|\nabla W\|_{L^2}$; W^- is global and scatters as $t \to -\infty$, and W^+ blows up in finite negative time, at least if $d \geq 5$.

The dynamics of the radial solutions with the energies slightly above $E(W)$ in the 3D case was studied by Nakanishi and Roy [86]. In continuation of the previous results of Nakanishi and Schlag [87,88] for the energy subcritical Klein–Gordon and Schrödinger equations and Krieger, Nakanishi, and Schlag [56] for the energy critical wave equation, Nakanishi and Roy proved that any radial $\dot{H}^1$ solution to (5.9) with $E(u) \leq E(W) + \varepsilon^2$, $\varepsilon \ll 1$, can stay $\dot{H}^1$-close to the ground state family $\mathcal{S}$ only on an interval of time, although it can be the entire lifespan. Once the solution leaves a neighborhood of $\mathcal{S}$, it either scatters or blows up (in the latter case, one has to assume in addition that $u_0 \in L^2$). Furthermore, all four combinations of scattering and blow-up forward/ backward in time occur for large sets of initial data. One might expect a similar result to hold in higher dimensions. The solutions that stay $\dot{H}^1$-close to $\mathcal{S}$ forward (backward) in time are expected to form a codimension one center-stable (center-unstable) manifold that divides a neighborhood of $\mathcal{S}$ into two regions exhibiting blow-up and scattering, respectively, forward (backward) in time (see Krieger, Nakanishi, and Schlag [57] for the corresponding result for the energy critical nonlinear wave equation). In low dimensions, the near ground state solutions can exhibit nontrivial dynamical behavior, including, along with scattering to the ground states, finite and infinite time type II blow-up. In dimension 3, the examples of infinite time near ground state blow-up at prescribed power law rate were constructed in [89]. Combining the linear analysis around W that we developed in [89] with the construction of approximate solutions of [91], one gets also $H^1_{\text{rad}}(\mathbb{R}^3) \cap \dot{H}^{1+\nu-}(\mathbb{R}^3)$ finite time blow-up solutions of the form

$$\begin{aligned} &u(t) = W_{\alpha(t),\lambda(t)} + u^* + o_{\dot{H}^1}(1), \quad \text{as } t \to 0, \\ &\lambda(t) = t^{-1/2-\nu}, \quad \alpha(t) = \alpha_0 \ln t, \quad \|u^*\|_{\dot{H}^1 \cap \dot{H}^{1+\nu-}} \ll 1, \quad \nu > 0, \quad \alpha_0 \in \mathbb{R}. \end{aligned}$$

Similar results can be proved for $d = 4$. As in the case of Schrödinger maps (5.1), these blow-up dynamics are closely related to the slow decay of the ground state in low dimensions and are expected to disappear starting from $d = 7$ (for $d = 5, 6$, finite or infinite near ground state concentration is still expected). Some partial results in this direction were obtained in [92], where we showed that for any $d \geq 7$, radial solutions staying in a neighborhood of $\mathcal{S}$ are global and scatter to a fixed ground state as soon as the linearization around W satisfies some suitable spectral assumptions that we were able to prove for d sufficiently large. In the parabolic setting, much more complete results are available. Namely, for the energy critical nonlinear heat equation in dimension $d \geq 7$, Collot, Merle, and Raphaël [15] obtained a complete classification of the dynamics for initial data $\dot{H}^1$ close to W (without radial symmetry assumption), showing that both the set of initial data leading to blow-up in the ODE type I regime and the set of initial data leading to global solutions that dissipate as $t \to +\infty$ are open in $\dot{H}^1$ and are separated by a codimension one set of global solutions that converge as $t \to +\infty$ to a fixed ground state.

5.3. Radial multibubble dynamics

In the breakthrough paper [25], Duyckaerts, Kenig, and Merle obtained a complete classification of radial, energy bounded solutions of the 3D focusing energy critical nonlinear wave equation:

$$\begin{aligned} u_{tt} &= \Delta u + u^5, \quad (t,x) \in \mathbb{R} \times \mathbb{R}^3, \\ (u, \partial_t u)|_{t=0} &= (u_0, u_1) \in \dot{H}^1_{\mathrm{rad}}(\mathbb{R}^3) \times L^2_{\mathrm{rad}}(\mathbb{R}^3), \end{aligned} \tag{5.11}$$

showing that they asymptotically decompose into a finite sum of scale separated ground states and a radiation term which solves the linear wave equation. More precisely, one has

Theorem 5.3 (Soliton resolution for (5.11), Duyckaerts, Kenig, Merle [25]). *Let $(u_0, u_1) \in \dot{H}^1_{\mathrm{rad}} \times L^2_{\mathrm{rad}}$ and $(u, \partial_t u) \in C(]T_-, T_+[, \dot{H}^1 \times L^2)$ be the corresponding maximal solution to (5.11). Then one of the following holds:*

(i) *Type I blow-up: $T_+ < +\infty$ and $\|(u(t), \partial_t u(t))\|_{\dot{H}^1 \times L^2} \xrightarrow{t \to T_+} \infty$.*

(ii) *Type II blow-up: $T_+ < +\infty$ and there exist $(v_0, v_1) \in \dot{H}^1 \times L^2$, an integer $J \in \mathbb{N} \setminus \{0\}$, and for each $1 \leq j \leq J$, a sign $\varepsilon_j \in \{-1, 1\}$ and a positive function $\lambda_j(t)$ defined for t close to T_+, verifying*

$$\frac{\lambda_j}{\lambda_{j+1}}(t) \xrightarrow{t \to T_+} \infty, \quad \forall 1 \leq j \leq J, \quad \lambda_{J+1}(t) = (T_+ - t)^{-1},$$

such that

$$u(t) = \sum_{j=1}^{J} \varepsilon_j \lambda_j^{1/2}(t) W\big(\lambda_j(t) \cdot\big) + v_0 + o_{\dot{H}^1}(1), \quad \partial_t u(t) = v_1 + o_{L^2}(1),$$

as $t \to T_+$.

(iii) *Global solution: $T_+ = +\infty$ and there exist a solution v_L of the linear wave equation, an integer $J \in \mathbb{N}$, and for each $1 \leq j \leq J$, a sign $\varepsilon_j \in \{-1, 1\}$ and a positive function $\lambda_j(t)$ defined for t sufficiently large, verifying*

$$\frac{\lambda_j}{\lambda_{j+1}}(t) \xrightarrow{t \to +\infty} \infty, \quad \forall 1 \leq j \leq J, \quad \lambda_{J+1}(t) = t^{-1},$$

such that

$$\begin{aligned} u(t) &= \sum_{j=1}^{J} \varepsilon_j \lambda_j^{1/2}(t) W\big(\lambda_j(t) \cdot\big) + v_L(t) + o_{\dot{H}^1}(1), \\ \partial_t u(t) &= \partial_t v_L(t) + o_{L^2}(1), \quad \text{as } t \to +\infty. \end{aligned}$$

Later similar results were proved for the radial energy critical NLW equation in all odd dimensions and in dimension 4, and for critical equivariant wave maps, see Duyckaerts, Kenig, and Merle [26], Duyckaerts, Kenig, Martel, and Merle [24], Jendrej and Lawrie [47], as well as the references therein.

In view of the above results, a natural question is to determine which type of configurations of solitons and radiation can really occur. A similar question can be asked the NLS equation. In dimensions $d = 3, 4, 5$, for both the radial energy critical wave and radial energy critical Schrödinger equations, no examples with $J \geq 2$ are known. For the energy critical wave equation in dimension $d \geq 6$ and for the energy critical Schrödinger equation in dimension $d \geq 7$, global ($T_+ = +\infty$) radial pure ($v_L = 0$) two-bubble solutions, with one bubble developing at scale 1 and the other concentrating at infinite time, were constructed by Jendrej [44,46].

5.4. Further generalizations

Blow-up by concentration of stationary states can also occur in the energy supercritical models. Among the known examples are the focusing energy supercritical NLS and NLW equations. The focusing energy supercritical NLS,

$$iu_t = -\Delta u - |u|^{2p}u, \quad x \in \mathbb{R}^d, \quad d \geq 3, \quad p > \frac{2}{d-2}, \tag{5.12}$$

has a two-parameter family of smooth radial stationary solutions $\varphi_{\alpha,\lambda}(x) = e^{i\alpha}\lambda^{\frac{1}{p}}\varphi(\lambda x)$, $\alpha \in \mathbb{R}$, $\lambda > 0$, where φ solves

$$\Delta\varphi + \varphi^{2p+1} = 0, \quad \varphi > 0, \quad \varphi(0) = 1,$$

and has the following behavior at infinity:

$$\varphi(x) \sim \frac{c_{p,d}}{|x|^{\frac{1}{p}}}, \quad c_{p,d}^{2p} = \frac{(d-2)p-1}{p^2}, \quad \text{as } |x| \to \infty.$$

Merle, Raphael, and Rodnianski [78] proved that in dimension $d \geq 11$, for generic integers p satisfying $p > p^*(d) = \frac{2}{d-4-2\sqrt{d-1}}$, equation (5.12) admits radial blow-up solutions of the form

$$u(t,x) \approx \lambda^{\frac{1}{p}}(t)\varphi\big(\lambda(t)x\big), \quad \lambda(t) \sim (T-t)^{-\kappa(p,d)l}, \quad l \in \mathbb{N} \setminus \{0\},$$

arising from C^∞ compactly supported initial data. Here $\kappa(p,d) > 0$ is an explicit constant. The H^s norms of these solutions remain bounded if $0 \leq s < s_c$, while the critical norm blows up logarithmically, $\|u\|_{H^{s_c}} \sim \sqrt{|\ln(T-t)|}$, as $t \to T$. The corresponding result for the energy supercritical nonlinear heat equation (in the same range of parameters and with the same sequence of blow-up rates) goes back to the work of Herrero and Velázquez [38], see also Mizoguchi [85]. The numerology $d \geq 11$, $p > p^*(d)$ is related to the stability properties of φ; for the radial energy supercritical heat equation, it is known (under some mild additional assumptions) that no type II blow-up occurs outside this range, see Matano– Merle [69] and the references therein. The analysis of [78] was extended to the energy supercritical wave equation by Collot [14]. One can also use the approach of [60,91] to construct near φ blow-up solutions of finite Sobolev regularity with a continuum of power-type blow-up rates for both energy supercritical NLS and NLW equations.

Another example that we would like to mention is the hyperbolic vanishing mean curvature flow in the Minkowski space $\mathbb{R}^{2n,1}$ that we considered in [1] in the case of $n = 4$. The minimal hypersurfaces in $\mathbb{R}^{2n}$ are stationary solutions of the corresponding Cauchy problem. It is known that $\mathbb{R}^8$ is foliated by a scaling-invariant family of smooth birotational invariant minimal hypersurfaces asymptotic at infinity to the Simons cone:

$$C_4 = \big\{(x_1, \dots, x_8) \in \mathbb{R}^8, x_1^2 + \dots + x_4^2 = x_5^2 + \dots + x_8^2\big\}.$$

In [1], we showed that this family of minimal hypersurfaces generates finite time blow-up for the hyperbolic vanishing mean curvature flow, again with a continuum of prescribed power-type blow-up rates. In the parabolic setting, that is, for the mean curvature flow, a similar result (but with a sequence of specific blow-up rates) was proved much earlier by Velázquez [110].

6. FINITE TIME BLOW-UP FOR THE ENERGY SUPERCRITICAL DEFOCUSING NLS

Consider the energy supercritical defocusing nonlinear Schrödinger equation

$$iu_t = -\Delta u + |u|^{2p}u, \quad x \in \mathbb{R}^d, \quad p > \frac{2}{d-2}, \quad d \geq 3. \tag{6.1}$$

The question whether finite time blow-up occurs for (6.1) remained completely open up to very recently. On the one hand, numerical simulations, global well-posedness results for the log-supercritical equations (see, e.g., Tao [107]), the nonexistence of soliton-like solutions, and the expected nonexistence of the self-similar blow-up supported the hypothesis of global well-posedness. On the other hand, in [108] Tao exhibited examples of the energy supercritical defocusing NLS systems for which finite time blow-up does occur.

A decisive breakthrough has been achieved recently by Merle, Raphaël, Rodnianski, and Szeftel [79] who showed that in dimension $5 \leq d \leq 9$ the energy supercritical NLS (6.1), at least for certain choices of p, admits finite time blow-up solutions arising from C^∞ well-localized initial data. The construction of [79] employs the hydrodynamic formulation of the NLS equation that relates (6.1) to a compressible Euler equation via the Madelung transform, $u = \sqrt{\rho}e^{i\varphi}$. In a companion paper [80], Merle, Raphaël, Rodnianski, and Szeftel proved that the underlying compressible Euler equation has a family of self-similar blow-up solutions with C^∞ profiles that well approximate the NLS dynamics and thus can be used to construct finite time blow-up solutions for (6.1). The smoothness of the Eulerian self-similar solutions plays an important role in the analysis of [79]. In contrast to the focusing energy supercritical blow-up regime discussed in Section 5.4, where all subcritical H^s norms remain bounded, the solutions constructed in [79] satisfy

$$\|u(t)\|_{H^s} \overset{t\to T}{\longrightarrow} +\infty, \quad \forall s > s^*, \tag{6.2}$$

for some $1 < s^* < s_c$, the growth in (6.2) being polynomial. The recent results of Bulut [8] indicate (6.2) as a general feature of the energy supercritical defocusing blow-up.

REFERENCES

[1] H. Bahouri, A. Marachli, and G. Perelman, Blow-up dynamics for the hyperbolic vanishing mean curvature flow of surfaces asymptotic to a Simons cone. *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 12, 3801–3887.

[2] Y. Bahri, Y. Martel, and P. Raphaël, Self-similar blow-up profiles for slightly supercritical nonlinear Schrödinger equations. *Ann. Henri Poincaré* **22** (2021), no. 5, 1701–1749.

[3] I. Bejenaru, A. Ionescu, C. Kenig, and D. Tataru, Global Schrödinger maps in dimensions $d \geq 2$: small data in the critical Sobolev spaces. *Ann. of Math. (2)* **173** (2011), no. 3, 1443–1506.

[4] I. Bejenaru, A. Ionescu, C. Kenig, and D. Tataru, Equivariant Schrödinger maps in two spatial dimensions. *Duke Math. J.* **162** (2013), no. 11, 1967–2025.

[5] I. Bejenaru and D. Tataru, Near soliton evolution for equivariant Schrödinger Maps in two spatial dimensions. *Mem. Amer. Math. Soc.* **228** (2014), no. 1069.

[6] J. Bourgain, Global wellposedness of defocusing critical nonlinear Schrödinger equation in the radial case. *J. Amer. Math. Soc.* **12** (1999), 145–171.

[7] J. Bourgain and W. Wang, Construction of blowup solutions for the nonlinear Schrödinger equation with critical nonlinearity. *Ann. Sc. Norm. Pisa Cl. Sci. (4)* **25** (1997), no. 1–2, 197–215.

[8] A. Bulut, Blow-up criteria below scaling for defocusing energy-supercritical NLS and quantitative global scattering bounds. 2020, arXiv:2001.05477.

[9] S. Burzio and J. Krieger, Type II blow up solutions with optimal stability properties for the critical focussing nonlinear wave equation on $\mathbb{R}^{3+1}$. 2018, arXiv:1709.06408. To appear in *Mem. Amer. Math. Soc.*

[10] Th. Cazenave, *Semilinear Schrödinger equations*. Courant Lect. Notes Math. 10, American Mathematical Society, Courant Institute of Mathematical Sciences, 2003.

[11] N-H. Chang, J. Shatah, and K. Uhlenbeck, Schrödinger maps. *Comm. Pure Appl. Math.* **53** (2000), no. 5, 590–602.

[12] J. Colliander, M. Keel, G. Staffilani, H. Takaoka, and T. Tao, Global well-posedness and scattering for the energy-critical nonlinear Schrödinger equation in $\mathbb{R}^3$. *Ann. of Math. (2)* **167** (2008), no. 3, 767–865.

[13] J. Colliander and P. Raphaël, Rough blowup solutions to the L^2 critical NLS. *Math. Ann.* **345** (2009), no. 2, 307–366.

[14] C. Collot, Type II blow up manifolds for the energy supercritical wave equation. *Mem. Amer. Math. Soc.* **252** (2018), no. 1205.

[15] C. Collot, F. Merle, and P. Raphal, Dynamics near the ground state for the energy critical nonlinear heat equation in large dimensions. *Comm. Math. Phys.* **352** (2017), no. 1, 215–285.

[16] B. Dodson, Global well-posedness and scattering for the mass critical nonlinear Schrodinger equation with mass below the mass of the ground state. *Adv. Math.* **285** (2015), 1589–1618.

[17] B. Dodson, Global well-posedness and scattering for the defocusing, L^2-critical, nonlinear Schrödinger equation when $d = 2$. *Duke Math. J.* **165** (2016), no. 18, 3435–3516.

[18] B. Dodson, Global well-posedness and scattering for the focusing, cubic Schrödinger equation in dimension $d = 4$. *Ann. Sci. Éc. Norm. Supér. (4)* **52** (2019), no. 1, 139–180.

[19] B. Dodson, Global well-posedness for the defocusing, cubic nonlinear Schrödinger equation with initial data in a critical space. 2020, arXiv:2004.09618.

[20] B. Dodson, A determination of the blowup solutions to the focusing NLS with mass equal to the mass of the soliton. 2021, arXiv:2106.02723.

[21] B. Dodson, A determination of the blowup solutions to the focusing, quintic NLS with mass equal to the mass of the soliton. 2021, arXiv:2104.11690.

[22] R. Donninger, M. Huang, J. Krieger, and W. Schlag, Exotic blowup solutions for the u^5 focusing wave equation in $\mathbb{R}^3$. *Michigan Math. J.* **63** (2014), no. 3, 451–501.

[23] R. Donninger and J. Krieger, Nonscattering solutions and blowup at infinity for the critical wave equation. *Math. Ann.* **357** (2013), no. 1, 89–163.

[24] Th. Duyckaerts, C. Kenig, Y. Martel, and F. Merle, Soliton resolution for critical co-rotational wave maps and radial cubic wave equation. 2021, arXiv:2103.01293.

[25] Th. Duyckaerts, C. Kenig, and F. Merle, Classification of radial solutions of the focusing, energy-critical wave equation. *Cambridge J. Math.* **1** (2013), no. 1, 75–144.

[26] Th. Duyckaerts, C. Kenig, and F. Merle, Soliton resolution for the radial critical wave equation in all odd space dimensions. 2019, arXiv:1912.07664.

[27] Th. Duyckaerts and F. Merle, Dynamic of threshold solutions for energy-critical NLS. *Geom. Funct. Anal.* **18** (2009), no. 6, 1787–1840.

[28] C. Fan, Log-log blow up solutions blow up at exactly m points. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **34** (2017), no. 6, 1429–1482.

[29] G. Fibich, N. Gavish, and X-P. Wang, Singular ring solutions of critical and supercritical nonlinear Schrödinger equations. *Phys. D* **231** (2007), 55–86.

[30] G. Fibich, F. Merle, and P. Raphaël, Proof of a spectral property related to the singularity formation for the L^2 critical nonlinear Schrödinger equation. *Phys. D* **220** (2006), no. 1, 1–13.

[31] S. Filippas, M. A. Herrero, and J. J. L. Velázquez, Fast blow-up mechanisms for sign-changing solutions of a semilinear parabolic equation with critical nonlinearity. *Proc. R. Soc. Lond. Ser. A* **456** (2000), no. 2004, 2957–2982.

[32] C. Gao and J. Krieger, Optimal polynomial blow up range for critical wave maps. *Commun. Pure Appl. Anal.* **14** (2015), no. 5, 1705–1741.

[33] R. T. Glassey, On the blowing up of solutions of the Cauchy Problem for nonlinear Schrödinger equations. *J. Math. Phys.* **18** (1977), no. 9, 1794–1797.

[34] G. Grillakis, On nonlinear Schrödinger equations. *Comm. Partial Differential Equations* **25** (2000), 1827–1844.

[35] S. Gustafson, K. Kang, and T-P. Tsai, Schrödinger flow near harmonic maps. *Comm. Pure Appl. Math.* **60** (2007), no. 4, 463–499.

[36] S. Gustafson, K. Kang, and T-P. Tsai, Asymptotic stability of harmonic maps under the Schrödinger flow. *Duke Math. J.* **145** (2008), no. 3, 537–583.

[37] S. Gustafson, K. Nakanishi, and T-P. Tsai, Asymptotic stability, concentration and oscillations in harmonic map heat flow, Landau–Lifschitz and Schrödinger maps on $\mathbb{R}^2$. *Comm. Math. Phys.* **300** (2010), no. 1, 205–242.

[38] M. A. Herrero and J. J. L. Velázquez, Explosion des solutions d'équations paraboliques semilinéaires supercritiques. *C. R. Acad. Sci., Sér. 1 Math.* **319** (1994), 141–145.

[39] M. Hillairet and P. Raphäel, Smooth type II blow up solutions to the four-dimensional energy critical wave equation. *Anal. PDE* **5** (2012), 777–829.

[40] T. Hmidi and S. Keraani, Blowup theory for the critical nonlinear Schrödinger equations revisited. *Int. Math. Res. Not.* **46** (2005), 2815–2828.

[41] J. Holmer, G. Perelman, and S. Roudenko, A solution to the focusing 3d NLS that blows up on a contracting sphere. *Trans. Amer. Math. Soc.* **367** (2015), no. 6, 3847–3872.

[42] J. Holmer and S. Roudenko, On blow-up solutions to the 3D cubic nonlinear Schrödinger equation. *Appl. Math. Res. Express. AMRX* Art. ID abm004 (2007), 31 pp.

[43] J. Holmer and S. Roudenko, A class of solutions to the 3D cubic nonlinear Schrödinger equation that blows up on a circle. *Appl. Math. Res. Express. AMRX* **1** (2011), 23–94.

[44] J. Jendrej, Construction of two-bubble solutions for the energy-critical NLS. *Anal. PDE* **10** (2017), no. 8, 1923–1959.

[45] J. Jendrej, Construction of type II blow-up solutions for the energy-critical wave equation in dimension 5. *J. Funct. Anal.* **272** (2017), 866–917.

[46] J. Jendrej, Construction of two-bubble solutions for energy-critical wave equations. *Amer. J. Math.* **141** (2019), no. 1, 55–118.

[47] J. Jendrej and A. Lawrie, Soliton resolution for equivariant wave maps. 2021, arXiv:2106.10738.

[48] C. Kenig and F. Merle, Global well-posedness, scattering and blow-up for the energy-critical, focusing, non-linear Schrödinger equation in the radial case. *Invent. Math.* **166** (2006), no. 3, 645–675.

[49] C. Kenig and F. Merle, Scattering for $H^{1/2}$ bounded solutions to the cubic, defocusing NLS in 3 dimensions. *Trans. Amer. Math. Soc.* **362** (2010), no. 4, 1937–1962.

[50] R. Killip, T. Tao, and M. Visan, The cubic nonlinear Schrödinger equation in two dimensions with radial data. *J. Eur. Math. Soc. (JEMS)* **11** (2009), no. 6, 1203–1258.

[51] R. Killip and M. Visan, Energy-supercritical NLS: critical $\dot{H}^s$-bounds imply scattering. *Comm. Partial Differential Equations* **35** (2010), no. 6, 945–987.

[52] R. Killip and M. Visan, The focusing energy-critical nonlinear Schrödinger equation in dimensions five and higher. *Amer. J. Math.* **132** (2010), no. 2, 361–424.

[53] R. Killip, M. Visan, and X. Zhang, The mass-critical nonlinear Schrödinger equation with radial data in dimensions three and higher. *Anal. PDE* **1** (2008), no. 2, 229–266.

[54] J. Krieger and S. Miao, On the stability of blowup solutions for the critical corotational wave-map problem. *Duke Math. J.* **169** (2020), no. 3, 435–532.

[55] J. Krieger, S. Miao, and W. Schlag, A stability theory beyond the co-rotational setting for critical Wave Maps blow up. 2020, arXiv:2009.08843.

[56] J. Krieger, K. Nakanishi, and W. Schlag, Global dynamics away from the ground state for the energy-critical nonlinear wave equation. *Amer. J. Math.* **135** (2013), no. 4, 935–965.

[57] J. Krieger, K. Nakanishi, and W. Schlag, Center-stable manifold of the ground state in the energy space for the critical wave equation. *Math. Ann.* **361** (2015), no. 1–2, 1–50.

[58] J. Krieger and W. Schlag, Non-generic blow-up solutions for the critical focusing NLS in 1-D. *J. Eur. Math. Soc. (JEMS)* **11** (2009), no. 1, 1–125.

[59] J. Krieger and W. Schlag, Full range of blow up exponents for the quintic wave equation in three dimensions. *J. Math. Pures Appl.* **9** (2014), 873–900.

[60] J. Krieger, W. Schlag, and D. Tataru, Renormalization and blow up for charge one equivariant critical wave maps. *Invent. Math.* **171** (2008), 543–615.

[61] J. Krieger, W. Schlag, and D. Tataru, Slow blow up solutions for the $H^1(\mathbb{R}^3)$ critical focusing semi-linear wave equation. *Duke Math. J.* **147** (2009), 1–53.

[62] M. J. Landman, G. C. Papanicolaou, C. Sulem, and P. L. Sulem, Rate of blowup for solutions of the nonlinear Schrödinger equation at critical dimension. *Phys. Rev. A (3)* **38** (1988), no. 8, 3837–3843.

[63] A. Lawrie and S.-J. Oh, A refined threshold theorem for $(1 + 2)$-dimensional wave maps into surfaces. *Comm. Math. Phys.* **342** (2016), no. 3, 989–999.

[64] B. J. LeMesurier, G. Papanicolaou, C. Sulem, and P. Sulem, Local structure of the self-focusing singularity of the nonlinear Schrödinger equation. *Phys. D* **32** (1988), 210–226.

[65] D. Li and X. Zhang, Dynamics for the energy critical nonlinear Schrödinger equation in high dimensions. *J. Funct. Anal.* **256** (2009), 1928–1961.

[66] I. Martel, F. Merle, P. Raphaël, and J. Szeftel, Near soliton dynamics and the formation of singularities in L^2-critical problems. *Russian Math. Surveys* **69** (2014), no. 2, 261–290.

[67] Y. Martel and D. Pilod, Finite point blowup for the critical generalized Korteweg–de Vries equation. 2021, arXiv:2107.00268.

[68] Y. Martel and P. Raphaël, Strongly interacting blow up bubbles for the mass critical nonlinear Schrödinger equation. *Ann. Sci. Éc. Norm. Supér. (4)* **51** (2018), no. 3, 701–737.

[69] H. Matano and F. Merle, Classification of type I and type II behaviors for a supercritical nonlinear heat equation. *J. Funct. Anal.* **256** (2009), no. 4, 992–1064.

[70] H. McGahagan, An approximation scheme for Schrödinger maps. *Comm. Partial Differential Equations* **32** (2007), 375–440.

[71] F. Merle, Construction of solutions with exactly k blow-up points for the Schrödinger equation with critical nonlinearity. *Comm. Math. Phys.* **129** (1990), no. 2, 223–240.

[72] F. Merle, Determination of blow-up solutions with minimal mass for nonlinear Schrödinger equations with critical power. *Duke Math. J.* **69** (1993), no. 2, 427–454.

[73] F. Merle and P. Raphaël, The blow up dynamics and upper bound on the blow up rate for the critical nonlinear Schrödinger equation. *Ann. of Math.* **161** (2005), 157–222.

[74] F. Merle and P. Raphaël, Profiles and quantization of the blow up mass for critical nonlinear Schrödinger equation. *Comm. Math. Phys.* **253** (2005), 675–704.

[75] F. Merle and P. Raphaël, On a sharp lower bound on the blow-up rate for the L^2 critical nonlinear Schrödinger equation. *J. Amer. Math. Soc.* **19** (2006), 37–90.

[76] F. Merle and P. Raphaël, Blow up of the critical norm for some radial L^2 supercritical nonlinear Schrödinger equations. *Amer. J. Math.* **130** (2008), no. 4, 945–978.

[77] F. Merle, P. Raphaël, and I. Rodnianski, Blow up dynamics of smooth equivariant solutions to the energy critical Schrödinger map. *Invent. Math.* **193** (2013), 3249–365.

[78] F. Merle, P. Raphaël, and I. Rodnianski, Type II blow up for the energy supercritical NLS. *Cambridge J. Math.* **352** (2015), 439–617.

[79] F. Merle, P. Raphaël, I. Rodnianski, and J. Szeftel, On blow up for the energy super critical defocusing nonlinear Schrödinger equations. 2019, arXiv:1912.11005. To appear in Invent. Math.

[80] F. Merle, P. Raphaël, I. Rodnianski, and J. Szeftel, On smooth self similar solutions to the compressible Euler equations. 2019, arXiv:1912.10998.

[81] F. Merle, P. Raphaël, and J. Szeftel, Stable self similar blow dynamics for slightly L^2 supercritical NLS equations. *Geom. Funct. Anal.* **20** (2010), no. 4, 1028–1071.

[82] F. Merle, P. Raphaël, and J. Szeftel, Instability of Bourgain–Wang solutions for the L^2 critical NLS. *Amer. J. Math.* **135** (2013), no. 4, 967–1017.

[83] F. Merle, P. Raphaël, and J. Szeftel, Collapsing ring blow up solutions to the L^2 supercritical NLS. *Duke Math. J.* **163** (2014), no. 2, 369–431.

[84] C. Miao, J. Murphy, and J. Zheng, The defocusing energy-supercritical NLS in four space dimensions. *J. Funct. Anal.* **267** (2014), no. 6, 1662–1724.

[85] N. Mizoguchi, Rate of type II blowup for a semilinear heat equation. *Math. Ann.* **339** (2007), no. 4, 839–877.

[86] K. Nakanishi and T. Roy, Global dynamics above the ground state for the energy-critical Schrödinger equation with radial data. *Commun. Pure Appl. Anal.* **15** (2016), no. 6, 2023–2058.

[87] K. Nakanishi and W. Schlag, Global dynamics above the ground state energy for the cubic NLS equation in 3D. *Calc. Var. Partial Differential Equations* **44** (2012), no. 1–2, 1–45.

[88] K. Nakanishi and W. Schlag, Global dynamics above the ground state for the nonlinear Klein–Gordon equation without a radial assumption. *Arch. Ration. Mech. Anal.* **203** (2012), no. 3, 809–851.

[89] C. Ortoleva and G. Perelman, Non-dispersive vanishing and blow up at infinity for the energy critical nonlinear Schrödinger equation in $\mathbb{R}^3$. *St. Petersburg Math. J.* **25** (2014), no. 2, 271–294.

[90] G. Perelman, On the blow-up phenomenon for the critical nonlinear Schrödinger equation in 1D. *Ann. Henri Poincaré* **2** (2001), 605–673.

[91] G. Perelman, Blow up dynamics for equivariant critical Schrödinger maps. *Comm. Math. Phys.* **330** (2014), 69–105.

[92] G. Perelman, Near ground state dynamics for the energy critical NLS. In preparation.

[93] P. Raphaël, Stability of the log-log bound for blow up solutions to the critical non linear Schrödinger equation. *Math. Ann.* **331** (2005), 577–609.

[94] P. Raphaël, Existence and stability of a solution blowing up on a sphere for a L^2 supercritical nonlinear Schrödinger equation. *Duke Math. J.* **134** (2006), no. 2, 199–258.

[95] P. Raphaël and I. Rodnianski, Stable blow up dynamics for the critical corotational wave maps and equivariant Yang–Mills problems. *Publ. Math. Inst. Hautes Études Sci.* **115** (2012), 1–122.

[96] P. Raphaël and J. Szeftel, Standing ring blow up solutions to the N-dimensional quintic nonlinear Schrödinger equation. *Comm. Math. Phys.* **290** (2009), no. 3, 973–996.

[97] I. Rodnianski and J. Sterbenz, On the formation of singularities in the critical $O(3)$ σ-model. *Ann. of Math. (2)* **172** (2010), no. 1, 187–242.

[98] E. Ryckman and M. Visan, Global well-posedness and scattering for the defocusing energy-critical nonlinear Schrödinger equation in $\mathbb{R}^{1+4}$. *Amer. J. Math.* **129** (2007), no. 1, 1–60.

[99] A. I. Smirnov and G. M. Fraiman, The interaction representation in the self-focusing theory. *Phys. D* **51** (1991), 2–15.

[100] J. Sterbenz and D. Tataru, Regularity of wave maps in dimension 2 +1. *Comm. Math. Phys.* **298** (2010), no. 1, 231–264.

[101] M. Struwe, Equivariant wave maps in two space dimensions. *Comm. Pure Appl. Math.* **56** (2003), no. 7, 815–823.

[102] C. Sulem and P. L. Sulem, *The nonlinear Schrödinger equation. Self-focusing and wave collapse*. Appl. Math. Sci. 139, Springer, New York, 1999.

[103] P. L. Sulem, C. Sulem, and C. Bardos, On the continuous limit for a system of continuous spins. *Comm. Math. Phys.* **107** (1986), no. 3, 431–454.

[104] T. Tao, Global well-posedness and scattering for the higher-dimensional energy-critical non-linear Schrödinger equation for radial data. *New York J. Math.* **11** (2005), 57–80.

[105] T. Tao, *Nonlinear dispersive equations*. CBMS Reg. Conf. Ser. Math. 106, American Mathematical Society, Providence, RI, 2006.

[106] T. Tao, A (concentration-)compact attractor for high-dimensional non-linear Schrödinger equations. *Dyn. Partial Differ. Equ.* **4** (2007), no. 1, 1–53.

[107] T. Tao, Global regularity for a logarithmically supercritical defocusing nonlinear wave equation for spherically symmetric data. *J. Hyperbolic Differ. Equ.* **4** (2007), no. 2, 259–265.

[108] T. Tao, Finite time blowup for a supercritical defocusing nonlinear Schrödinger system. *Anal. PDE* **11** (2018), no. 2, 383–438.

[109] T. Tao, M. Visan, and X. Zhang, Global well-posedness and scattering for the defocusing mass-critical nonlinear Schrödinger equation for radial data in high dimensions. *Duke Math. J.* **140** (2007), no. 1, 165–202.

[110] J. J. L. Velázquez, Curvature blow-up in perturbations of minimal cones evolving by mean curvature flow. *Ann. Sc. Norm. Super. Pisa* **21** (1994), 595–628.

[111] M. Visan, The defocusing energy-critical nonlinear Schrödinger equation in higher dimensions. *Duke Math. J.* **138** (2007), no. 2, 281–374.

[112] M.-I. Weinstein, Nonlinear Schrödinger equations and sharp interpolation estimates. *Comm. Math. Phys.* **87** (1983), 567–576.

[113] K. Yang, S. Roudenko, and Y. Zhao, Blow-up dynamics and spectral property in the L^2-critical nonlinear Schrödinger equation in high dimensions. *Nonlinearity* **31** (2018), no. 9, 4354–4392.

[114] L. Zwiers, Standing ring blowup solutions for cubic nonlinear Schrödinger equations. *Anal. PDE* **4** (2011), no. 5, 677–727.

GALINA PERELMAN

Laboratoire d'Analyse et de Mathématiques Appliquées UMR 8050, Université
Paris-Est Créteil, 61 avenue du Général de Gaulle, 94010 Créteil Cedex, France,
galina.perelman@u-pec.fr

ON THE ASYMPTOTICS FOR MINIMIZERS OF DONALDSON FUNCTIONAL IN TEICHMÜLLER THEORY

GABRIELLA TARANTELLO

ABSTRACT

We discuss the asymptotic behavior of minimizers for a Donaldson functional of interest in Teichmüller theory. For example, such minimizers allow one to parametrize the moduli space of constant mean curvature immersions of a closed surface S of genus $\mathfrak{g} \geq 2$ into a 3-manifold with sectional curvature -1, by elements of the tangent bundle of the Teichmüller space of S. The minimizers are governed by a system of PDEs which include a Gauss equation of Liouville type and a holomorphic κ-differential.

In our asymptotic analysis, we face the difficulty to describe the possible blow-up behavior of minimizers, especially when it occurs at a point where different zeroes of the holomorphic κ-differential coalesce. Therefore, we need to pursue accurate estimates of the blow-up profile of solutions for Liouville type equations, in the "collapsing" case.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 35B; Secondary 32G15, 35J20, 35B40, 35B44, 53C42

KEYWORDS

Minimal immersions, asymptotic CMC immersions, blow-up analysis, Liouville equations

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3880–3901
DOI 10.4171/ICM2022/91

1. INTRODUCTION

In this note we discuss the asymptotics for minimizers of a Donaldson-type functional whose relevance in Teichmüller theory was pointed out in [29] and [33]. Such minimizers are governed by the system of equations (1.3) below, which includes a Liouville-type equation (as a Gauss consistency condition) and a holomorphic κ-differential, $\kappa \geq 2$, over a closed surface of genus $\mathfrak{g} \geq 2$.

Since such a holomorphic κ-differential admits $2\kappa(\mathfrak{g}-1)$ zeroes counted with multiplicity, to pursue the asymptotics of such minimizers, we must keep handy the detailed blow-up analysis and profile estimates developed for solutions of Liouville-type equations involving a weight function with a finite number of zeroes of integral multiplicity (see [3, 10, 11, 67]).

We recall that Liouville-type equations arise in many contexts of interest in mathematics and physics, and have attracted much attention after their first encounter by Liouville in his model of Field Theory. Since then, a rich literature is now available revealing the many facets of Liouville-type equations and the crucial role they played towards a successful development of Liouville Field Theory, see [54].

In [47], Liouville furnished a local formula for solutions of Liouville equations in terms of a meromorphic complex function, the so-called "developing map." In this way, Liouville equations were introduced into the realm of Complex Analysis and Algebraic Geometry. In fact, exploring the solvability of Liouville equations has led to tackle many fundamental issues about modular functions and forms, normal families, Fuchsian, Lamé, and Painlevé equations, and about various moduli spaces, see [7, 13–18, 28, 40, 42] and the references therein.

One may focus also on Liouville equations involving Dirac measures, whose poles replace the role of the zeroes. Indeed, the poles will correspond to the zeroes of the weight function which appears in the equation governing the "regular" part of the solution.

In (bidimensional) abelian Gauge Field Theory, at a self-dual regime, we have that vortex configurations are governed by the Bogomolny equations. They involve a (gauge invariant) Cauchy–Riemann equation for the (complex valued) Higgs field. Thus, the (gauge independent) zeroes of the Higgs field are isolated with integral multiplicity and identify the so-called "vortex-points." As a consequence, around a vortex-point we can confirm the "quantization" properties for the electric and magnetic fields, as already observed experimentally (e.g., in superconductivity).

Taubes showed how to express Bogomolny's self-dual equations in the form of Liouville-type equations with Dirac measures supported exactly at the vortex points, see [34]. By virtue of Taubes' approach, it has been possible to obtain a rigorous description of self-dual vortices for various models proposed in the context of Maxwell–Chern–Simons–Higgs theory, Electroweak theory, Comics strings, etc. We refer to the monographs [59, 68] for details.

We mention that, the analytical construction of physically meaningful vortices has motivated the accurate blow-up analysis and profile estimates for solutions of (singular) Liouville equations, contained in [8–11, 38].

From the geometrical side, such an analysis has helped also tackle the classical "uniformization" problem of surfaces with conical singularities prescribed along a given "divisor." In this direction, the most delicate situation occurs when the prescribed conical angle is bigger than 2π. For smaller angles, a complete description of conical metrics with constant Gauss curvature is contained in [48, 62, 63]. On the other hand, for the standard 2-sphere $S^2 = \mathbb{C} \cup \{\infty\}$, it is yet not clear when a spherical metric with prescribed conical singularities and relative angles (bigger than 2π) exists. Clearly, beside the constraint dictated by the Gauss–Bonnet theorem, there are other less obvious obstructions to prevent the existence of such (spherical) metrics. For example, in the case of two singularities, only the "American" football is possible, where both conical angles must coincide.

There is a rich literature concerning spherical metrics on S^2 (see, e.g., [12, 19, 22–27, 49–51, 57, 65, 69, 70] and the references therein), where different points of view have been adopted and yielded to interesting (partial) results. Only recently Mondello–Panov [52, 53] have identified (almost) sharp necessary and sufficient conditions on the conical angles so that a corresponding spherical metric exists. The sharp results in [52, 53] are established using strategies and techniques developed in algebraic geometry. At the moment, such results seem out of reach by mere analytical techniques. On the other hand, a blow-up approach to solutions of the singular Liouville equation over the flat torus has permitted to reveal surprising results, where nonexistence or (sharp) existence results may hold, according to the "geometry" of the periodic cell domain. Thus, for example, a flat torus with a *square* lattice and a single singularity with conical angle 4π cannot admit a metric with constant Gauss curvature, while this is possible for a *rhombus* lattice, see [42]. Many other surprising phenomena have been identified for the moduli space of tori and their metrics with conical singularities and constant Gauss curvature, see [15] and [42].

For those and other reasons, it has emerged the need to describe what happens when singularities (i.e., vortex or conical points) coalesce into a single point. Naturally, such an investigation furnishes a better grasp about the uniformization problem, see [51]. But also it helps in the understanding of *non-abelian* self-dual vortices which are described in terms of *systems* of Liouville equations (see [4, 36, 37, 43–46]). Indeed, it is difficult to have a firm grasp about the blow-up of solutions for systems, especially when various components blow-up at the same point, but with different blow-up rates. In such a situation, "concentration" phenomena introduce terms in the equations which behave as Dirac measures whose poles, however, may "collapse" together, see [35, 41].

We encounter an analogous "collapsing" issue in the asymptotic description of *minimizers* for the Donaldson functional, considered in [29] and [33]. Such a functional is inspired by [20, 21, 56], and relates to the representation of the fundamental group of a closed surface into various character varieties, or to the parametrization of the moduli space of minimal or constant mean curvature (CMC) immersions into a 3-manifold with constant sectional curvature -1, see [29, 33].

To be more precise, for a given oriented closed surface S with genus $\mathfrak{g} \geq 2$, we denote by $\mathcal{T}_{\mathfrak{g}}(S)$ the Teichmüller space of conformal structures on S, modulo biholomorphisms in the homotopy class of the identity.

For minimal immersions, Uhlenbeck in [64] proposed a parametrization of the corresponding moduli space in terms of elements of the cotangent bundle of $\mathcal{T}_{\mathfrak{g}}(S)$, described by pairs $(X,\alpha) \in \mathcal{T}_{\mathfrak{g}}(S) \times C_2(X)$, where $C_2(X)$ is the space of holomorphic quadratic differentials on X. In this way, minimal immersions are sought with assigned second fundamental form $\mathrm{II} = \mathrm{Re}(\alpha)$, simply by solving the Gauss equation of Liouville type for the conformal factor of the pullback metric on X from the minimal immersion. However, as discussed in [31,32], such an immersion may not exist, or when it exists, it may not be unique (see also [30]). So, by this approach, one does not obtain a one-to-one correspondence between a minimal immersion and the pair (X,α).

On the contrary, as we shall see below, we have a better chance when we choose to parametrize minimal or (CMC) immersions of S in terms of elements of the tangent bundle $\mathcal{T}_{\mathfrak{g}}(S)$.

To this purpose, for given $X \in \mathcal{T}_{\mathfrak{g}}(S)$, we let $T_X^{1,0}$ denote the holomorphic tangent bundle of X and define $E = \otimes^{\kappa-1} T_X^{1,0}$ with $\kappa \geq 2$. Moreover, letting $A^0(E)$ be the space of smooth sections of E and $A^{0,1}(X,E)$ the space of $(0,1)$-forms of X valued on E, we consider the $(0,1)$-Dolbeault cohomology group $\mathcal{H}^{0,1}(X,E) = A^{0,1}(X,E)/\bar{\partial}(A^0(E))$, where $\bar{\partial} : A^0(E) \to A^{0,1}(X,E)$ is the d-bar operator.

Using the Hodge star operator $*_E : A^{0,1}(X,E) \to A^{1,0}(X,E^*)$ and Serre duality theorem, we know that $C_\kappa(X)$, the space of holomophic κ-differential on X, satisfyies:

$$C_\kappa(X) \simeq (\mathcal{H}^{0,1}(X,E))^*,$$

see ([66]). Therefore, for $\kappa = 2$, we can use the pair $(X,[\beta]) \in \mathcal{T}_{\mathfrak{g}}(S) \times \mathcal{H}^{0,1}(X,E)$ to parametrize the tangent bundle of $\mathcal{T}_{\mathfrak{g}}(S)$.

At this point, we consider on X the unique hyperbolic metric g_X with induced norm $|\cdot|$ and volume element dA. Also for $\beta \in A^{0,1}(X,E)$ the corresponding norm (in local coordinates) is given by $\|\beta\| = |\beta|(z)(g_X)^{\frac{\kappa-2}{2}}$.

Moreover, every $\beta \in A^{0,1}(X,E)$ admits the (unique) decomposition $\beta = \beta_0 + \bar{\partial}\eta$, with *harmonic* $\beta_0 \in A^{0,1}(X,E)$ and $\eta \in A^0(E)$. Therefore the class $[\beta] \in \mathcal{H}^{0,1}(X,E)$ is uniquely identified by its harmonic representative β_0 with respect to the metric g_X.

Thus, for any pair $(X,[\beta])$ and $t > 0$, we define the Donaldson functional:

$$D_t(u,\eta) = \int_X \left(\frac{1}{4}|\nabla u|^2 - u + te^u + 4e^{(\kappa-1)u}\|\beta_0 + \bar{\partial}\eta\|^2 \right) dA, \tag{1.1}$$

with the function u and the section η in the appropriate Sobolev spaces.

As observed in [29], it is possible to construct a (CMC) immersion with constant c, directly from a critical point of the Donaldson functional D_t with

$$t = 1 - c^2 > 0 \quad \text{and} \quad \kappa = 2. \tag{1.2}$$

Indeed, in this case, if (u, η) is a critical point of D_t, then it satisfies

$$\begin{cases} \Delta u + 2 - 2te^u - 8(\kappa - 1)e^{(\kappa-1)u}\|\beta_0 + \bar{\partial}\eta\|^2 = 0 \quad \text{in } X, \\ \bar{\partial}(e^{(\kappa-1)u} *_E (\beta_0 + \bar{\partial}\eta)) = 0. \end{cases} \tag{1.3}$$

Therefore, one may check that, if (1.2) holds, then $(X, e^u g_X)$ can be immersed as a (CMC) surface with constant $\pm c$ into a suitable hyperbolic 3-manifold $M^3 \simeq S \times \mathbb{R}$ with second fundamental form given by $\mathrm{II} = \mathrm{Re}(\alpha)$ and $\alpha = 8e^u *_E (\beta_0 + \bar{\partial}\eta) \in C_2(X)$, see [29,32,33] for details. Interestingly, as discussed in [33], system (1.3) can be recasted as Hitchin's selfduality equations for a suitable nilpotent $SL(2,\mathbb{C})$-Higgs bundle (of rank 2) and for this reason we refer to D_t as a Donaldson functional.

As also anticipated in [29], the following holds:

Theorem 1 ([33]). *For given $c \in (-1, 1)$, there is a one-to-one correspondence between the space of constant mean curvature c immersions into a 3-manifold of constant sectional curvature -1 and the tangent bundle of $\mathcal{T}_{\mathfrak{g}}(S)$, the latter parametrized by the pairs*

$$\big(X, [\beta]\big) \in \mathcal{T}_{\mathfrak{g}}(S) \times \mathcal{H}^{0,1}(X, E), \quad E = T_X^{1,0}.$$

Theorem 1 is a particular case of a more general result established in [33], showing that, for all $t > 0$ and $[\beta] \in \mathcal{H}^{0,1}(X, E)$, the Donaldson functional D_t admits a *unique* critical point (u_t, η_t), which is smooth and corresponds to the *global* minimum of D_t.

Such a uniqueness result yields also several interesting algebraic consequences. For example (for $\kappa = 2$ and $c = 0$), we derive a one-to-one correspondence between minimal immersions of S into a (germ of) hyperbolic 3-manifold and the irreducible representation of $\pi_1(S)$ into the group $\mathrm{PSL}(2,\mathbb{C})$ of the (orientation preserving) isometry group of $\mathbb{H}^3$.

On the grounds of Theorem 1, we can adventure to investigate the existence of (CMC) immersions with constant c reaching the limiting values $c = \pm 1$. Thus, for (u_c, η_c), the (unique) global minimum of D_t with $t = 1 - c^2$ and $\kappa = 2$, we can investigate if it survives the passage to the limit, as $|c| \to 1^-$. But we run immediately into trouble, since u_c could "blow up", as $|c| \to 1^-$. In fact, by using the blow-up analysis developed for solutions of Liouville equations, we find that actually blow-up can only occur around a finite number of (blow-up) points. We face a particularly delicate situation, when the blow-up point occurs at the "collapsing" of different zeroes of the holomorphic quadratic differential $\alpha_c = e^{u_c} *_E (\beta_0 + \bar{\partial}\eta_c) \in C_2(X)$. Recall that any holomorphic quadratic differential admits $4(\mathfrak{g} - 1) \geq 4$ zeroes in X (counted with multiplicity).

Thus, we devote the following sections to illustrate such a new scenario where, as pointed out in [35] and [41], we have to handle the new phenomenon of "blow-up without concentration." We present the recent results contained in [60,61]. Interestingly, when we deal with blow-up solutions carrying the least possible 'blow-up" mass 8π (see (3.19) and (3.20)), the pointwise estimates we obtain in the collapsing case are in striking analogy with the sharp "single bubble" estimates obtained in [8] and [38] for the nonvanishing (hence noncollapsing) case. Observe that no "bubble" is available in the "collapsing" situation.

By using the full power of the whole system (1.3), beyond the information encompassed by the mere Liouville equation, we are able to provide a useful description of (CMC) immersions with constant c "close" to ± 1 in some interesting cases, see Theorems 8 and 9.

In particular, we show that, for genus $\mathfrak{g} = 2$ and $[\beta] \neq 0$, the Donaldson functional at $t = 0$ is always bounded from below. This is a nontrivial information, since for $[\beta] = 0$ and $t = 0$, $D_{t=0}$ is always unbounded.

The seminal contribution contained in this note awaits improvements and some geometrical interpretation. We hope that our discussion will stimulate further investigation and new ideas in the pursuit of more complete results.

2. BLOW-UP AT COLLAPSING ZEROES: LOCAL ANALYSIS

Let $\Omega \subset \mathbb{R}^2$ be an open, bounded, and regular set, and consider the sequence:

$$\eta_k \in C^2(\Omega) \cap C^0(\overline{\Omega}),$$

satisfying the following Liouville-type problem:

$$\begin{cases} -\Delta \eta_k = W_k e^{\eta_k} \quad \text{in } \Omega, & (2.1) \\ \max\limits_{\partial\Omega} \eta_k - \min\limits_{\partial\Omega} \eta_k \leq C, & (2.2) \\ \displaystyle\int_\Omega W_k e^{\eta_k} \leq C, & (2.3) \end{cases}$$

with a weight function $W_k \geq 0$.

After the pioneering work of Brezis–Merle [6], a vast literature is now available, concerning the asymptotic behavior of η_k (possibly along a subsequence), as $k \to +\infty$, according to various assumptions on W_k and its vanishing behavior, see [2, 9, 11, 39, 55, 59]. Motivated by our applications, here we shall take W_k to satisfy

$$W_k \geq 0 \quad \text{and} \quad \|W_k\|_{L^\infty(\Omega)} + \int_\Omega \frac{1}{(W_k)^{\varepsilon_0}} \leq C, \quad \text{for some } \varepsilon_0 > 0. \tag{2.4}$$

As in [6], we say that η_k admits a *blow-up* point at $z_0 \in \Omega$, if

$$\exists z_k \to z_0 \quad \text{with } \eta_k(z_k) \to +\infty, \quad \text{as } k \to \infty \tag{2.5}$$

(possibly along a subsequence), and the value

$$\sigma(z_0) = \lim_{r\to 0} \liminf_{k\to+\infty} \int_{B_r(z_0)} W_k e^{\eta_k} \tag{2.6}$$

is called the "*blow-up mass*" of η_k at z_0.

The following result was pointed out in [61], as a general version of previous results contained in [2, 6, 55, 59]. We hope it can be useful in other contexts as well.

Proposition 2.1. *Let η_k satisfy (2.1)–(2.3) with $W_k \to W$ uniformly in $C^0_{\mathrm{loc}}(\Omega)$, and assume that (2.4) holds. Then (along a subsequence) η_k satisfies one of the following alternatives, as $k \to +\infty$:*

(i) $\eta_k \to -\infty$ *uniformly on compact sets of* Ω*;*

(ii) $\eta_k \to \eta_0$ *in* $C^0_{\mathrm{loc}}(\Omega)$*, with* η_0 *satisfying*

$$\begin{cases} -\Delta \eta_0 = W e^{\eta_0} & \text{in } \Omega, \\ \int_\Omega W e^{\eta_0} \le C; \end{cases}$$

(iii) *(blow-up) there exists a finite set* $\mathcal{S}$ *of blow-up points of* η_k *in* Ω*. Moreover, either*

("concentration") $\quad W_k e^{\eta_k} \rightharpoonup \sum_{q \in \mathcal{S}} \sigma(q)\delta_q$ *weakly in the sense of measures,*

and in particular $\eta_k \to -\infty$*, uniformly on compact sets of* $\Omega \setminus \mathcal{S}$*;*

or

("no concentration") $\quad \eta_k \to \eta_0$ *in* $C^0_{\mathrm{loc}}(\Omega \setminus \mathcal{S})$*,*

$$W_k e^{\eta_k} \rightharpoonup \sum_{q \in \mathcal{S}} \sigma(q)\delta_q + W e^{\eta_0}$$

weakly in the sense of measures, and η_0 *satisfies*

$$\begin{cases} -\Delta \eta_0 = W e^{\eta_0} + \sum_{q \in \mathcal{S}} \sigma(q)\delta_q & \text{in } \Omega, \\ \int_\Omega W e^{\eta_0} \le C. \end{cases} \tag{2.7}$$

Moreover, the blow-up mass satisfies $\sigma(q) \ge 4\pi$*,* $\forall q \in \mathcal{S}$*.*

Clearly, when alternative (iii) holds, in order to better understand the behavior of η_k around a blow-up point $q \in \mathcal{S}$, it is crucial to identify the specific value of the blow-up mass $\sigma(q)$ in (2.6).

In this respect, we recall the result of Li–Shafrir [39] and Bartolucci–Tarantello [2] in case,

$$W_k(x) = |x - p_k|^{2\alpha_k} h_k(x) \quad \text{in } B_r(q), \tag{2.8}$$

for $r > 0$ sufficiently small, with $p_k \in B_r(q)$ and

$$\begin{aligned} &h_k \to h \quad \text{uniformly with } 0 < a \le h \le b \text{ and } |\nabla h_k| \le A; \\ &0 \le \alpha_k \to \alpha, \quad p_k \to q, \text{as } k \to +\infty. \end{aligned} \tag{2.9}$$

Theorem 2 ([2, 39]). *If* η_k *in Proposition* 2.1 *satisfies alternative* (iii) *and for some* $q \in \mathcal{S}$ *the weight function* W_k *satisfies* (2.8)–(2.9)*, then* (iii)(a) *holds, in the sense that blow-up occurs with a "concentration" property. Furthermore,*

(i) *if* $W(q) > 0$ *(i.e.,* $\alpha_k \equiv 0$ *and* (2.9)*) then* $\sigma(q) = 8\pi$*,*

(ii) *if* $\alpha > 0$ *in* (2.9) *then* $\sigma(q) = 8\pi(1 + \alpha)$*.*

Therefore, we focus on a blow-up point $q \in \mathcal{S}$ with $W(q) = 0$ and q being the accumulation point of different zeroes of W_k (collapsing zeroes). In view of the applications we have in mind, we assume that the zeroes of W_k have integral multiplicity.

In [35] this situation was handled in case only two zeroes of W_k coalesce at q, while the general case was treated in [61], see also [36], [37]. The following "quantization" property for the "blow-up" mass holds:

Theorem 3 ([35,61]). *Suppose that η_k in Proposition 2.1 satisfy alternative* (iii). *Let $q \in \mathcal{S}$ and assume that, for $r > 0$ sufficiently small, we have*

$$W_k(x) = \left(\prod_{j=1}^{s} |x - p_{j,k}|^{2\alpha_j}\right) h_k(x), \quad \textit{for } x \in B_r(q) \textit{ and } s \geq 2, \tag{2.10}$$

and h_k satisfies (2.9) *in $B_r(q)$, $\alpha_j \in \mathbb{N}$ and $p_{j,k} \to q$, as $k \to +\infty$, $\forall j = 1, \ldots, s$. Then $\sigma(q) \in 8\pi\mathbb{N}$.*

The "local" results above can be used to describe the asymptotic behavior of solutions for Liouville-type equations on a compact Riemann surface (X, g). Denote by $d_g(\cdot, \cdot)$ the distance in (X, g). We consider a sequence $\xi_k \in C^{2,\alpha}(X)$ satisfying

$$-\Delta \xi_k = R_k e^{\xi_k} - f_k \quad \text{in } X, \tag{2.11}$$

where

$$R_k(z) = \left(\prod_{j=1}^{N} \left(d_g(z, z_{j,k})\right)^{2\alpha_j}\right) g_k(z), \quad z \in X; \tag{2.12}$$

$$g_k \in C^1(X) : a \leq g_k \leq b, \quad |\nabla g_k| \leq A \text{ and } g_k \to g_0 \text{ in } C^0(X); \tag{2.13}$$

$$z_{j,k} \in X : z_{j,k} \neq z_{l,k}, \quad j \neq l \in \{1, \ldots, N\} \text{ and } z_{j,k} \to z_j, j = 1, \ldots, N; \tag{2.14}$$

$$f_k \in C^{0,\alpha}(X), \quad f_k \to f_0 \text{ in } L^p(X), p > 1, \int_X f_0 \, dA \neq 0. \tag{2.15}$$

As before, we assume that

$$\alpha_j \in \mathbb{N}, \quad j = 1, \ldots, N. \tag{2.16}$$

In particular, we have that $R_k \to R_0$ uniformly in X, as $k \to +\infty$, with

$$R_0(z) = \left(\prod_{j=1}^{N} \left(d_g(z, z_j)\right)^{2\alpha_j}\right) g_0(z).$$

We denote by

$$Z = \left\{z \in X : R_0(z) = 0\right\} \tag{2.17}$$

the zero set of R_0. Clearly, $Z = \{z_1, \ldots, z_N\}$ with the point z_j given in (2.14) for $j = 1, \ldots, N$. We must keep in mind that such points are *not* necessarily distinct, as different zeroes of R_k could coalesce to the same zero of R_0. Therefore, we let Z_0 be the set (possibly empty) of such "collapsing" zeroes, namely

$$\begin{aligned} Z_0 = \big\{&z \in Z : \exists s \geq 2, 1 \leq j_1 < \cdots < j_s \leq N \text{ such that} \\ &z = z_{j_1} = \cdots = z_{j_s} \text{ and } z \notin Z \setminus \{z_{j_1}, \ldots, z_{j_s}\}\big\}. \end{aligned} \tag{2.18}$$

By combining the "local" results stated above, we can establish the following:

Theorem 4 (**[61]**). *Let ξ_k satisfy* (2.11) *and assume* (2.12)–(2.16). *Then, along a subsequence, one of the following alternatives holds:*

(i) *(compactness)* $\xi_k \to \xi_0$ *in* $C^2(X)$ *with*

$$-\Delta\xi_0 = R_0 e^{\xi_0} - f_0 \quad \textit{in } X, \tag{2.19}$$

(ii) *(blow-up) there exists a finite blow-up set*

$$\mathcal{S} = \left\{q \in X : \exists q_k \to q \textit{ and } \xi_k(q_k) \to +\infty,\ \textit{as } k \to +\infty\right\}$$

such that ξ_k *is uniformly bounded in* $C^2_{\mathrm{loc}}(X \setminus \mathcal{S})$ *and, as* $k \to +\infty$,

(a) *either (blow-up with concentration)*

$$\begin{aligned}&\xi_k \to -\infty \quad \textit{uniformly on compact sets of } X \setminus \mathcal{S};\\ &R_k e^{\xi_k} \rightharpoonup \sum_{q\in\mathcal{S}} \sigma(q)\delta_q \quad \textit{weakly in the sense of measures}, \sigma(q) \in 8\pi\mathbb{N}.\end{aligned} \tag{2.20}$$

In particular, $\int_X f_0\, dA \in 8\pi\mathbb{N}$ *in this case.*

(b) *or (blow-up without concentration)*

$$\begin{aligned}&\xi_k \to \xi_0 \quad \textit{in } C^2_{\mathrm{loc}}(X \setminus \mathcal{S});\\ &R_k e^{\xi_k} \rightharpoonup R_0 e^{\xi_0} + \sum_{q\in\mathcal{S}} \sigma(q)\delta_q \quad \textit{weakly in the sense of measures};\\ &-\Delta\xi_0 = R_0 e^{\xi_0} + \sum_{q\in\mathcal{S}} \sigma(q)\delta_q - f_0 \quad \textit{in } X,\ \sigma(q) \in 8\pi\mathbb{N}.\end{aligned}$$

Furthermore, in case alternative (ii)(b) *holds,* $\mathcal{S} \subset Z_0$ *and so any blow-up point occurs at a collapsing of zeroes of* R_k.

See **[61]** for details. As discussed in **[35]** and **[41]**, all the alternatives of Theorem 4 can actually occur.

Remark 2.1. If in (ii) we have $\mathcal{S} \setminus Z_0 \neq \emptyset$, then blow-up always occurs with the "concentration" property. So (2.20) holds and, by Theorem 2, for $q \in \mathcal{S} \setminus Z_0$, we have:

(1) $\sigma(q) = 8\pi$, if $q \notin Z$;

(2) $\sigma(q) = 8\pi(1+\alpha_j)$, if $q = z_j \in Z \setminus Z_0$.

As a direct consequence of Theorem 4, we may extend to the "collapsing" case the "compactness" result, well known to hold in the "non-collapsing" situation:

Corollary 2.1. *Under the assumption of Theorem* 4, *if*

$$\limsup_{k\to+\infty} \int_X R_k e^{\xi_k}\, dA < 8\pi,$$

then alternative (i) *holds.*

Next, we wish to provide more precise information around $q \in \mathcal{S} \cap Z_0$, a blow-up point of "collapsing" zeroes of R_k. To this purpose, we "localize" our analysis by introducing in X local holomorphic coordinates around q centered at the origin. Thus, with the obvious manipulations (see, e.g., [1, 2, 38]), and with abuse of notation, for $r > 0$ small, in $B_r = \{x \in \mathbb{R}^2 : |x| < r\}$ we may consider a sequence $\xi_k \in C^{2,\alpha}(B_r) \cap C(\overline{B_r})$ satisfying

$$\begin{cases} -\Delta \xi_k = W_k e^{\xi_k} \quad \text{in } B_r, & (2.21) \\ \max_{\partial B_r} \xi_k - \min_{\partial B_r} \xi_k \leq C, \quad \int_{B_r} W_k e^{\xi_k} \leq C, & (2.22) \\ \max_{\overline{B}_r} \xi_k = \xi_k(0) \to +\infty, \quad \text{as } k \to +\infty, & (2.23) \end{cases}$$

with

$$\begin{aligned} & W_k(x) = \left(\prod_{j=1}^{s} |x - p_{j,k}|^{2\alpha_j} \right) h_k(x), \quad \text{where } h_k \text{ satisfies (2.9) in } B_r; \\ & s \geq 2, \quad \alpha_j \in \mathbb{N}, p_{j,k} \neq p_{l,k} \text{ for } j \neq l; \\ & p_{j,k} \to 0, \quad \text{as } k \to +\infty, \ \forall j = 1, \ldots, s. \end{aligned} \tag{2.24}$$

Let us just recall that the bounded oscillation property stated in (2.22) follows from the global problem (2.11) by means of the Green representation formula. We have

$$|\nabla W_k| \leq A \quad \text{and} \quad W_k \to W \text{ in } C^0_{\text{loc}}(B_r), \text{ as } k \to +\infty, \tag{2.25}$$

with

$$W(x) = |x|^{2\alpha} h(x) \quad \text{and} \quad \alpha = \sum_{j=1}^{s} \alpha_j \in \mathbb{N}. \tag{2.26}$$

Furthermore, by taking $r > 0$ smaller if necessary, we may assume that zero is the *only* blow-up point of ξ_k in B_r, that is,

$$\forall 0 < \delta < r \ \exists C_\delta > 0 : \max_{\overline{B}_r \setminus B_\delta} \xi_k \leq C_\delta. \tag{2.27}$$

Clearly, under the assumptions above, Theorem 3 applies to ξ_k and implies the following for the "blow-up" mass:

$$\sigma := \lim_{\delta \to 0^+} \left(\lim_{k \to +\infty} \int_{B_\delta(0)} W_k e^{\xi_k} \right) \in 8\pi\mathbb{N}. \tag{2.28}$$

Here, we focus on the case of the *least* "blow-up" mass, namely when (2.28) holds with

$$\sigma = 8\pi. \tag{2.29}$$

Interestingly, in this case we are able to provide sharp pointwise estimates for ξ_k in B_r. This should be considered a first relevant step. Indeed, the analysis of multiple "blow-up," where $\sigma = 8\pi m$ with $m \in \mathbb{N}$ and $m \geq 2$, typically reduces to the case $\sigma = 8\pi$ after multiple rescaling, unless one ends up with a blow-up point at a "noncollapsing" zero of W, described in (ii) of Theorem 2. But in the latter case one can take advantage of the recent estimates in [3] and [67] to complete the analysis. Also we mention [35], where blow-up was analyzed when "collapsing" occurs between two zeroes, i.e., when $s = 2$ in (2.24).

The following estimates were derived in [61].

Theorem 5 ([61]). *Let ξ_k satisfy the assumptions above. If* (2.29) *holds, then*

(i) $\xi_k(0) = -(\min_{\partial B_r} \xi_k + 2\sum_{j=1}^{s} 2\alpha_j \log|p_{j,k}|) + O(1);$

(ii) $\xi_k(x) = \log \frac{e^{\xi_k(0)}}{(1+\frac{1}{8}W_k(0)e^{\xi_k(0)}|x|^2)^2} + O(1);$

(iii) $\int_{B_r} |\nabla \xi_k|^2 = -16\pi(\min_{\partial B_r} \xi_k + \sum_{j=1}^{s} 2\alpha_j \log|p_{j,k}|) + O(1).$

It is interesting to compare the above estimates with those available in [8] and [38] (see, e.g., Theorem 0.3 in [38]) for solutions of (2.21)–(2.23), when (2.25) holds with $W(0) > 0$ (instead of (2.26) as considered here). In this case, (2.29) is automatically satisfied (see (i) of Theorem 2) and the estimate (ii) of Theorem 5 is the striking exact analogue of the pointwise estimate provided in Theorem 0.3 of [38]. Furthermore, by considering the sequence

$$u_k(x) = \xi_k(x) + \sum_{j=1}^{s} 2\alpha_j \log|x - p_{j,k}|,$$

$$\text{satisfying } -\Delta u_k = h_k e^{u_k} - 4\pi \sum_{j=1}^{s} \alpha_j \delta_{p_j,k} \text{ in } B_r,$$

we realize that the estimate (i) stated for ξ_k in Theorem 5 reduces just to the following "sup + inf" estimate of Harnack type [5] for u_k:

$$u_k(0) + \min_{\partial B_r} u_k = O(1), \tag{2.30}$$

which was established in this form in [58] when the origin is a "noncollapsing" zero of W. Therefore, we expect that the estimate (2.30) should remain valid in the "collapsing" case as well, without the assumption (2.29).

We shall use those estimates to describe the asymptotic behavior of minimizers of the Donaldson functional, considered in [29,33].

3. ASYMPTOTICS FOR MINIMIZERS OF THE DONALDSON FUNCTIONAL

Let S be a smooth, closed, oriented surface of genus $\mathfrak{g} \geq 2$, and denote by $\mathcal{T}_{\mathfrak{g}}(S)$ the Teichmüller space of conformal structures on S, modulo biholomorphisms in the homotopy class of the identity.

We fix a conformal structure $X \in \mathcal{T}_{\mathfrak{g}}(S)$ and denote by g_X the corresponding hyperbolic metric on X, which will be used as the background metric, with norm $|\cdot|$ and volume element dA.

On (X, g_X) we consider a Donaldson functional assigned in terms of a pair of (conformal) data $(X, [\beta]) \in \mathcal{T}_{\mathfrak{g}}(S) \times \mathcal{H}^{0,1}(X, E)$, where $E = \otimes^{k-1} T_X^{1,0}$ with $\kappa \geq 2$ and $T_X^{1,0}$ the holomorphic tangent bundle of X and $\mathcal{H}^{0,1}(X, E) = A^{0,1}(X, E)/\bar{\partial}(A^0(E))$ is the $(0,1)$-Dolbeault cohomology group. We recall that $A^{0,1}(X, E)$ is the space of $(0,1)$-forms in X

valued in E, $A^0(E)$ is the space of smooth sections of E and $\bar\partial : A^0(E) \to A^{0,1}(X, E)$ is the d-bar operator. For $\beta \in A^{0,1}(X, E)$, we have the decomposition $\beta = \beta_0 + \bar\partial\eta$, with β_0 a unique *harmonic* $(0, 1)$-form valued on E and $\eta \in A^0(E)$. So the class $[\beta] \in \mathcal{H}^{0,1}(X, E)$ is uniquely identified by its harmonic representative β_0. We also recall that, by means of the Hodge star operator $*_E : A^{0,1}(E) \to A^{1,0}(E^*)$ and by Serre's duality theorem (see [66]), for any class $[\beta] = [\beta_0 + \bar\partial\eta] \in \mathcal{H}^{0,1}(X, E)$ with β_0 harmonic, we can uniquely identify $*_E\beta_0$ with a holomorphic κ-differential on X. In other words, denoting by $C_\kappa(X)$ the space of κ-holomorphic differentials, we have that $C_\kappa(X) \simeq (\mathcal{H}^{0,1}(X, E))^*$. Moreover, the linear complex space $C_\kappa(X)$ is finite dimensional and $\dim_{\mathbb{C}} C_\kappa(X) = (2\kappa - 1)(\mathfrak{g} - 1)$. Since $\mathcal{T}_{\mathfrak{g}}(S)$ is a complex cell of dimension $3(\mathfrak{g} - 1)$, we find that, for $\kappa = 2$, the pair $(X, [\beta]) \in \mathcal{T}_{\mathfrak{g}}(S) \times \mathcal{H}^{0,1}(X, E)$ can be used to parametrize the tangent bundle of $\mathcal{T}_{\mathfrak{g}}(S)$.

In addition, recall that in local holomorphic coordinates $\{z\}$, any $\alpha \in C_\kappa(X)$ takes the expression $\alpha = h(dz)^k$, with h holomorphic. In this way, a zero for α is well understood, and actually, it is known that α admits $2\kappa(\mathfrak{g} - 1)$ zeroes in X, counted with multiplicity.

At this point, for a given pair $(X, [\beta])$ and $t > 0$, we define the *Donaldson functional*

$$D_t(u, \eta) = \int_X \left(\frac{|\nabla u|^2}{4} - u + te^u + 4e^{(\kappa-1)u} \|\beta_0 + \bar\partial\eta\|^2 \right) dA \tag{3.1}$$

with "natural" (convex) domain

$$\Lambda = \left\{ (u, \eta) \in H^1(X) \times W^{1,2}(X, E) : \int_X e^{(\kappa-1)u} \|\beta_0 + \bar\partial\eta\|^2 \, dA < +\infty \right\}.$$

Here, $H^1(X)$ and $W^{1,2}(X, E)$ are the usual Sobolev spaces. Clearly, the functional D_t is bounded from below in Λ.

In [33], the authors have shown that, for any $[\beta] \in \mathcal{H}^{0,1}(X, E)$ and $t > 0$, the functional D_t attains its infimum on Λ at a *smooth* pair (u_t, η_t) satisfying

$$\begin{cases} \Delta u + 2 - 2te^u - 8(\kappa - 1)e^{(\kappa-1)u} \|\beta_0 + \bar\partial\eta\|^2 = 0 & \text{in } X, \\ \bar\partial\big(e^{(\kappa-1)u} *_E (\beta_0 + \bar\partial\eta)\big) = 0. \end{cases} \tag{3.2}$$

More importantly, it is possible to show the *unique* solvability of (3.2).

Theorem 6 ([33]). *For given $t > 0$ and $[\beta] \in \mathcal{H}^{0,1}(X, E)$, the functional D_t admits a unique critical point (u_t, η_t), which corresponds to its global minimum in Λ. Furthermore, (u_t, η_t) is smooth and it is the only solution of* (3.2).

Such a uniqueness result implies relevant information about the moduli space of minimal, constant mean curvature, and Lagrangean immersions into hyperbolic 3-manifolds, and also about the irreducible representation of the fundamental group $\pi_1(S)$ in various character varieties. We refer to [33] and the references therein for more details. Here, we only mention the following consequence of Theorem 6 about the immersion of constant mean curvature (CMC) surfaces:

Corollary 3.1 ([29, 33]). *For a given $c \in (-1, 1)$, there is a one-to-one correspondence between the space of constant mean curvature c immersions of S into 3-manifolds of constant sectional curvature -1 and the tangent bundle of $\mathcal{T}_{\mathfrak{g}}(S)$, the latter parametrized by the* *pair*

$$(X,[\beta]) \in \mathcal{T}_{\mathfrak{g}}(S) \times \mathcal{H}^{0,1}(X,E).$$

Clearly, Corollary 3.1 is a direct consequence of Theorem 6, once we take $\kappa = 2$ and $t = 1 - c^2 > 0$. We refer the reader to [33] for details.

For $X \in \mathcal{T}_{\mathfrak{g}}(S)$ fixed, by this approach, one may be tempted to look for (CMC) immersions of X with constant $c = \pm 1$, simply by taking $t = 1 - c^2$ and by following the solution (u_t, η_t) to the limit, as $t \to 0^+$. However, this requires a rather delicate analysis. Indeed, it is not even clear for which data $(X,[\beta])$, the functional $D_0 = D_{t=0}$, given by

$$D_0(u,\eta) = \int_X \left(\frac{|\nabla u|^2}{4} - u + 4e^{(\kappa-1)u} \|\beta_0 + \bar{\partial}\eta\|^2 \right) dA, \tag{3.3}$$

is bounded from below in Λ. Notice, for example, that for $t > 0$ and $[\beta] = 0$ (i.e., $\beta_0 = 0$),

$$u_t = \log \frac{1}{t}, \quad \eta_t = 0 \text{ and } D_t(u_t,\eta_t) = \log t \to -\infty, \quad \text{as } t \to 0^+.$$

Obviously, for $\beta_0 = 0$ and $t = 0$, the system of equations (3.2) admits no solutions.

On the other hand, for $[\beta] \neq 0$, the following has been established in [60].

Theorem 7 ([60]). *Let $[\beta] \neq 0$ and assume that D_0 admits a critical point (u_0, η_0) (or equivalently, the system (3.2) for $t = 0$ is solvable). Then D_0 is bounded from below in Λ and (u_0,η_0) is unique, smooth, and it corresponds to the global minimum of D_0 in Λ.*

Therefore, as anticipated, to find out if D_0 admits a critical point, we need to analyze the convergence of (u_t,η_t) (the global minimum of D_t), as $t \to 0^+$. We shall see that the failure of convergence of (u_t,η_t) (along a subsequence) is due to "blow-up" phenomena.

For $\kappa = 2$, such an asymptotic analysis allows us to obtain information about (CMC)-immersions, when the constant c approaches the limiting values ± 1. On the other hand, when $\kappa \geq 2$, such an asymptotic analysis permits to follow the behavior of the global minimizer $(u_\lambda, \eta_\lambda)$ of the Donaldson functional

$$D(u,\eta) = \int_X \left(\frac{|\nabla u|^2}{4} - u + e^u + 4e^{(\kappa-1)u} \|\lambda\beta_0 + \bar{\partial}\eta\|^2 \right) dA \tag{3.4}$$

along the $(0,1)$-Dolbeault cohomology classes $[\lambda\beta]$, as λ varies in $(0,+\infty)$ and $[\beta] \neq 0$ is fixed in $\mathcal{H}^{0,1}(X,E)$. Indeed, via the transformations

$$\begin{gathered} t = \lambda^{-\frac{2}{\kappa-1}}, \quad u_t = u_\lambda + \frac{2}{\kappa-1}\log\lambda, \quad \eta_t = \frac{1}{\lambda}\eta_\lambda, \quad \text{and} \\ D_t(u_t,\eta_t) = D(u_\lambda,\eta_\lambda) - 4\pi(\mathfrak{g}-1)\log\lambda^{\frac{2}{\kappa-1}}, \end{gathered} \tag{3.5}$$

we can recast the analysis of (u_λ,η_λ) (the global minimum of D in (3.4)), as $\lambda \to +\infty$, to the analysis of (u_t,η_t) (the global minimum of D_t in (3.1)), as $t \to 0^+$.

To start, we notice that, by the strict positivity of the Hessian D_t'' at (u_t,η_t) (see [29,33]) and the Implicit Function Theorem, we can show the C^2-dependence of (u_t,η_t) with respect to $t \in (0,+\infty)$. We refer for details to [60], where it is also shown that the expression

$t \int_X e^{u_t}\, dA$ is increasing, as a function of $t \in (0, +\infty)$. Since, after integration over X of the first equation in (3.2), we have

$$t \int_X e^{u_t}\, dA + 4(\kappa - 1) \int_X e^{(k-1)u_t} \|\beta_0 + \bar{\partial}\eta_t\|^2\, dA = 4\pi(\mathrm{g} - 1), \tag{3.6}$$

we may conclude that

$$\rho_t\big([\beta]\big) := 4(\kappa - 1) \int_X e^{(\kappa-1)u_t} \|\beta_0 + \bar{\partial}\eta_t\|^2\, dA \in \big(0, 4\pi(\mathrm{g} - 1)\big) \tag{3.7}$$

is *decreasing* in $(0, +\infty)$. These facts lead us to ask the following question:

Question 1. Can we identify the value

$$\rho\big([\beta]\big) = \rho(\beta_0) = \lim_{t\to 0^+} 4(\kappa - 1) \int_X e^{(\kappa-1)u_t} \|\beta_0 + \bar{\partial}\eta_t\|^2\, dA \tag{3.8}$$

in terms of the given cohomology class $[\beta] = [\beta_0 + \bar{\partial}\eta] \in \mathcal{H}^{0,1}(X, E)$?

To emphasize the relevance of the value $\rho([\beta])$ in (3.8), we observe that, for $[\beta] = [\beta_0 + \bar{\partial}\eta] \neq 0$, the interval $(0, \rho([\beta]))$ provides the range of the (decreasing) function

$$\rho_t\big([\beta]\big) = 4(\kappa - 1) \int_X e^{(\kappa-1)u_t} \|\beta_0 + \bar{\partial}\eta_t\|^2\, dA, \quad \text{as } t \text{ varies in } (0, +\infty).$$

We summarize the following consequences of the above discussion:

Proposition 3.1. *Given $[\beta] \in \mathcal{H}^{0,1}(X, E)$, there hold:*

(i) $\rho([\beta]) \in [0, 4\pi(\mathrm{g} - 1)]$ *and* $\rho([\beta]) = 0 \iff [\beta] = 0$*;*

(ii) *If* $[\beta] \neq 0$*, then for every* $0 < \rho < \rho([\beta])$*, there exists a unique* $\lambda \in (0, +\infty)$ *such that* $\rho = 4(\kappa - 1) \int_X e^{(\kappa-1)u_\lambda} \|\lambda\beta_0 + \bar{\partial}\eta_\lambda\|^2\, dA$*, where* $(u_\lambda, \eta_\lambda)$ *is the global minimum (and unique critical point) for* D *in* (3.4)*.*

Letting $c_t = D_t(u_t, \eta_t) = \min_\Lambda D_t$, we see that it is increasing for $t \in (0, +\infty)$ and therefore,

$$D_0 \quad \text{is bounded from below on } \Lambda \iff \inf_{t>0} c_t = \lim_{t\to 0^+} c_t = c_0 > -\infty \tag{3.9}$$

and $\inf_\Lambda D_0 = c_0$. More importantly, it has been proved in [60] that the following holds:

Proposition 3.2 ([60])**.** *If D_0 is bounded from below on Λ then $\rho([\beta]) = 4\pi(\mathrm{g} - 1)$.*

Now the delicate questions are the following:

Question 2. (i) If $\rho([\beta]) = 4\pi(\mathrm{g} - 1)$, is it true that D_0 is bounded from below in Λ?

(ii) If D_0 is bounded from below in Λ, for which $[\beta] \neq 0$ is the infimum attained?

In order to investigate the questions raised above, we set

$$\beta_t = \beta_0 + \bar{\partial}\eta_t \in A^{0,1}(X, E) \quad \text{and} \quad \alpha_t = e^{u_t} *_E \beta_t. \tag{3.10}$$

By virtue of the second equation in (3.2), we know that

$$\alpha_t \in C_\kappa(X) = \{\alpha \in A^{1,0}(X, E^*) : \bar{\partial}\alpha = 0\}, \quad \alpha_t \neq 0,$$

namely $\alpha_t \neq 0$ is a holomorphic κ-differential in X, and so it admits $2\kappa(\mathfrak{g}-1)$ zeroes in X, counted with multiplicity. Moreover, since $C_\kappa(X)$ is finite dimensional, all norms of α_t are equivalent. Let

$$s_t \in \mathbb{R} : e^{(\kappa-1)s_t} = \|\alpha_t\|^2_{L^\infty} \quad \text{and} \quad \hat{\alpha}_t = \frac{\alpha_t}{\|\alpha_t\|_{L^\infty}} = e^{-\frac{(k-1)s_t}{2}}\alpha_t. \tag{3.11}$$

Then, as $t \to 0^+$ (along a subsequence), we have $\hat{\alpha}_t \to \hat{\alpha}_0$ with

$$\hat{\alpha}_0 \in C_\kappa(X) \quad \text{and} \quad \|\hat{\alpha}_0\|_{L^\infty} = 1.$$

So, also $\hat{\alpha}_0$ must vanish at $2\kappa(\mathfrak{g}-1)$ points (counted with multiplicity), which correspond to the limits of the zeroes of $\hat{\alpha}_t$ (along a subsequence). Obviously, different zeroes of $\hat{\alpha}_t$ could coalesce into the same zero of $\hat{\alpha}_0$. It is shown in [60] that, in order to describe the asymptotic behavior of (u_t, η_t) satisfying (3.2), it is possible to use the blow-up analysis discussed in Section 2 for (a subsequence of)

$$\xi_t = -u_t + s_t. \tag{3.12}$$

With this information, it is possible to obtain the following (nontrivial) lower bound:

Proposition 3.3 ([60]). *For $[\beta] \in \mathcal{H}^{0,1}(X, E) \setminus \{0\}$, there holds:*

$$\rho([\beta]) \geq \frac{4\pi}{\kappa - 1} \quad \textit{with } \rho([\beta]) \textit{ in } (3.8).$$

For details, we refer the interested reader to [60].

Next, along a suitable sequence $t_k \to 0^+$, we are going to analyze more closely the sequence $\xi_k = \xi_{t_k}$ in (3.12). To reduce technicalities, from now on we focus on the case

$$\kappa = 2. \tag{3.13}$$

We let $u_k = u_{t_k}$, $\eta_k = \eta_{t_k}$, $s_k = s_{t_k}$, and $\alpha_k = \alpha_{t_k}$, so that the function

$$\xi_k = -(u_k - s_k) \tag{3.14}$$

satisfies

$$-\Delta\xi_k = 8\|\hat{\alpha}_k\|^2 e^{\xi_k} - f_k \quad \text{in } X, \tag{3.15}$$

with $f_k = 2(1 - t_k e^{u_k})$ and $\hat{\alpha}_k = e^{-\frac{s_k}{2}}\alpha_k$ satisfying $\|\hat{\alpha}_k\|_{L^\infty} = 1$. By the maximum principle, we also know that $\|f_k\|_{L^\infty(X)} \leq 2$. So, along the given sequence, we can further assume that

$$f_k \to f_0 \quad \text{in } L^p(X), p > 1, \quad \text{and} \quad \hat{\alpha}_k \to \hat{\alpha}_0 \in C_2(X), \|\hat{\alpha}_0\|_{L^\infty} = 1.$$

So, for $N = 4(\mathfrak{g}-1)$ (recall (3.13)), we let $Z = \{z_1, \ldots, z_N\}$ be the set of zeroes of $\hat{\alpha}_0$, repeated according to their multiplicity. Clearly, the set Z is formed by the limit points of the zeroes of $\hat{\alpha}_k$, which may coalesce into the same zero of $\hat{\alpha}_0$. Thus, we let Z_0 the set (possibly empty) of such "collapsing" zeroes of $\hat{\alpha}_k$, as defined in (2.18).

Theorem 4 applies to ξ_k and (possibly along a subsequence) implies that:

(i) either (compactness) $\xi_k \to \xi_0$ in $C^2(X)$, as $k \to +\infty$, and D_0 is bounded from below and attains its infimum in Λ;

(ii) or (blow-up) ξ_k admits a finite blow-up set

$$\mathcal{S} = \{q_1, \dots, q_n : 1 \le n \le \mathfrak{g} - 1\},$$

and we may have "blow-up with concentration," or "blow-up without concentration," as described respectively in parts (ii)(a) and (ii)(b) of Theorem 4 (with $R_k = 8\|\hat{\alpha}_k\|^2$ and $R_0 = \|\hat{\alpha}_0\|^2$).

At this point, by exploiting the full power of the whole system (3.2), it is possible to provide a careful description of the minimizer (u_k, η_k) of D_{t_k} in case of blow-up.

We start to discuss the case where we assume that $\mathcal{S} \cap Z = \emptyset$, namely no blow-up point coincides with a zero of $\hat{\alpha}_0$. In this situation, by Remark 2.1, we know that only "blow-up with concentration" occurs [6,38]. Therefore,

$$8e^{u_k}\|\alpha_k\|^2 = 8\|\hat{\alpha}_k\|^2 e^{\xi_k} \rightharpoonup 8\pi \sum_{l=1}^{n} \delta_{q_l}, \quad \text{and we obtain } \rho\big([\beta]\big) = 4\pi n.$$

To proceed further, we follow [60], and for a given set $P = \{x_1, \dots, x_\nu\} \subset X$ with $1 \le \nu \le (\mathfrak{g} - 1)$, we introduce the following subspace of $C_2(X)$:

$$Q_2[P] = Q_2\big[\{x_1, \dots, x_\nu\}\big] = \big\{\alpha \in C_2(X) : \alpha \text{ vanishes exactly at the set } P\big\}.$$

By the Riemann–Roch theorem, we have

$$\dim_{\mathbb{C}} Q_2\big[\{x_1, \dots, x_\nu\}\big] = 3(\mathfrak{g} - 1) - \nu.$$

In [60] it has been shown that the following holds:

Theorem 8 ([60]). *Assume that ξ_k blows up (in the sense of (ii) above). If* (3.13) *holds and*

$$\mathcal{S} \cap Z = \emptyset, \tag{3.16}$$

then (along a subsequence), as $k \to +\infty$,

$$\begin{aligned}
&\alpha_k \to \alpha_0 \in C_2(X) \quad \textit{with } \alpha_0 \neq 0 \textit{ vanishing exactly at } Z,\\
&e^{-u_k} \rightharpoonup 4\pi \sum_{q \in \mathcal{S}} \frac{1}{\|\alpha_0\|^2(q)} \delta_q \quad \textit{weakly in the sense of measures}, &(3.17)\\
&c_k = D_{t_k}(u_k, \eta_k) = -4\pi(\mathfrak{g} - 1 - n)d_k + O(1), \quad \textit{with } d_k = \fint_X u_k \, dA \to +\infty,\\
&\int_X \beta_0 \wedge \alpha \, dA = 0, \quad \forall \alpha \in Q_2[\mathcal{S}]. &(3.18)
\end{aligned}$$

Furthermore, $\rho([\beta]) = \int_X \beta_0 \wedge \alpha_0 \, dA = 4\pi n$.

Remark 3.1. Since $\dim_{\mathbb{C}} Q_2[\mathcal{S}] = 3(\mathfrak{g} - 1) - n$, the orthogonality condition (3.18), together with the estimate (3.17) for the global minimizer of D_{t_k}, seems to indicate that ξ_k should admit only *one* blow-up point ($n = 1$), where the holomorphic quadratic differential $*_E \beta_0$ does not vanish.

When (3.16) holds, the estimate (3.17) allows us to answer Question 2 posed above. Indeed, if $\rho([\beta]) = 4\pi(\mathfrak{g} - 1)$ then $n = \mathfrak{g} - 1$, and therefore, by using (3.17), we find that

D_0 is bounded from below in Λ. However, the analysis above seems to suggests that D_0 may not attain its infimum in Λ.

Next we wish to acquire some useful information about the blow-up behavior of (u_k, η_k) when we no longer assume (3.16). By taking advantage of the blow-up analysis developed in Section 2, we focus to the case where blow-up occurs with the “least” blow-up mass. More precisely, for the blow-up mass

$$\sigma(q) = \lim_{r\to 0^+}\left(\lim_{k\to+\infty}\int_{B_r(q)} e^{u_k}\|\beta_0 + \bar{\partial}\eta_k\|^2 dA\right) \in 8\pi\mathbb{N}, \quad \forall q \in \mathcal{S}, \tag{3.19}$$

we assume that

$$\sigma(q) = 8\pi, \quad \forall q \in \mathcal{S}. \tag{3.20}$$

Remark 3.2. When (3.20) holds, it is shown in [60] that every blow-up point $q \in \mathcal{S} \cap Z$ must correspond to a collapsing of zeroes, that is,

$$\mathcal{S} \cap Z = \mathcal{S} \cap Z_0. \tag{3.21}$$

For $q_l \in \mathcal{S}$ and $r > 0$ sufficiently small, let

$$x_{k,l} \in B_r(q_l) : \xi_k(x_{k,l}) = \max_{B_r(q_l)} \xi_k \to +\infty \quad \text{and} \quad x_{k,l} \to q_l, \text{ as } k \to +\infty, \tag{3.22}$$

and set

$$\mu_{k,l} = \|\alpha_k\|^2(x_{k,l}). \tag{3.23}$$

In [60], it is shown that the following holds:

Theorem 9 ([60]). *Assume (3.20) and suppose that $\mathcal{S} \cap Z \neq \emptyset$. Then (along a subsequence)*

$$s_k \to +\infty, \quad \text{as } k \to +\infty.$$

Moreover, there exists a set of indices $J \subseteq \{1, \dots, n\}$ such that, as $k \to +\infty$,

(i) *$\forall l \in J$ we have $q_l \in \mathcal{S} \cap Z = \mathcal{S} \cap Z_0$ and $\mu_{k,l} \to \mu_l > 0$,*

$$e^{-u_k} \rightharpoonup 4\pi \sum_{l\in J} \frac{1}{\mu_l}\delta_{q_l} \quad \text{weakly in the sense of measures;}$$

(ii) *$\int_X \beta_0 \wedge \alpha \, dA = 0$, $\forall \alpha \in C_2(X)$ vanishing at $\mathcal{S}_0 = \{q_l \in \mathcal{S} : l \in J\} \subset Z_0$. In particular, $\int_X \beta_0 \wedge \hat{\alpha}_0 \, dA = 0$;*

(iii) *$\mu_{k,l} \to +\infty$, as $k \to +\infty$, $\forall l \in \{1, \dots, n\} \setminus J$ (if not empty),*

$$c_k = D_{t_k}(u_k, \eta_k) = -4\pi(\mathrm{g} - 1 - n)d_k - \sum_{l\in\{1,\dots,n\}\setminus J} \log(\mu_{k,l}) + O(1),$$

$$\text{with } d_k = \fint_X u_k \, dA \to +\infty. \tag{3.24}$$

We can reveal a clearer relation between Theorems 8 and 9, when J covers the full set of possible indices, namely when $J = \{1, \dots, n\}$, which is reasonable, as we expect that $n = 1$. With the above notations, the following holds:

Corollary 3.2. *Under the assumptions of Theorem 9, if in part 2 we have*

$$J = \{1, \dots, n\},$$

then $\mathcal{S} \subset Z_0$. Moreover, as $k \to +\infty$,

(i) $e^{-u_k} \to 4\pi \sum_{l=1}^{n} \frac{1}{\mu_l} \delta_q$ *weakly in the sense of measures;*

(ii) $c_k = -4\pi(\mathfrak{g} - 1 - n)d_k + O(1)$ *with* $d_k = ⨍_X u_k \, dA \to +\infty$,

and

(iii) $\int_X \beta_0 \wedge \alpha \, dA = 0$, $\forall \alpha \in Q_2[\mathcal{S}]$.

When $\mathfrak{g} = 2$, $\mathcal{S}$ contains at most one single point ($n = 1$), and by virtue of Propositions 3.1 and 3.3, we know that

$$\rho\big([\beta]\big) = 4\pi, \quad \text{for every } [\beta] \in \mathcal{H}^{0,1}(X, E) \setminus \{0\}.$$

Thus, as a consequence of Corollary 2.1 or Theorem 8 and Corollary 3.2, we obtain

Corollary 3.3. *For the genus $\mathfrak{g} = 2$, the functional D_0 in (3.3) is bounded from below, whenever $[\beta] \neq 0$.*

As a final observation, we add that in Theorem 9 it should be possible to remove the assumption (3.20). However, when (3.20) is no longer valid then also (3.21) cannot be expected to hold (recall Remark 3.2) and so we could end up with a blow-up point $q \in Z \setminus Z_0$. Namely, blow-up can occur at a zero of $\hat{\alpha}_0$ which *does not* coincide with a "collapsing" of zeroes of $\hat{\alpha}_k$. As well known, in this case one needs to deal with a "multiple bubble" situation where, after rescaling, the "bubbles" are symmetrically placed (see [3]). This fact causes some "cancelation" phenomena that prevents to obtain, as in [60], a nice control on the sequence s_k. It is likely that the new sharper estimates obtained by Wei–Zhang in [67] may help resolve such difficulties.

ACKNOWLEDGMENTS

We thank Dr. Martin Mayer for his precious assistance and useful comments during the preparation of the manuscript.

FUNDING

This work was partially supported by MIUR excellence project, Department of Mathematics, University of Rome "Tor Vergata" CUP E83C18000100006, and by "Beyond Borders" research project "Variational approaches to PDE."

REFERENCES

[1] D. Bartolucci, C. C. Chen, C. S. Lin, and G. Tarantello, Profile of blow-up solutions to mean field equations with singular data. *Comm. Partial Differential Equations* **29** (2004), no. 7–8, 1241–1265.

[2] D. Bartolucci and G. Tarantello, Liouville type equations with singular data and their applications to periodic multivortices for the electroweak theory. *Comm. Math. Phys.* **229** (2002), no. 1, 3–47.

[3] D. Bartolucci and G. Tarantello, Asymptotic blow-up analysis for singular Liouville type equations with applications. *J. Differential Equations* **262** (2017), no. 7, 3887–3931.

[4] L. Battaglia, A. Jevnikar, A. Malchiodi, and D. Ruiz, A general existence result for the Toda system on compact surfaces. *Adv. Math.* **285** (2015), 937–979.

[5] H. Brezis, Y. Y. Li, and I. Shafrir, A sup + inf inequality for some nonlinear elliptic equations involving exponential nonlinearities. *J. Funct. Anal.* **115** (1993), no. 2, 344–358.

[6] H. Brezis and F. Merle, Uniform estimates and blow-up behavior for solutions of $-\Delta u = V(x)e^u$ in two dimensions. *Comm. Partial Differential Equations* **16** (1991), no. 8–9, 1223–1253.

[7] C. L. Chai, C. S. Lin, and C. L. Wang, Mean field equations, hyperelliptic curves and modular forms: I. *Cambridge J. Math.* **3** (2015), no. 1–02, 127–274.

[8] C. C. Chen and C. S. Lin, Sharp estimates for solutions of multi-bubbles in compact Riemann surfaces. *Comm. Pure Appl. Math.* **55** (2002), no. 6, 728–771.

[9] C. C. Chen and C. S. Lin, Topological degree for a mean field equation on Riemann surfaces. *Comm. Pure Appl. Math.* **56** (2003), no. 12, 1667–1727.

[10] C. C. Chen and C. S. Lin, Mean field equations of Liouville type with singular data: sharper estimates. *Discrete Contin. Dyn. Syst.* **28** (2010), no. 3, 1237–1272.

[11] C. C. Chen and C. S. Lin, Mean field equation of Liouville type with singular data: topological degree. *Comm. Pure Appl. Math.* **68** (2015), no. 6, 887–947.

[12] Q. Chen, W. Wang, Y. Wu, and B. Xu, Conformal metrics with constant curvature one and finitely many conical singularities on compact Riemann surfaces. *Pacific J. Math.* **273** (2015), no. 1, 75–100.

[13] Z. Chen, T. J. Kuo, and C. S. Lin, The geometry of generalized Lamé equation, I. *J. Math. Pures Appl.* **127** (2019), 89–120.

[14] Z. Chen, T. J. Kuo, and C. S. Lin, The geometry of generalized Lamé equation, II: existence of premodular forms and application. *J. Math. Pures Appl.* **132** (2019), 251–272.

[15] Z. Chen, T. J. Kuo, and C. S. Lin, Simple zero property of some holomorphic functions on the moduli space of tori. *Sci. China Math.* **62** (2019), no. 11, 2089–2102.

[16] Z. Chen, T. J. Kuo, C. S. Lin, and K. Takemura, On reducible monodromy representations of some generalized Lamé equation. *Math. Z.* **288** (2018), 679–688.

[17] Z. Chen, T. J. Kuo, C. S. Lin, and K. Takemura, Real-root property of the spectral polynomial of the Treibich–Verdier potential and related problems. *J. Differential Equations* **264** (2018), no. 8, 5408–5431.

[18] Z. Chen, T. J. Kuo, C. S. Lin, and C. L. Wang, Green function, Painlevé VI equation, and Eisenstein series of weight one. *J. Differential Geom.* **108** (2018), no. 2, 185–241.

[19] S. Dey, Spherical metrics with conical singularities on 2-spheres. *Geom. Dedicata* **196** (2018), 53–61.

[20] S. K. Donaldson, Anti self-dual Yang–Mills connections over complex algebraic surfaces and stable vector bundles. *Proc. Lond. Math. Soc.* **50** (1985), no. 1, 1–26.

[21] S. K. Donaldson, Twisted harmonic maps and the self-duality equations. *Proc. Lond. Math. Soc.* **55** (1987), no. 1, 127–131.

[22] A. Eremenko, Metrics of positive curvature with conic singularities on the sphere. *Proc. Amer. Math. Soc.* **132** (2004), no. 11, 3349–3355.

[23] A. Eremenko, Co-axial monodromy. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **20** (2020), no. 2, 619–634.

[24] A. Eremenko, Metrics of constant positive curvature with four conic singularities on the sphere. *Proc. Amer. Math. Soc.* **148** (2020), no. 9, 3957–3965.

[25] A. Eremenko, A. Gabrielov, and A. Hinkkanen, Exceptional solutions to the Painlevé VI equation. *J. Math. Phys.* **58** (2017), no. 1, 012701, 8 pp.

[26] A. Eremenko, A. Gabrielov, and V. Tarasov, Metrics with conic singularities and spherical polygons. *Illinois J. Math.* **58** (2014), no. 3, 739–755.

[27] A. Eremenko, A. Gabrielov, and V. Tarasov, Spherical quadrilaterals with three non-integer angles. *Zh. Mat. Fiz. Anal. Geom.* **12** (2016), no. 2, 134–167.

[28] A. Eremenko and V. Tarasov, Fuchsian equations with three non-apparent singularities. *SIGMA Symmetry Integrability Geom. Methods Appl.* **14** (2018), no. 058, 12 pp.

[29] K. Goncalves and K. Uhlenbeck, Moduli space theory for constant mean curvature surfaces immersed in space-forms. *Comm. Anal. Geom.* **15** (2007), 299–305.

[30] Z. Huang, J. Loftin, and M. Lucia, Holomorphic cubic differentials and minimal Lagrangian surfaces in $\mathbb{CH}^2$. *Math. Res. Lett.* **20** (2013), no. 3, 501–520.

[31] Z. Huang and M. Lucia, Minimal immersions of closed surfaces in hyperbolic three-manifolds. *Geom. Dedicata* **158** (2012), 397–411.

[32] Z. Huang, M. Lucia, and G. Tarantello, Bifurcation for minimal surface equation in hyperbolic 3-manifolds. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **38** (2021), no. 2, 243–279.

[33] Z. Huang, M. Lucia, and G. Tarantello, Donaldson Functional in Teichmüller Theory. (2021), submitted for publication.

[34] A. Jaffe and C. Taubes, *Vortices and monopoles, structure of static gauge theories*. Prog. Phys. 2, Boston–Basel–Stuttgart, Birkhäuser, 1980.

[35] Y. Lee, C. S. Lin, G. Tarantello, and W. Yang, Sharp estimates for solutions of mean field equations with collapsing singularity. *Comm. Partial Differential Equations* **42** (2017), no. 10, 1549–1597.

[36] Y. Lee, C. S. Lin, J. Wei, and W. Yang, Degree counting and shadow system for Toda system of rank two: one bubbling. *J. Differential Equations* **264** (2018), no. 7, 4343–4401.

[37] Y. Lee, C. S. Lin, W. Yang, and L. Zhang, Degree counting for Toda system with simple singularity: one point blow up. *J. Differential Equations* **268** (2020), no. 5, 2163–2209.

[38] Y. Y. Li, Harnack type inequality: the method of moving planes. *Comm. Math. Phys.* **200** (1999), no. 2, 421–444.

[39] Y. Y. Li and I. Shafrir, Blow-up analysis for solutions of $-\Delta u = Ve^u$ in dimension two. *Indiana Univ. Math. J.* **43** (1994), no. 4, 1255–1270.

[40] C. S. Lin, Z. Nie, and J. Wei, Toda systems and hypergeometric equations. *Trans. Amer. Math. Soc.* **370** (2018), no. 11, 7605–7626.

[41] C. S. Lin and G. Tarantello, When "blow-up" does not imply "concentration": a detour from Brezis–Merle's result. *C. R. Math. Acad. Sci. Paris* **354** (2016), no. 5, 493–498.

[42] C. S. Lin and C. L. Wang, Elliptic functions, Green functions and the mean field equations on tori. *Ann. of Math.* **2(172)** (2010), no. 2, 911–954.

[43] C. S. Lin, J. C. Wei, W. Yang, and L. Zhang, On rank-2 Toda systems with arbitrary singularities: local mass and new estimates. *Anal. PDE* **11** (2018), no. 4, 873–898.

[44] C. S. Lin, J. C. Wei, and L. Zhang, Classification of blowup limits for SU(3) singular Toda systems. *Anal. PDE* **8** (2015), no. 4, 807–837.

[45] C. S. Lin, J. Wei, and C. Zhao, Asymptotic behavior of SU(3) Toda system in a bounded domain. *Manuscripta Math.* **137** (2012), no. 1–2, 1–18.

[46] C. S. Lin, J. Wei, and C. Zhao, Sharp estimates for fully bubbling solutions of a SU(3) Toda system. *Geom. Funct. Anal.* **22** (2012), no. 6, 1591–1635.

[47] J. Liouville, Sur 'equation aux derivées partielles $\frac{\partial^2 \log \lambda}{\partial u \partial v} \pm \frac{\lambda}{2a^2} = 0$. *J. Math. Pures Appl.* **8** (1853), 71–72.

[48] F. Luo and G. Tian, Liouville equation and spherical convex polytopes. *Proc. Amer. Math. Soc.* **116** (1992), no. 4, 1119–1129.

[49] R. Mazzeo and H. Weiss, Teichmüller theory for conic surfaces. In *Geometry, Analysis and Probability*, pp. 127–164, Progr. Math. 310, Birkhäuser, 2017.

[50] R. Mazzeo and X. Zhu, Conical metrics on Riemann surfaces, I: the compactified configuration space and regularity. *Geom. Topol.* **24** (2020), 309–372.

[51] R. Mazzeo and X. Zhu, Conical Metrics on Riemann Surfaces, II: Spherical Metrics. *Int. Math. Res. Not.* **rnab011** (2021). DOI 10.1093/imrn/rnab011.

[52] G. Mondello and D. Panov, Spherical metrics with conical singularities on a 2-sphere: angle constraints. *Int. Math. Res. Not.* **16** (2016), 4937–4995.

[53] G. Mondello and D. Panov, Spherical surfaces with conical points: systole inequality and moduli spaces with many connected components. *Geom. Funct. Anal.* **29** (2019), no. 4, 1110–1193.

[54] Y. Nakayama and L. Field, Theory: a decade after the revolution. *Internat. J. Modern Phys. A* **19** (2004), no. 17, 18, 2771–2930.

[55] H. Ohtsuka and T. Suzuki, Blow-up analysis for Liouville type equation in self-dual gauge field theories. *Commun. Contemp. Math.* **7** (2005), no. 2, 177–205.

[56] C. T. Simpson, Constructing variations of Hodge structure using Yang–Mills theory and applications to uniformization. *J. Amer. Math. Soc.* **1** (1988), no. 4, 867–918.

[57] J. Song, Y. Cheng, B. Li, and B. Xu, Drawing cone spherical metrics via Strebel differentials. *Int. Math. Res. Not.* **11** (2020), 3341–3363.

[58] G. Tarantello, A Harnack inequality for Liouville-type equations with singular sources. *Indiana Univ. Math. J.* **54** (2005), no. 2, 599–615.

[59] G. Tarantello, *Selfdual gauge field vortices an analytical approach*. Progr. Nonlinear Differential Equations Appl. 72, Birkhäuser, Basel, 2008.

[60] G. Tarantello, Asymptotics for minimizers of the Donaldson functional in Teichmüller theory. (2021) Preprint.

[61] G. Tarantello, On the blow-up analysis at collapsing poles for solutions of singular Liouville type equations. (2021) Preprint.

[62] M. Troyanov, Metrics of constant curvature on a sphere with two conical singularities. In *Differential geometry (Peñiscola, 1988)*, pp. 296–306, Lecture Notes in Math. 1410, Springer, Berlin, 1989.

[63] M. Troyanov, Prescribing curvature on compact surfaces with conical singularities. *Trans. Amer. Math. Soc.* **324** (1991), no. 2, 793–821.

[64] K. Uhlenbeck, Closed minimal surfaces in hyperbolic 3-manifolds. In *Seminar on minimal submanifolds*, pp. 147–168, Ann. of Math. Stud. 103, Princeton Univ. Press, Princeton, NJ, 1983.

[65] M. Umehara and K. Yamada, Metrics of constant curvature 1 with three conical singularities on the 2-sphere. *Illinois J. Math.* **44** (2000), no. 1, 72–94.

[66] C. Voisin, *Hodge theory and complex algebraic geometry. I.* Cambridge Stud. Adv. Math., Cambridge, 2007.

[67] J. Wei and L. Zhang, Estimates for Liouville equation with quantized singularities. *Adv. Math.* **380** (2021).

[68] Y. Yang, *Solitons in field theory and nonlinear analysis*. Springer Monogr. Math., Springer, New York, 2001.

[69] X. Zhu, Spherical conic metrics and realizability of branched covers. *Proc. Amer. Math. Soc.* **147** (2019), no. 4, 1805–1815.

[70] X. Zhu, Rigidity of a family of spherical conical metrics. *New York J. Math.* **26** (2020), 272–284.

GABRIELLA TARANTELLO

Dipartimento di Matematica, Universita' di Roma "Tor Vergata", Via della Ricerca Scientifica 1, 00133 Roma, Italy, tarantel@mat.uniroma2.it

HYDRODYNAMIC STABILITY AT HIGH REYNOLDS NUMBER

DONGYI WEI AND ZHIFEI ZHANG

ABSTRACT

The hydrodynamic stability theory is mainly concerned with how laminar flows become unstable and transit to turbulence at high Reynolds number. To shed some light on the transition mechanism, Trefethen et al. [Science 261(1993)] proposed the transition threshold problem: *how much disturbance will lead to the instability of the flow and the dependence of disturbance on the Reynolds number*. Many effects such as 3D lift-up, inviscid damping, enhanced dissipation, and boundary layer play a crucial role in determining the transition threshold. In this note, we will first survey some important progress on linear inviscid damping and enhanced dissipation for shear flows. Then we will outline key ingredients in our proof of transition threshold for the 3D Couette flow in a finite channel.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 76E05; Secondary 35B25, 35B35, 35Q30, 76E30

KEYWORDS

Hydrodynamic stability, shear flow, inviscid damping, enhanced dissipation, transition threshold

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3902–3922
DOI 10.4171/ICM2022/31

1. INTRODUCTION

The hydrodynamic stability has been an active field in the fluid mechanics since Reynolds's experiment in 1883 [43]. This field focuses on how the laminar flows become unstable and transit to turbulence [20,46,57]. A fundamental model describing the motion of the incompressible fluid is the Navier–Stokes (NS) equations:

$$\begin{cases} \partial_t v - \nu \Delta v + v \cdot \nabla v + \nabla p = 0, \\ \nabla \cdot v = 0, \end{cases} \tag{1.1}$$

where $v = (v^1(t,x,y,z), v^2(t,x,y,z), v^3(t,x,y,z))$ is the velocity, $p(t,x,y,z)$ is the pressure, and $\nu = \mathrm{Re}^{-1} > 0$ (Re Reynolds number) is the viscosity coefficient. Let us recall some well-known laminar solutions of (1.1): the plane Couette flow $(y,0,0)$, the plane Poiseuille flow $(1-y^2,0,0)$, and the pipe Poiseuille flow $(0,0,1-r^2)$ with $r^2 = x^2 + y^2$. Our aim is to study the stability of these laminar flows at high Reynolds number, i.e., $\mathrm{Re} \gg 1$.

The plane Couette flow is spectrally stable for any Reynolds number $\mathrm{Re} \geq 0$ [44]. It has been a folklore conjecture that the pipe Poiseuille flow is spectrally stable for any Reynolds number. Recently, we (jointly with Chen) [15] proved that the pipe Poiseuille flow is spectrally stable at high Reynolds number. On the other hand, the experiments and numerics observed that these flows could be unstable and transit to turbulence for small but finite perturbations when the Reynolds number exceeds some critical number [13,22,42]. In addition, some laminar flows such as plane Poiseuille flow become turbulent at a much lower Reynolds number than that predicted by the eigenvalue analysis. These are the so-called Sommerfeld paradoxes. The resolution of these paradoxes is a long-standing problem in fluid mechanics. For many works dedicated to resolving these paradoxes, see [13] and references therein.

Trefethen et al. [48] provided an explanation about the linear instability via the ε-pseudospectra of the linearized NS operator $\mathcal{L}$ defined by

$$\sigma_\varepsilon(\mathcal{L}) = \left\{\lambda \in \mathbb{C} : \left\|(\lambda - \mathcal{L})^{-1}\right\| \geq \varepsilon^{-1}\right\}.$$

For the plane Couette flow, the spectrum of the linearized operator lies in the stable lower half-plane, but the pseudospectrum extends significantly into the upper half-plane. The pseudomode may be excited to a substantial amplitude by a very small input. This phenomenon is due to the nonnormality of the linear operator. Now the psuedospectrum has become an important concept in the study of nonnormal operators [47]. Li and Lin [35] provided a resolution from the following point of view: there is a sequence of linearly unstable shears which approach the linear shear in the kinetic energy norm but not in the enstrophy norm, and such linear instabilities offer an initiator for the transition from the linear shear to turbulence.

To shed some light on the transition mechanism to turbulence, Trefethen et al. [48] proposed the *transition threshold problem*: *how much disturbance will lead to the instability of the flow and the dependence of disturbance on the Reynolds number*. This idea may be traced back to Kelvin [30]. The following mathematical version was formulated by Bedrossian, Germain, and Masmoudi [7]:

Given a norm $\|\cdot\|_X$, find a $\beta = \beta(X)$ so that

$$\|u_0\|_X \leq \mathrm{Re}^{-\beta} \;\Rightarrow \text{stability},$$
$$\|u_0\|_X \gg \mathrm{Re}^{-\beta} \;\Rightarrow \text{instability}.$$

The exponent β is referred to as the transition threshold. It was conjectured in [48] that

> *"Notwithstanding these qualifications, we conjecture that transition to turbulence of eigenvalue-stable shear flows proceeds analogously to our model in that the destabilizing mechanism is essentially linear in the sense described above and the amplitude threshold for transition is $O(\mathrm{Re}^{\gamma})$ for some $\gamma < -1$."*

Later on, a lot of works were devoted to estimating β (see [13] and the references therein). To the best of our knowledge, the community never reached a consensus on what the thresholds should be. Numerical results by Lundbladh, Henningson, and Reddy [38] indicated that for the plane Couette flow, $\beta = 1$ for streamwise perturbation and $\beta = \frac{5}{4}$ for oblique perturbation; for the plane Poiseuille flow, $\beta = \frac{7}{4}$ for both streamwise and oblique perturbations. Asymptotic analysis results by Chapman [13] showed that for the plane Couette flow, $\beta = 1$ for streamwise and oblique perturbation; for the plane Poiseuille flow, $\beta = \frac{3}{2}$ for streamwise perturbation and $\beta = \frac{5}{4}$ for oblique perturbation.

In the absence of a physical boundary, Bedrossian, Germain, and Masmoudi (BGM) made important progress on the transition threshold problem for the 3D Couette flow in a series of works [5, 6, 8]. It was shown that $\beta \leq 1$ for the perturbations in Gevrey class and $\beta \leq \frac{3}{2}$ for the perturbations in Sobolev space. In [52], we improved the result of [6] to $\beta \leq 1$ in Sobolev space, which means that the regularity of the initial data (at least above H^2-regularity) does not play an important role in determining the transition threshold. In the presence of a physical boundary, the boundary layer could affect the stability of the flow at the high Reynolds number regime. To understand the boundary layer effect, we (jointly with Chen and Li) [14] studied the transition threshold problem for the 2D Couette flow in a finite channel $\mathbb{T} \times [-1, 1]$. We established various space–time estimates for the linearized NS system by developing the robust resolvent estimate method. Based on this work and [52], we (jointly with Chen) [16] proved that the transition threshold $\beta \leq 1$ in the Sobolev space for the 3D Couette flow in a finite channel $\mathbb{T} \times [-1, 1] \times \mathbb{T}$. Therefore, the transition threshold for the 3D Couette flow is inconsistent with the value (some $\beta > 1$) conjectured in [48] even in the presence of the boundary layer effect. The main reason may be that the infinite-dimensional mixing effects and special null structures in the nonlinearity suppress most of the nonlinear interactions rather than giving what could be predicted by the toy model in [48].

Both BGM's and our works show that these linear effects, namely 3D lift-up, inviscid damping, enhanced dissipation, and boundary layer, play a crucial role in determining the transition from a laminar to turbulent flow at high Reynolds number. In this note, we will first survey some recent important progress about linear inviscid damping and enhanced dissipation for shear flows. Then we will outline some key ingredients in our proof of transition threshold for the 3D Couette flow in a finite channel.

2. LINEAR INVISCID DAMPING FOR SHEAR FLOWS

We consider the 2D linearized Euler equation around shear flow $(u(y), 0)$ in a finite channel $\Omega = \{(x, y) : x \in \mathbb{T}, y \in [-1, 1]\}$:

$$\partial_t \omega + \mathcal{L}\omega = 0, \quad \omega|_{t=0} = \omega_0(x, y), \tag{2.1}$$

where $\mathcal{L} = u(y)\partial_x + u''(y)\partial_x(-\Delta)^{-1}$ and ω is the vorticity. Taking the Fourier transform with respect to x, the linearized Euler equation (2.1) in terms of the stream function ψ (i.e., $\Delta\psi = \omega$) is reduced to

$$\partial_t \widehat{\psi} + i\alpha\mathcal{R}_\alpha\widehat{\psi} = 0, \tag{2.2}$$

where $\mathcal{R}_\alpha\widehat{\psi} = -(\partial_y^2 - \alpha^2)^{-1}(u''(y) - u(\partial_y^2 - \alpha^2))\widehat{\psi}$.

For the Couette flow (i.e., $u(y) = y$), Orr [41] observed an important phenomenon that the velocity will tend to 0 as $t \to \infty$, although the Euler equation is a conserved system. This phenomenon is the so-called *inviscid damping*, which is the analogue in hydrodynamics of Landau damping [32]; see [45] for similar phenomena in various systems. For general shear flows, the problem is challenging due to the presence of the nonlocal operator $u''(y)\partial_x(-\Delta)^{-1}$. In this case, the linear dynamics is associated with the singularities at the critical layer $u = c$ of the solution of the Rayleigh equation

$$(u - c)(\Phi'' - \alpha^2\Phi) - u''\Phi = f.$$

Based on the Laplace transform and singularity analysis of the solution ϕ at the critical layer, Case [12] gave the first prediction of linear damping for monotone shear flows. However, Case's argument does not work for nonmonotone flows. Bouchet and Morita [11] may be the first to study the linear damping for nonmonotone shear flows. Based on Laplace tools and numerical computations, they found a new dynamic mechanism, i.e., *vorticity depletion phenomena*. Assume that for large time

$$\widehat{\omega}(t, \alpha, y) \sim \omega_\infty(y)\exp\bigl(-i\alpha u(y)t\bigr) + O(t^{-\gamma}).$$

The vorticity depletion means that $\omega_\infty(y)$ vanishes at stationary points of $u(y)$. This is another important mechanism leading to the damping for nonmonotone shear flows. Based on this observation and using stationary phase expansion, they predicted similar decay rates of the velocity as in the monotone case.

In a series of works [53–55], we (jointly with Zhao) confirmed Case's prediction on linear damping for monotone shear flows and Bouchet–Morita's prediction for nonmonotone shear flows, including Poiseuille and Kolmogorov flows. Let us review these results. The first result is the linear inviscid damping for monotone flows [53].

Theorem 2.1. *Let $u(y) \in C^4([0, 1])$ be a monotone function. Suppose that the linearized operator $\mathcal{L}$ has no embedding eigenvalues. Assume that $\int_T \omega_0(x, y)dx = 0$ and $P_{\mathcal{L}}\omega_0 = 0$, where $P_{\mathcal{L}}$ is the spectral projection to $\sigma_d(\mathcal{L})$. Then it holds that*

1. *If $\omega_0(x, y) \in H_x^{-1}H_y^1$, then*

$$\|V(t)\|_{L^2} \le \frac{C}{\langle t\rangle}\|\omega_0\|_{H_x^{-1}H_y^1};$$

2. *If $\omega_0(x,y) \in H_x^{-1}H_y^2$, then*

$$\|V^2(t)\|_{L^2} \le \frac{C}{\langle t\rangle^2}\|\omega_0\|_{H_x^{-1}H_y^2}.$$

Now we introduce a class of nonmonotone flows denoted by $\mathcal{K}$, which consists of the functions $u(y)$ satisfying $u(y) \in H^3(-1,1)$ and $u''(y) \neq 0$ for critical points (i.e., $u'(y)=0$) and $u'(\pm 1) \neq 0$. For the flows in $\mathcal{K}$, we prove the following linear inviscid damping result and confirm the vorticity depletion phenomenon [54].

Theorem 2.2. *Assume that $u(y) \in \mathcal{K}$ and the linearized operator $\mathcal{R}_\alpha$ has no embedding eigenvalues. Assume that $\widehat{\omega}_0(\alpha,y) \in H_y^1(-1,1)$ and $P_{\mathcal{R}_\alpha}\widehat{\psi}_0(\alpha,y)=0$, where ψ_0 is the stream function and $P_{\mathcal{R}_\alpha}$ is the spectral projection to $\sigma_d(\mathcal{R}_\alpha)$. Then it holds that*

$$\|\hat{V}(\cdot,\alpha,\cdot)\|_{L_t^2L_y^2} + \|\partial_t\hat{V}(\cdot,\alpha,\cdot)\|_{L_t^2L_y^2} \le C_\alpha\|\widehat{\omega}_0(\alpha,\cdot)\|_{H_y^1}.$$

In particular, $\lim_{t\to+\infty}\|\hat{V}(t,\alpha,\cdot)\|_{L_y^2}=0$. *If $u'(y_0)=0$, then*

$$\lim_{t\to+\infty}\widehat{\omega}(t,\alpha,y_0)=0.$$

Remark 2.1. For a class of symmetric shear flows, including the Poiseuille and Kolmogorov flows, we can obtain the explicit decay estimates as in the monotone case [54, 55]. A very interesting question is to prove the explicit decay estimates for general flows in $\mathcal{K}$.

The proof of Theorem 2.1 is based on the representation formula of the solution. Let Ω_ϵ be a simply connected domain including the spectrum $\sigma(\mathcal{R}_\alpha)$ of $\mathcal{R}_\alpha$. Then the solution $\widehat{\psi}(t,\alpha,y)$ is given by the following Dunford integral:

$$\widehat{\psi}(t,\alpha,y) = \frac{1}{2\pi i}\int_{\partial\Omega_\epsilon} e^{-i\alpha tc}(c-\mathcal{R}_\alpha)^{-1}\widehat{\psi}(0,\alpha,y)dc.$$

Let $\Phi(\alpha,y,c)$ be the solution of the inhomogeneous Rayleigh equation with $f(\alpha,y,c)=\frac{\widehat{\omega_0}(\alpha,y)}{i\alpha(u-c)}$ and $c \in \Omega_\epsilon$:

$$\Phi'' - \alpha^2\Phi - \frac{u''}{u-c}\Phi = f, \quad \Phi(-1)=\Phi(1)=0. \tag{2.3}$$

Then we find that

$$(c-\mathcal{R}_\alpha)^{-1}\widehat{\psi}(0,\alpha,y) = i\alpha\Phi(\alpha,y,c).$$

Therefore, we have

$$\widehat{\psi}(t,\alpha,y) = \frac{1}{2\pi}\int_{\partial\Omega_\epsilon}\alpha\Phi(\alpha,y,c)e^{-i\alpha ct}dc. \tag{2.4}$$

Thus, the key ingredient of the proof is reduced to solving the inhomogeneous Rayleigh equation (2.3) and deriving uniform estimates of the solution Φ in ϵ. For this, we need to construct two independent solutions to the homogeneous Rayleigh equation for $c \in \Omega_\epsilon$:

$$\phi'' - \alpha^2\phi - \frac{u''}{u-c}\phi = 0.$$

Our idea is as follows. Let $\phi = (u(y) - c)\phi_1$. Then ϕ_1 satisfies

$$\big((u(y) - c)^2 \phi_1'\big)' = \alpha^2 \phi_1 \big(u(y) - c\big)^2.$$

If $\phi_1(y_c, c) = 1$ and $\phi_1'(y_c, c) = 0$ at y_c, then we have

$$\begin{aligned}\phi_1(y, c) &= 1 + \int_{y_c}^{y} \frac{\alpha^2}{(u(y') - c)^2} \int_{y_c}^{y'} \phi_1(z, c)\big(u(z) - c\big)^2 dz dy' \\ &= 1 + \alpha^2 T\phi_1(y, c).\end{aligned}$$

Assume that u is monotone and let $y_c = u^{-1}(c_r)$ with $c_r = \mathrm{Re} c$. The following estimate is crucial: there exists a constant C independent of A so that

$$\left\| \frac{Tf(y, c)}{\cosh A(y - y_c)} \right\|_{L^\infty_{y,c}} \leq \frac{C}{A^2} \left\| \frac{f(y, c)}{\cosh A(y - y_c)} \right\|_{L^\infty_{y,c}}.$$

Then $\phi_1(y, c) = \sum_{k=0}^{\infty} (\alpha^2 T)^k (1)$ by taking A large enough.

The proof of Theorem 2.2 is based on the limiting absorption principle. Consider the inhomogeneous Rayleigh equation:

$$(u - c)(\Phi'' - \alpha^2 \Phi) - u''\Phi = \omega, \quad \Phi(-1) = \Phi(1) = 0,$$

where $c \in \Omega \setminus D_0$, $D_0 = \mathrm{Ran}\, u$. Using blow-up analysis and a compactness argument, we prove the limiting absorption principle for shear flows $u \in \mathcal{K}$.

Proposition 2.1. *If $\mathcal{R}_\alpha$ has no embedding eigenvalues, then there exists an ϵ_0 such that for $c \in \Omega_{\epsilon_0} \setminus D_0$, Φ has the the following uniform bound:*

$$\|\Phi\|_{H^1(-1,1)} \leq C \|\omega\|_{H^1(-1,1)}.$$

Here C is a constant independent of ϵ_0. Moreover, there exists $\Phi_\pm(\alpha, y, c) \in H^1_0(-1, 1)$ for $c \in \mathrm{Ran}\, u$, such that $\Phi(\alpha, \cdot, c \pm i\epsilon) \to \Phi_\pm(\alpha, \cdot, c)$ in $C([-1, 1])$ as $\epsilon \to 0+$ and

$$\big\|\Phi_\pm(\alpha, \cdot, c)\big\|_{H^1(-1,1)} \leq C \|\omega\|_{H^1(-1,1)}.$$

From (2.4) and Plancherel's formula, we infer that

$$\begin{aligned}\big\|\hat{V}(t, \alpha, y)\big\|^2_{H^1_t L^2_y} &= \int_{\mathbb{R}} \big(\big\|\hat{V}(t, \alpha, \cdot)\big\|^2_{L^2_y} + \big\|\partial_t \hat{V}(t, \alpha, \cdot)\big\|^2_{L^2_y}\big) dt \\ &\leq C \int_{\mathrm{Ran}\, u} \big\|\widetilde{\Phi}(\alpha, \cdot, c)\big\|^2_{H^1_y} dc \leq C \big\|\widehat{\omega}_0(\alpha, \cdot)\big\|^2_{H^1_y}.\end{aligned}$$

For monotone shear flows, we (jointly with Zhu) also developed the vector field method in the sprit of wave equation [56]. The idea is as follows. We first proved the space–time estimate of the velocity via the limiting absorption principle. Consider

$$\partial_t \omega + i\alpha \mathcal{R}'_\alpha \omega = f, \quad \mathcal{R}'_\alpha \omega = -\big(u''\big(\partial_y^2 - \alpha^2\big)^{-1} - u\big)\omega.$$

Using the limiting absorption principle, we can prove that

$$\begin{aligned}&\big\|\omega(T)\big\|^2_{L^2} + \alpha^2 \int_0^T \big(\big\|\partial_y \psi(t)\big\|^2_{L^2} + \alpha^2 \big\|\psi(t)\big\|^2_{L^2}\big) dt \\ &\quad \leq C \big\|\omega(0)\big\|^2_{L^2} + C\alpha^{-2} \int_0^T \big(\big\|\partial_y f(t)\big\|^2_{L^2} + \alpha^2 \big\|f(t)\big\|^2_{L^2}\big) dt = RHS. \qquad (2.5)\end{aligned}$$

Moreover, if $f(t,0) = f(t,1) = 0$, then we also have

$$\alpha \int_0^T \left(\left|\partial_y \psi(t,0)\right|^2 + \left|\partial_y \psi(t,1)\right|^2\right) dt \leq RHS. \tag{2.6}$$

Then we introduce the vector field $X = (1/u')\partial_y + i\alpha t$, which commutes with $\partial_t + i\alpha u$. We denote

$$\omega_1 = X\omega, \quad \psi_2 = -\left(\partial_y^2 - \alpha^2\right)^{-1}\left(\partial_y \omega / u'\right), \quad \psi_3 = \psi_2 - \partial_y \psi / u'.$$

Then we have

$$\partial_t \omega_1 + i\alpha \mathcal{R}'_{\alpha,\beta}\omega_1 = -i\alpha(u'''/u')\psi + i\alpha u'' \psi_3.$$

Based on the space–time estimate (2.5) and (2.6), we can obtain a uniform estimate for $\|X\omega\|_{L^2}$, which implies that $\|V(t)\|_{L^2} \leq C\langle t\rangle^{-1}$. More work is needed to prove $\|V^2(t)\|_{L^2} \leq C\langle t\rangle^{-2}$. See Section 2 in [56] for the details.

Finally, let us mention some recent important results on linear inviscid damping [4,23,58,59] and nonlinear inviscid damping [10,19,27–29,37,39]. However, when the boundary effect is involved, nonlinear inviscid damping is still a challenging problem [58].

3. LINEAR ENHANCED DISSIPATION FOR KOLMOGOROV FLOW

Let us first consider the diffusion–convection equation in $\mathbb{T} \times \mathbb{R}$:

$$\partial_t \omega - \nu \Delta \omega + y \partial_x \omega = 0.$$

Introduce new variables $(\bar{x}, y) = (x - ty, y)$ and set $\widetilde{\omega}(t, \bar{x}, y) = \omega(t, x, y)$. Then the solution $\widehat{\widetilde{\omega}}(t,k,\xi) = \int_{\mathbb{T}\times\mathbb{R}} \widetilde{\omega}(t,x,y) e^{-2\pi ikx - i2\pi\xi y} dx dy$ takes the form

$$\widehat{\widetilde{\omega}}_{\neq}(t,k,\xi) = e^{-\nu(2\pi)^2 \int_0^t (k^2 + (\xi - k\tau)^2) d\tau} \widehat{\omega}_{\neq}(0,k,\xi).$$

Due to $\int_0^t (k^2 + (\xi - k\tau)^2) d\tau \geq k^2 t^3/12$, we deduce that

$$\left\|\omega_{\neq}(t)\right\|_{L^2} \leq e^{-c\nu t^3} \left\|\omega_{\neq}(0)\right\|_{L^2} \leq C e^{-c\nu^{1/3} t} \left\|\omega_{\neq}(0)\right\|_{L^2}.$$

Here the exponent νt^3 gives a dissipation time scale $\nu^{-1/3}$, which is much shorter than the dissipation time scale ν^{-1}. We refer to this phenomenon as the *enhanced dissipation*, which is also due to the mixing mechanism.

We are concerned with the enhanced dissipation phenomenon for the linearized Navier–Stokes equations around shear flows. In this note, we will review some progress on the enhanced dissipation estimates for the linearized 2D NS equations in the torus $\mathbb{T}_{2\pi\delta} \times \mathbb{T}_{2\pi}$ around the Kolmogorov flow $(-e^{-\nu t} \cos y, 0)$, which is a solution of the 2D NS equations:

$$\partial_t \omega + \mathcal{L}_\nu(t)\omega = 0, \quad \omega|_{t=0} = \omega_0(x,y), \tag{3.1}$$

where $\mathcal{L}_\nu(t) = -\nu\Delta - e^{-\nu t} \cos y \partial_x (1 + \Delta^{-1})$. Beck and Wayne [2] considered the following model equation by removing the nonlocal part Δ^{-1} of $\mathcal{L}_\nu(t)$:

$$\partial_t \omega - \nu \Delta \omega - e^{-\nu t} \cos y \partial_x \omega = 0.$$

Using the hypocoercivity method in [49], they proved the enhanced dissipation rate of the solution in some Banach space X (see (3.7) in [2]): for any $t \in [0, \tau/\nu]$,

$$\|\omega(t)\|_X \le Ce^{-M\sqrt{\nu t}}\|\omega_0\|_X.$$

Based on numerical results, Beck and Wayne [2] conjectured that the same decay result should hold for $\mathcal{L}_\nu(t)$. In a series of works [34,51,55], we have developed three approaches to solve this conjecture: resolvent estimate method, wave operator method, and hypocoercivity method.

In [51], by developing the hypocoercivity method from [2], we proved the following enhanced dissipation results.

Theorem 3.1. *Given $\delta \in (0,1)$ and $\tau > 0$, there exist constants $c_1 > 0$, $C > 0$ such that if ω satisfies (3.1) with $\omega_0 \in L^2$ and $\int_{\mathbb{T}_{2\pi\delta}} \omega_0(x,y)dx = 0$, then it holds that, for $0 < t \le \tau/\nu$,*

$$\|\omega(t)\|_{L^2} \le Ce^{-c_1\sqrt{\nu t}}\|\omega_0\|_{L^2},$$
$$\|V(t)\|_{\dot{H}^1_x L^2_y} \le \frac{Ce^{-c_1\sqrt{\nu t}}}{\sqrt{1+\nu t^3}}\|\omega_0\|_{L^2}.$$

When $\delta = 1$, it holds that, for $0 < t \le \tau/\nu$,

$$\|(I-P_1)\omega(t)\|_{L^2} \le Ce^{-c_1\sqrt{\nu t}}\|(I-P_1)\omega_0\|_{L^2},$$
$$\|(I-P_1)V(t)\|_{\dot{H}^1_x L^2_y} \le \frac{Ce^{-c_1\sqrt{\nu t}}}{\sqrt{1+\nu t^3}}\|(I-P_1)\omega_0\|_{L^2}.$$

Here P_1 is the orthogonal projection to the space W_1 spanned by $\{\cos x, \sin x\}$.

Remark 3.1. Here the enhanced dissipation rate is smaller than that for the Couette flow. This leads to conjecture that, for stable monotone shear flows to the Euler equations, the enhanced dissipation rate should be $\nu^{\frac{1}{3}}$, and the rate should be $\nu^{\frac{1}{2}}$ for stable shear flows with nondegenerate critical points.

Remark 3.2. In addition to the important application to the transition threshold problem [34,36,55], the enhanced dissipation also plays an important role for the suppression of blow-up in the Keller–Segel system [9,24,31] and axisymmetrization of 2D viscous vortices [21]. Let us refer to [3,17,18,23] and the references therein for more relevant works.

Taking Fourier transform with respect to x to (3.1), we obtain

$$\partial_t\widehat{\omega} + \mathcal{L}_\nu(\alpha,t)\widehat{\omega} = 0, \quad \mathcal{L}_\nu = \nu\big(-\partial_y^2 + \alpha^2\big) - i\alpha e^{-\nu t}\cos y\big(1 + \big(\partial_y^2 - \alpha^2\big)^{-1}\big).$$

We write

$$A = \sin y\big(1 + \big(\partial_y^2 - \alpha^2\big)^{-1}\big), \quad B = \cos y\big(1 + \big(\partial_y^2 - \alpha^2\big)^{-1}\big), \quad \gamma(t) = \alpha e^{-\nu t}.$$

Next we introduce an important inner product structure

$$\langle u, w\rangle_* = \big\langle u, w - \big(\alpha^2 - \partial_y^2\big)^{-1}w\big\rangle.$$

An important observation is that under this inner product, the operators A and B are symmetric, i.e.,

$$\langle u, Aw\rangle_* = \langle Au, w\rangle_*, \quad \langle u, Bw\rangle_* = \langle Bu, w\rangle_*.$$

Moreover, for $|\alpha| > 1$, the norm $\|u\|_* = \langle u, u\rangle_*^{\frac{1}{2}}$ is equivalent to the usual L^2-norm:

$$(1-\alpha^{-2})\|u\|_{L^2}^2 \le \|u\|_*^2 \le \|u\|_{L^2}^2.$$

We introduce the energy functional:

$$E_0(t) = \|\widehat{\omega}(t)\|_*^2, \quad E_1(t) = \|\partial_y\widehat{\omega}(t)\|_*^2, \quad E_2(t) = \|\partial_y^2\widehat{\omega}(t)\|_*^2,$$
$$\mathcal{E}_1(t) = \mathrm{Re}\big\langle \partial_y\widehat{\omega}(t), iA\widehat{\omega}(t)\big\rangle_*, \quad \mathcal{E}_2(t) = \|\widehat{\omega}(t)\|_*^2 - \|B\widehat{\omega}(t)\|_*^2.$$

Then we construct the total energy functional as follows:

$$\Phi(t) = E_0(t) + \alpha_0 \nu t E_1(t) + \beta_0 \nu t^2 \mathcal{E}_1(t) + \gamma_0 \nu t^3 \mathcal{E}_2(t)$$

with the constants α_0, β_0, γ_0 depending on $\gamma(0)$ so that

$$\Phi'(t) \le -c\big|\gamma(0)\big|^2 \nu^2 t^3 E_0(t), \quad \Phi(t) \ge E_0(t).$$

Then the bound

$$E_0(t) \le \big(1 + c_2\big|\gamma(0)\big|^2 \nu^2 t^4\big)^{-1} E_0(0)$$

follows from the fact that $E_0(t)$ is decreasing in t. Once the polynomial decay is obtained, the exponential decay can be proved by iteration. Compared with [2], the key difference is that we introduce the new inner product and time dependent weights. This modification is also very effective in removing the logarithmic loss in [2] when achieving the dissipation in the usual L^2-norm.

In [55], we (jointly with Zhao) used the wave operator method. This idea was first introduced in [33] to study the pseudospectral bound of the Oseen vortices operator. The aim is to construct a wave operator $\mathbb{D}$ so that

$$\mathbb{D}\cos y\big(1 + \big(\partial_y^2 - \alpha^2\big)^{-1}\big)\omega = \cos y\mathbb{D}\omega.$$

Then $w = \mathbb{D}\omega$ satisfies

$$\partial_t w - \nu\big(\partial_y^2 - \alpha^2\big)w - i\alpha e^{-\nu t}\cos y w = -\nu\big[\partial_y^2, \mathbb{D}\big]\omega.$$

Moreover, the wave operator $\mathbb{D}$ we constructed has the following important properties:

- $\|\mathbb{D}(\omega)\|_{L^2}^2 = \langle \omega, \omega + (\partial_y^2 - \alpha^2)^{-1}\omega\rangle$;
- There exists a constant C independent of α so that

$$\big\|\sin y\mathbb{D}(\omega)\big\|_{L^2}^2 \ge \|\partial_y\psi\|_{L^2}^2 + (\alpha^2 - 1)\|\psi\|_{L^2}^2,$$
$$\big\|\partial_y\mathbb{D}(\omega)\big\|_{L^2} \le C|\alpha|^{\frac{1}{2}}\|\omega\|_{H^1},$$
$$\big\|\partial_y^2\mathbb{D}(\omega)\big\|_{L^2} \le C|\alpha|^{\frac{3}{2}}\|\omega\|_{H^2},$$

  where $-(\partial_y^2 - \alpha^2)\psi = \omega$;

- Commutator estimate holds:

$$\left\|\sin y\left[\partial_y^2, \mathbb{D}\right]\omega\right\|_{L^2} \leq C\left(|\alpha|\|\omega\|_{L^2} + \|\partial_y\omega\|_{L^2}\right).$$

The construction of the wave operator was motivated by our study of linear inviscid damping. More precisely, we may write the solution of (2.2) in the form

$$e^{-i\alpha t\mathcal{R}_\alpha}\psi_0 = \frac{1}{2\pi i}\int_{\operatorname{Ran} u} e^{-i\alpha tc}\Gamma(y,c)\tilde{\mathbb{D}}[\omega_0](c)dc.$$

An important observation is that

$$-\tilde{\mathbb{D}}\left[\left(\partial_y^2 - \alpha^2\right)e^{-i\alpha t\mathcal{R}_\alpha}\psi_0\right](c) = e^{-i\alpha tc}\tilde{\mathbb{D}}[\omega_0](c).$$

Taking the time derivative at $t = 0$, we get

$$-\tilde{\mathbb{D}}\left[\left(\partial_y^2 - \alpha^2\right)\mathcal{R}_\alpha\psi_0\right](c) = c\tilde{\mathbb{D}}[\omega_0](c),$$

which implies, by taking $c = u(y)$, that

$$\mathbb{D}\left[u\omega_0 + u''\psi_0\right] = u(y)\mathbb{D}[\omega_0].$$

Here $u(y) = -\cos y$, $\mathbb{D}[\omega_0](y) = \Lambda_o(y)\tilde{\mathbb{D}}[\omega_0](u(y))$ if ω_0 is odd and $\mathbb{D}[\omega_0](y) = \Lambda_e(y)\tilde{\mathbb{D}}[\omega_0](u(y))$ if ω_0 is even. See Section 2.2 in [55] for the details.

In [34], we (jointly with Li) used the resolvent estimate method developed in [33] to prove the enhanced dissipation estimates for the linearized NS equations with time-independent coefficient:

$$\partial_t\omega + \mathcal{L}_\nu\omega = 0, \quad \mathcal{L}_\nu = -\nu\Delta - \sin y\partial_x(1 + \Delta^{-1}).$$

The key ingredient is to establish the following resolvent estimate: given $0 < \nu \leq 1$ and $|\beta| > 1$, there exists a constant $C > 0$, independent of ν, λ, β, such that

$$\left\|(L_\nu - i\lambda)w\right\|_{L^2} \geq C\nu^{\frac{1}{2}}|\beta|^{\frac{1}{2}}(1 - \beta^{-2})\|w\|_{L^2}, \tag{3.2}$$

where $L_\nu w = -\nu\partial_y^2 w + i\beta(\sin yw + \sin y\varphi)$ with $(\partial_y^2 - \beta^2)\varphi = w$.

In order to deduce the semigroup bound from (3.2), we use Gearhart–Prüss-type lemma for an m-accretive operator proved by the first author [50].

Lemma 3.2. *Let H be an m-accretive operator in a Hilbert space X. Then it holds that, for any $t \geq 0$,*

$$\left\|e^{-tH}\right\| \leq e^{-t\Psi + \pi/2},$$

where $\Psi(H) = \inf\{\|(H - i\lambda)u\|; u \in D(H),\ \lambda \in \mathbb{R},\ \|u\| = 1\}$.

Now the operator L_ν is m-accretive with respect to the new inner product $\langle\cdot,\cdot\rangle_*$. From (3.2), we infer that $\Psi(L_\nu) \geq c\nu^{\frac{1}{2}}|\beta|^{\frac{1}{2}}(1 - \beta^{-2})$. Then it follows from Lemma 3.2 that $\|e^{-tL_\nu}\| \leq Ce^{-t\nu^{\frac{1}{2}}}$. In [26], the authors also derive the semigroup bound via establishing the pseudospectral bound of the linearized operator.

Next we give a simple sketch of the proof of (3.2). Notice that

$$(L_\nu - i\beta\lambda)w = -\nu\partial_y^2 w + i\beta\left(\sin y(w + \varphi) - \lambda w\right).$$

We introduce $u = w + \varphi$. Then it suffices to show that

$$\|\mathcal{L}_\lambda u\|_{L^2} \geq C|\nu\beta|^{\frac{1}{2}}\|u\|_{L^2},$$

where $(\partial_y^2 - \tilde{\beta}^2)\varphi = u$ with $\beta^2 - 1 = \tilde{\beta}^2$ and

$$\mathcal{L}_\lambda u = i\beta\big[(\sin y - \lambda)u + \lambda\varphi\big] - \nu\partial_y^2 u.$$

Consider the case of $\lambda > 1$. Integration by parts gives

$$\big|\mathrm{Im}\langle\mathcal{L}_\lambda u, u\rangle\big| = \beta\left(\int_0^{2\pi}(\lambda - \sin y)|u|^2 dy + \lambda\|\varphi'\|_{L^2}^2 + \lambda\tilde{\beta}^2\|\varphi\|_{L^2}^2\right),$$

which implies

$$\int_0^{2\pi}(\lambda - \sin y)|u|^2 dy + \lambda\|\varphi'\|_{L^2}^2 + \lambda\tilde{\beta}^2\|\varphi\|_{L^2}^2 \leq \beta^{-1}\|\mathcal{L}_\lambda u\|_{L^2}\|u\|_{L^2}.$$

Let $\delta \in (0, 1]$. Then we have

$$\begin{aligned}\|u\|_{L^2}^2 \leq \|u\|_{L^2(\frac{\pi}{2}+\delta,\frac{5\pi}{2}-\delta)}^2 + 2\delta\|u\|_{L^\infty}^2 &\lesssim \delta^{-2}\int_0^{2\pi}(\lambda - \sin y)|u|^2 dy + \delta\|u\|_{L^\infty}^2 \\ &\lesssim \beta^{-1}\delta^{-2}\|\mathcal{L}_\lambda u\|_{L^2}\|u\|_{L^2} \\ &\quad + \nu^{-\frac{1}{2}}\delta\|\mathcal{L}_\lambda u\|_{L^2}^{\frac{1}{2}}\|u\|_{L^2}^{\frac{3}{2}} + \delta\|u\|_{L^2}^2.\end{aligned}$$

Here we used the fact that

$$\|u\|_{L^\infty} \leq \|u'\|_{L^2}^{\frac{1}{2}}\|u\|_{L^2}^{\frac{1}{2}} + \|u\|_{L^2} \leq \nu^{-\frac{1}{4}}\|\mathcal{L}_\lambda u\|_{L^2}^{\frac{1}{4}}\|u\|_{L^2}^{\frac{3}{4}} + \|u\|_{L^2},$$

due to $\nu\|u'\|_{L^2}^2 = |\mathrm{Re}\langle\mathcal{L}_\lambda u, u\rangle|$. Taking $\delta = \beta^{-\frac{1}{4}}\nu^{\frac{1}{4}} \ll 1$, we infer

$$\|u\|_{L^2}^2 \lesssim (\beta\nu)^{-\frac{1}{2}}\|\mathcal{L}_\lambda u\|_{L^2}\|u\|_{L^2} + (\beta\nu)^{-\frac{1}{4}}\|\mathcal{L}_\lambda u\|_{L^2}^{\frac{1}{2}}\|u\|_{L^2}^{\frac{3}{2}},$$

which implies that

$$\|\mathcal{L}_\lambda u\|_{L^2} \gtrsim |\beta\nu|^{\frac{1}{2}}\|u\|_{L^2}.$$

The case of $|\lambda| < 1$ is much more difficult. Let $0 \leq y_1 \leq \frac{\pi}{2} \leq y_2 \leq \pi$ so that $\lambda = \sin y_1 = \sin y_2$. Let $\delta = \beta^{-\frac{1}{4}}\nu^{\frac{1}{4}} \ll 1$. Then we need to consider the following four types of energy estimates:

$$\begin{aligned}&\mathrm{Im}\langle\mathcal{L}_\lambda u, \chi_{(y_1,y_2)}u\rangle, \quad \mathrm{Im}\left\langle\mathcal{L}_\lambda u, \chi_{(y_1+\delta,y_2-\delta)}\frac{u}{\sin y - \lambda}\right\rangle, \\ &\mathrm{Im}\langle\mathcal{L}_\lambda u, \chi_{(y_2,y_1+2\pi)}u\rangle, \quad \mathrm{Im}\left\langle\mathcal{L}_\lambda u, \chi_{(y_2+\delta,y_1+2\pi-\delta)}\frac{u}{\sin y - \lambda}\right\rangle.\end{aligned}$$

See Section 3 in [34] for the details.

4. TRANSITION THRESHOLD PROBLEM FOR THE 3D COUETTE FLOW

We consider the transition threshold problem for the 3D Couette flow $U_*(y) = (y,0,0)$ in a finite channel $\Omega = \mathbb{T}\times[-1,1]\times\mathbb{T}$. We introduce the perturbation $u(t,x,y,z) = v(t,x,y,z) - U_*(y)$, which solves

$$\begin{cases}\partial_t u - \nu\Delta u + y\partial_x u + (u^2,0,0) + \nabla p^L + u\cdot\nabla u + \nabla p^{NL} = 0,\\ \nabla\cdot u = 0,\\ u(t,x,\pm 1,z) = 0, \quad u(0,x,y,z) = u_0(x,y,z).\end{cases} \tag{4.1}$$

Here the pressure p^L and p^{NL} are determined by

$$\begin{cases}\Delta p^L = -2\partial_x u^2,\\ \Delta p^{NL} = -\operatorname{div}(u\cdot\nabla u) = -\partial_i u^j\partial_j u^i,\\ \big(\partial_y p^L - \nu\Delta u^2\big)|_{y=\pm 1} = 0, \quad \partial_y p^{NL}|_{y=\pm 1} = 0.\end{cases} \tag{4.2}$$

We define

$$P_0 f = \bar{f} = \frac{1}{2\pi}\int_{\mathbb{T}} f(x,y,z)dx, \quad P_{\neq} f = f_{\neq} = f - P_0 f.$$

In [16], we prove the following stability result, which implies that the transition threshold $\beta \le 1$ for the 3D Couette flow in a finite channel.

Theorem 4.1. *Assume that* $u_0 \in H_0^1(\Omega)\cap H^2(\Omega)$ *with* $\operatorname{div} u_0 = 0$. *There exist constants* $\nu_0, c_0, \epsilon, C > 0$, *independent of* ν, *so that if* $\|u_0\|_{H^2} \le c_0\nu$, $0 < \nu \le \nu_0$, *then the solution* u *of the system* (4.1) *is global in time and satisfies the following stability estimates:*

- *(Uniform bounds and decay of the background streak)*

$$\begin{aligned}&\big\|\bar{u}^1(t)\big\|_{H^2} + \big\|\bar{u}^1(t)\big\|_{L^\infty} \le C\nu^{-1}\min(\nu t + \nu^{2/3}, e^{-\nu t})\|u_0\|_{H^2},\\ &\big\|\bar{u}^2(t)\big\|_{H^2} + \big\|\bar{u}^3(t)\big\|_{H^1} + \big\|(\bar{u}^2,\bar{u}^3)(t)\big\|_{L^\infty} \le Ce^{-\nu t}\|u_0\|_{H^2};\end{aligned}$$

- *(Rapid convergence to a streak)*

$$\begin{aligned}&\big\|(\partial_x,\partial_z)\partial_x u_{\neq}(t)\big\|_{L^2} + \big\|(\partial_x,\partial_z)\nabla u^2_{\neq}(t)\big\|_{L^2} + \big\|\big(\partial_x^2+\partial_z^2\big)u^3_{\neq}(t)\big\|_{L^2}\\ &\quad + \nu^{1/4}\big\|u^2_{\neq}(t)\big\|_{H^2} + \nu^{1/3}\big\|\big(u^1_{\neq},u^3_{\neq}\big)(t)\big\|_{H^1} + \big\|u^2_{\neq}(t)\big\|_{L^\infty}\\ &\quad + \nu^{1/6}\big\|\big(u^1_{\neq},u^3_{\neq}\big)(t)\big\|_{L^\infty} \le Ce^{-2\epsilon\nu^{1/3}t}\|u_0\|_{H^2},\\ &\|u_{\neq}\|_{L^\infty L^2} + \sqrt{\nu}\big\|t\big(u^1_{\neq},u^3_{\neq}\big)\big\|_{L^2L^2} + \big\|\nabla u^2_{\neq}\big\|_{L^\infty L^2}\\ &\quad + \big\|\nabla u^2_{\neq}\big\|_{L^2L^2} \le C\|u_0\|_{H^2}.\end{aligned}$$

Let us give some remarks on our result.

1. Global stability estimates in particular imply that

$$\big\|u(t)\big\|_{L^\infty} \le Cc_0 e^{-\nu t} \to 0 \quad \text{as} \quad t \to +\infty.$$

This means that the 3D Couette flow is nonlinearly stable in the L^∞-sense when the perturbation is $o(\nu)$ in H^2.

2. Our rigorous analysis shows that various linear effects (including 3D lift-up effect, boundary layer effect, inviscid damping, and enhanced dissipation) play a crucial role in determining the transition threshold. Surprisingly, the transition threshold obtained in this paper is consistent with that for the case of $\Omega = \mathbb{T} \times \mathbb{R} \times \mathbb{T}$ obtained in [52]. This shows that the 3D lift-up may be the main mechanism leading to the instability of the flow even in the presence of the boundary layer effect. Our explanation of this surprise result is that weak nonlinear interaction (or null structure of nonlinear terms) and good linear mechanisms (inviscid damping and enhanced dissipation) counteract the bad effect of the boundary layer.

3. The transition threshold problem is very interesting in an infinite channel $\Omega = \mathbb{R} \times [-1,1] \times \mathbb{T}$. In this case, we need to understand the long wave effect in the x variable. In fact, we conjecture that the threshold may be strictly less than 1 in this case.

4. The asymptotic analysis conducted in [13] indicates that the profile of shear flows may affect the transition threshold. From the results in [13], it seems reasonable to conjecture that the threshold $\beta \le \frac{3}{2}$ for the plane Poiseuille flow. In [34], Li, Wei, and Zhang proved that the threshold $\beta \le \frac{7}{4}$ for the 3D Kolmogorov flow. It is unclear whether one can improve it to $\beta \le \frac{3}{2}$.

5. The transition threshold problem for the pipe Poiseuille flow is completely open. This flow is probably the most interesting and important because it is close to the setting of the experiment conducted by Reynolds in 1883. The experimental result carried out by Hof, Juel, and Mullin [25] conclude that the minimum amplitude of a perturbation required to cause transition scales as the inverse of the Reynolds number, i.e., $O(\mathrm{Re}^{-1})$. The subsequent numerical result in [40] agrees with the experiment result in [25] for $\mathrm{Re} \gtrsim 4000$.

Now we give a sketch of some key ingredients of the proof.

First of all, we decompose the solution u into the zero mode $\bar{u}$ and nonzero mode $u_{\neq}$ due to their different behaviors. The zero mode $\bar{u}$ satisfies

$$(\partial_t - \nu\Delta)\bar{u}^1 + \bar{u}^2 + \overline{u \cdot \nabla u^1} = 0, \tag{4.3}$$

$$(\partial_t - \nu\Delta)\bar{u}^j + \partial_j \bar{p} + \big(\bar{u}^2\partial_y + \bar{u}^3\partial_z\big)\bar{u}^j + \overline{u_{\neq} \cdot \nabla u^j_{\neq}} = 0, \quad j = 2,3. \tag{4.4}$$

To estimate nonzero modes, we will use a formulation in terms of the shearwise velocity u^2 and vorticity $\omega^2 = \partial_z u^1 - \partial_x u^3$:

$$\begin{cases} \partial_t(\Delta u^2) - \nu\Delta^2 u^2 + y\partial_x \Delta u^2 + \big(\partial_x^2 + \partial_z^2\big)(u \cdot \nabla u^2) \\ \quad - \partial_y\big[\partial_x(u \cdot \nabla u^1) + \partial_z(u \cdot \nabla u^3)\big] = 0, \\ \partial_t \omega^2 - \nu\Delta\omega^2 + y\partial_x\omega^2 + \partial_z u^2 + \partial_z(u \cdot \nabla u^1) - \partial_x(u \cdot \nabla u^3) = 0, \\ \partial_y u^2(t,x,\pm 1,z) = u^2(t,x,\pm 1,z) = 0, \quad \omega^2(x,\pm 1,z) = 0. \end{cases} \tag{4.5}$$

The idea of using Δu^2 may go back to Kelvin's original paper [30]. The main advantage of using Δu^2 is that the equation of Δu^2 does not destroy the linear structure. This important point has played an important role in the works [6,52].

The linearized system of zero mode $\bar{u}$ becomes

$$(\partial_t - \nu\Delta)\bar{u}^1 + \bar{u}^2 = 0, \quad (\partial_t - \nu\Delta)\bar{u}^j + \partial_j \bar{p} = 0, \quad j = 2,3.$$

Then it is easy to see that $\|\bar{u}^1(t)\|_{L^2} \leq C(1+t)e^{-\nu t}\|u_0\|_{L^2}$. When $t \lesssim \nu^{-1}$, $\bar{u}^1$ grows linearly in time. This phenomenon is referred to as the *3D lift-up*. To keep $\bar{u}^1$ small, the perturbation u_0 should be as small as $o(\nu)$. From this point of view, our result seems optimal. It also turns out that the 3D lift-up is the worst mechanism leading to the instability.

To estimate Δu^2 and ω^2, we need to establish the space–time estimates for the following linearized system:

$$\begin{cases} \partial_t\omega - \nu\big(\partial_y^2 - \eta^2\big)\omega + iky\omega = -ikf_1 - \partial_y f_2 - i\ell f_3 - f_4, \\ \omega|_{y=\pm 1} = 0, \quad \omega|_{t=0} = \omega_{in}, \end{cases} \tag{4.6}$$

and

$$\begin{cases} \partial_t\omega - \nu\big(\partial_y^2 - \eta^2\big)\omega + iky\omega = F, \\ \big(\partial_y^2 - \eta^2\big)\varphi = \omega, \ \partial_y\varphi|_{y=\pm 1} = \varphi|_{y=\pm 1} = 0, \\ \omega|_{t=0} = \omega_{in}. \end{cases} \tag{4.7}$$

Here $\eta^2 = k^2 + \ell^2$. In [16], we establish the following space–time estimates.

Theorem 4.2. *Let ω be a solution of (4.6) with $f_4(t,\pm 1) = 0$ and $\omega_{in}(\pm 1) = 0$. Then there exists $\epsilon_1 > 0$ so that, for any $a \in [0,\epsilon_1]$,*

$$\begin{aligned} &\big\|e^{a\nu^{1/3}t}\omega\big\|_{L^\infty L^2}^2 + \nu\big\|e^{a\nu^{1/3}t}\omega'\big\|_{L^2L^2}^2 + \big(\nu\eta^2 + (\nu k^2)^{1/3}\big)\big\|e^{a\nu^{1/3}t}\omega\big\|_{L^2L^2}^2 \\ &\quad \leq C\big(\|\omega_{in}\|_{L^2}^2 + \nu^{-1}\big\|e^{a\nu^{1/3}t}f_2\big\|_{L^2L^2}^2 + \big(\eta|k|\big)^{-1}\big\|e^{a\nu^{1/3}t}\partial_y f_4\big\|_{L^2L^2}^2 \\ &\qquad + \eta|k|^{-1}\big\|e^{a\nu^{1/3}t}f_4\big\|_{L^2L^2}^2 + \min\big((\nu\eta^2)^{-1}, (\nu k^2)^{-1/3}\big)\big\|e^{a\nu^{1/3}t}(kf_1 + \ell f_3)\big\|_{L^2L^2}^2\big). \end{aligned}$$

Moreover, we have

$$\begin{aligned} &\big\|e^{a\nu^{1/3}t}\omega'\big\|_{L^\infty L^2}^2 + \nu\big\|e^{a\nu^{1/3}t}\omega''\big\|_{L^2L^2}^2 + \nu\eta^2\big\|e^{a\nu^{1/3}t}\omega'\big\|_{L^2L^2}^2 \\ &\quad \leq C\big\|\omega_{in}'\big\|_{L^2}^2 + C\nu^{-\frac{2}{3}}|k|^{\frac{2}{3}}\big(\|\omega_{in}\|_{L^2}^2 + \big(\eta|k|\big)^{-1}\big\|e^{a\nu^{1/3}t}\partial_y f_4\big\|_{L^2L^2}^2 \\ &\qquad + \eta|k|^{-1}\big\|e^{a\nu^{1/3}t}f_4\big\|_{L^2L^2}^2\big) + C\nu^{-1}\big(\big\|e^{a\nu^{1/3}t}(kf_1 + \ell f_3)\big\|_{L^2L^2}^2 \\ &\qquad + \nu^{-\frac{2}{3}}|k|^{\frac{2}{3}}\big\|e^{a\nu^{1/3}t}f_2\big\|_{L^2L^2}^2 + \big\|e^{a\nu^{1/3}t}\partial_y f_2\big\|_{L^2L^2}^2\big). \end{aligned}$$

Here $\omega' = \partial_y\omega$ and $\omega'' = \partial_y^2\omega$.

Theorem 4.3. *Let ω solve (4.7) with $\partial_y\varphi_{in}|_{y=\pm 1} = 0$ and $F = ikf_1 + \partial_y f_2 + i\ell f_3$. Then there exist $\epsilon_1 > 0$, $\nu_0 > 0$ so that, for any $a \in [0,\epsilon_1]$, $\nu \in (0,\nu_0)$,*

$$\begin{aligned} &|k\eta|^{\frac{1}{2}}\big\|e^{a\nu^{\frac{1}{3}}t}(\partial_y,\eta)\varphi\big\|_{L^2L^2} + \nu^{\frac{3}{4}}\big\|e^{a\nu^{\frac{1}{3}}t}\partial_y\omega\big\|_{L^2L^2} + \nu^{\frac{1}{2}}\eta\big\|e^{a\nu^{\frac{1}{3}}t}\omega\big\|_{L^2L^2} \\ &\quad + \eta\big\|e^{a\nu^{\frac{1}{3}}t}(\partial_y,\eta)\varphi\big\|_{L^\infty L^2} + \nu^{\frac{1}{4}}\big\|e^{a\nu^{\frac{1}{3}}t}\omega\big\|_{L^\infty L^2} \\ &\leq C\nu^{-\frac{1}{2}}\big\|e^{a\nu^{\frac{1}{3}}t}(f_1,f_2,f_3)\big\|_{L^2L^2} + C\big(\eta^{-1}\|\partial_y\omega_{in}\|_{L^2} + \|\omega_{in}\|_{L^2}\big). \end{aligned}$$

The proofs of Theorems 4.2 and 4.3 used the resolvent estimate method developed in [14]. The main idea is to separate the resolvent problem into two subproblems:

1. *The inhomogeneous problem with favorable boundary conditions*
   The good boundary conditions avoid the boundary terms caused by the integration by parts argument so that we can establish various resolvent estimates via the direct energy method by choosing suitable multipliers.

2. *The homogenous problem with nonvanishing boundary conditions*
   This step is to match the boundary conditions. We can first use the Airy function or the solution of a simple elliptic problem to construct an approximate solution. Then we can construct the solution to the homogenous problem via solving a perturbation problem with favorable boundary conditions.

The space–time estimates established in Theorems 4.2 and 4.3 encompass four kinds of important linear effects: heat diffusion, enhanced dissipation, inviscid damping, and boundary layer. These estimates should be enough to prove a transition threshold $\beta \le \frac{5}{3}$. To achieve the sharp threshold, we have to handle the problem in a quasilinear way. That is, we need to consider the full linearized 3D Navier–Stokes system around the flow $(V(y,z),0,0)$, which is a small perturbation of the Couette flow, i.e.,

$$\|V-y\|_{H^4} \le \varepsilon_0, \quad V(y,z)-y|_{y=\pm 1}=0,$$

with ε_0 small enough but independent of ν. We denote

$$\mathbb{A}_{\nu,V}u = \mathbb{P}\big(\nu\Delta u - V\partial_x u - \big(\partial_y V(u^2+\kappa u^3),0,0\big)\big),$$

here $\mathbb{P}$ is the Leray projection and $\kappa = \partial_z V/\partial_y V$. Then we study the following linearized system:

$$\partial_t u_{\neq} - \mathbb{A}_{\nu,V}u_{\neq} + \vec{g} = 0. \tag{4.8}$$

The key point is to exclude the unstable eigenvalues of the operator $\mathbb{A}_{\nu,V}$. This problem is highly nontrivial. Even for the following linearized equation:

$$\begin{cases} \partial_t w - \nu\Delta w + V\partial_x w = f, \\ \Delta\varphi = w, \quad \varphi|_{y=\pm 1} = \partial_y\varphi|_{y=\pm 1} = 0, \end{cases}$$

the linear stability when $V = V(y)$ is close to y was just proved by Almog and Helffer [1]. After applying the Fourier transform with respect to (t,x) and introducing $W = u^2+\kappa u^3$ and $U = u^3$, the problem is reduced to the following linearized system in terms of (W,U):

$$\begin{cases} -\nu\Delta W + ik\big(V(y,z)-\lambda\big)W - a(\nu k^2)^{1/3}W + (\partial_y+\kappa\partial_z)p^{L1} \\ \quad + G_1 + \nu(\Delta\kappa)U + 2\nu\nabla\kappa\cdot\nabla U = 0, \\ -\nu\Delta U + ik\big(V(y,z)-\lambda\big)U - a(\nu k^2)^{1/3}U + G_2 + \partial_z p^{L1} = 0, \\ W|_{y=\pm 1} = \partial_y W|_{y=\pm 1} = U|_{y=\pm 1} = 0, \end{cases} \tag{4.9}$$

where $\lambda \in \mathbb{R}$ and

$$\Delta p^{L1} = -2ik\partial_y VW, \quad \partial_x W = ikW, \ \partial_x U = ikU, \quad \partial_x p^{L1} = ikp^{L1}.$$

Theorem 4.4. *Let $W \in H^4(\Omega)$, $U \in H^2(\Omega)$ be a solution of* (4.9)*. Then there exist $\epsilon_1 > 0$, $\nu_0 > 0$ so that, for any $a \in [0, \epsilon_1]$, $\nu \in (0, \nu_0)$,*

$$\begin{aligned}&\nu^{\frac{1}{3}}\big(\big\|\partial_x^2 U\big\|_{L^2}^2 + \big\|\partial_x(\partial_z - \kappa\partial_y)U\big\|_{L^2}^2\big) + \nu\big(\big\|\nabla\partial_x^2 U\big\|_{L^2}^2 + \big\|\nabla\partial_x(\partial_z - \kappa\partial_y)U\big\|_{L^2}^2\big)\\ &\quad + \nu^{\frac{1}{3}}\|\partial_x\nabla W\|_{L^2}^2 + \nu\|\partial_x\Delta W\|_{L^2}^2 + \nu^{\frac{5}{3}}\|\partial_x\Delta U\|_{L^2}^2\\ &\quad \le C\nu^{-1}\big(\|\nabla G_1\|_{L^2}^2 + \|\partial_x G_2\|_{L^2}^2\big).\end{aligned}$$

In particular, this result shows that *the 3D linearized Navier–Stokes system* (4.8) *around the Couette flow is linearly stable.* This theorem is the key and most difficult part in the proof of nonlinear stability. The proof was motivated by our work [52]. The key point is to introduce a good unknown $W_g = W - W_s$, where W_s is the singular part of W. Then $w_g = \Delta W_g$ satisfies

$$-\nu\Delta w_g + ik\big(V(y,z) - \lambda\big)w_g - a(\nu k^2)^{1/3}w_g = \text{good terms}.$$

For nonlinear stability, we introduce the following energy functionals, which are suitable adaptations of those introduced in [52].

(1) *Energy functional of zero mode*. We first decompose $\bar{u}^1 = \bar{u}^{1,0} + \bar{u}^{1,\neq}$ with

$$\begin{aligned}&(\partial_t - \nu\Delta)\bar{u}^{1,0} + \bar{u}^2 + \bar{u}^2\partial_y\bar{u}^{1,0} + \bar{u}^3\partial_z\bar{u}^{1,0} = 0,\\ &(\partial_t - \nu\Delta)\bar{u}^{1,\neq} + \bar{u}^2\partial_y\bar{u}^{1,\neq} + \bar{u}^3\partial_z\bar{u}^{1,\neq} + \overline{u_{\neq}\cdot\nabla u_{\neq}^1} = 0,\\ &\bar{u}^{1,0}|_{t=0} = 0, \quad \bar{u}^{1,\neq}|_{t=0} = \bar{u}^1(0), \quad \bar{u}^{1,0}|_{y=\pm 1} = 0, \quad \bar{u}^{1,\neq}|_{y=\pm 1} = 0.\end{aligned}$$

The main reason for making this decomposition is that $\bar{u}^{1,\neq}$ has better decay in ν, and thus $\bar{u}^{1,\neq}\partial_x$ could be viewed as a perturbation. In this way, we avoid estimating the higher-order derivatives of nonzero modes. Then we introduce the following energy functional to control the zero mode:

$$E_1 = E_{1,0} + \nu^{-2/3}E_{1,\neq},$$

where

$$\begin{aligned}E_{1,0} &= \big\|\bar{u}^{1,0}\big\|_{L^\infty H^4} + \nu^{-1}\big\|\partial_t\bar{u}^{1,0}\big\|_{L^\infty H^2} + \nu^{-\frac{1}{2}}\big\|\partial_t\bar{u}^{1,0}\big\|_{L^2 H^3},\\ E_{1,\neq} &= \big\|\bar{u}^{1,\neq}\big\|_{L^\infty H^2} + \nu^{\frac{1}{2}}\big\|\nabla\bar{u}^{1,\neq}\big\|_{L^2 H^2},\end{aligned}$$

and the energy E_2 is defined by

$$\begin{aligned}E_2 = {}& \big\|\Delta\bar{u}^2\big\|_{L^\infty L^2} + \nu^{\frac{1}{2}}\big\|\nabla\Delta\bar{u}^2\big\|_{L^2L^2} + \nu^{\frac{1}{2}}\big\|\Delta\bar{u}^2\big\|_{L^2L^2} + \nu^{-\frac{1}{2}}\big\|\partial_t\nabla\bar{u}^2\big\|_{L^2L^2}\\ &+ \big\|\nabla\bar{u}^3\big\|_{L^\infty L^2} + \nu^{\frac{1}{2}}\big\|\Delta\bar{u}^3\big\|_{L^2L^2} + \nu^{\frac{1}{2}}\big\|\nabla\bar{u}^3\big\|_{L^2L^2} + \nu^{-\frac{1}{2}}\big\|\partial_t\bar{u}^3\big\|_{L^2L^2}\\ &+ \big\|\min\big((\nu^{\frac{2}{3}} + \nu t)^{\frac{1}{2}}, 1 - y^2\big)\Delta\bar{u}^3\big\|_{L^\infty L^2}\\ &+ \nu^{-\frac{1}{2}}\big\|\min\big((\nu^{\frac{2}{3}} + \nu t)^{\frac{1}{2}}, 1 - y^2\big)\nabla\partial_t\bar{u}^3\big\|_{L^\infty L^2}\\ &+ \nu^{\frac{1}{2}}\big\|\min\big((\nu^{\frac{2}{3}} + \nu t)^{\frac{1}{2}}, 1 - y^2\big)\nabla\Delta\bar{u}^3\big\|_{L^2L^2}.\end{aligned}$$

The estimates of E_1 and E_2 are based on direct energy estimates for the system (4.3) and (4.4).

(2) *Energy functional of nonzero mode (semilinear part).* We consider

$$E_3 = E_{3,0} + E_{3,1},$$

where $E_{3,0}$ and $E_{3,1}$ are defined by

$$\begin{aligned} E_{3,0} &= \nu^{\frac{1}{2}} \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} (\partial_x, \partial_z) \Delta u^2_{\neq} \big\|_{L^2L^2} + \nu^{\frac{3}{4}} \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \nabla \Delta u^2_{\neq} \big\|_{L^2L^2} \\ &\quad + \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} (\partial_x, \partial_z) \nabla u^2_{\neq} \big\|_{L^\infty L^2} + \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \partial_x \nabla u^2_{\neq} \big\|_{L^2L^2} \\ &\quad + \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \big(\partial_x^2 + \partial_z^2\big) u^3_{\neq} \big\|_{L^\infty L^2} + \nu^{\frac{1}{2}} \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \big(\partial_x^2 + \partial_z^2\big) \nabla u^3_{\neq} \big\|_{L^2L^2}, \\ E_{3,1} &= \nu^{\frac{1}{3}} \Big(\big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \nabla \omega^2_{\neq} \big\|_{L^\infty L^2} + \nu^{\frac{1}{2}} \big\| e^{2\epsilon\nu^{\frac{1}{3}}t} \Delta \omega^2_{\neq} \big\|_{L^2L^2} \Big). \end{aligned}$$

The estimate of E_3 is based on the space–time estimates for the coupled system (4.5) of $(\Delta u^2, \omega^2)$ via Theorems 4.2 and 4.3.

(3) *Energy functional of nonzero mode (quasilinear part).* We now treat

$$E_5 = \nu^{1/6} \big\| e^{3\epsilon\nu^{1/3}t} \partial_x^2 u^2_{\neq} \big\|_{L^2L^2} + \nu^{1/6} \big\| e^{3\epsilon\nu^{1/3}t} \partial_x^2 u^3_{\neq} \big\|_{L^2L^2},$$

which is vital to control some nonlinear interaction terms with the lift-up effect such as $\bar{u}^1 \partial_x u_{\neq}$ and $u^j_{\neq} \partial_j \bar{u}^1 (j = 2, 3)$. The estimate of E_5 relies on Theorem 4.4.

Based on the linear space–time estimates, combined with a nonlinear interaction estimate, we can derive the following uniform energy estimates:

$$\begin{aligned} &E_{1,0} \le C\nu^{-1} \big(\|u_0\|_{H^2} + E_2 + E_2 E_{1,0} \big), \\ &E_{1,\neq} \le C \big(\|u_0\|_{H^2} + \nu^{-1} E_2 E_{1,\neq} + \nu^{-\frac{4}{3}} E_3^2 \big), \\ &E_2 \le C(1 + \nu^{-1} E_2)^2 \big(\|u(0)\|_{H^2} + \nu^{-1} E_3^2 \big), \end{aligned}$$

and

$$\begin{aligned} &E_{3,0}^2 \le C \|u_0\|_{H^2}^2 + C \big(E_3^4/\nu^2 + E_2^2 E_3^2/\nu^2 + E_1^2 E_3 E_5 + E_1^2 E_3^{\frac{3}{2}} E_5^{\frac{1}{2}} \big), \\ &E_{3,1}^2 \le C \big(\|u_0\|_{H^2}^2 + \nu^{-2} E_3^4 + \nu^{-\frac{4}{3}} E_2^2 E_3^2 + E_1^2 E_3 E_5 + E_1^2 E_3^{\frac{7}{4}} E_5^{\frac{1}{4}} + E_1^2 E_3^{\frac{3}{2}} E_5^{\frac{1}{2}} \big), \end{aligned}$$

as well as

$$E_5^2 \le C E_6^2 \le C \|u_0\|_{H^2}^2 + C \big(E_1^2 + \nu^{-2} E_2^2 \big) E_6^2 + C\nu^{-2} E_3^4,$$

where E_6 is an auxiliary energy functional (see Section 14 in [16] for the definition of E_6). When the perturbation $\|u_0\|_{H^2} \le c_0 \nu$, E_1 is small due to the lift-up effect, while E_2, E_3, and E_5 are as small as $o(\nu)$.

ACKNOWLEDGMENTS

We thank Qi Chen, Te Li, and Weiren Zhao for many helpful comments when writing this note.

FUNDING

This work was partially supported by NSF of China under Grant 11425103.

REFERENCES

[1] Y. Almog and B. Helffer, On the stability of laminar flows between plates. *Arch. Ration. Mech. Anal.* **241** (2021), 1281–1401.

[2] M. Beck and C. E. Wayne, Metastability and rapid convergence to quasi-stationary bar states for the two-dimensional Navier–Stokes equations. *Proc. Roy. Soc. Edinburgh Sect. A* **143** (2013), 905–927.

[3] J. Bedrossian and M. Coti Zelati, Enhanced dissipation, hypoellipticity, and anomalous small noise inviscid limits in shear flows. *Arch. Ration. Mech. Anal.* **224** (2017), 1161–1204.

[4] J. Bedrossian, M. Coti Zelati, and V. Vicol, Vortex axisymmetrization, inviscid damping, and vorticity depletion in the linearized 2D Euler equations. *Ann. PDE* **5** (2019), Art. 4, 192 pp.

[5] J. Bedrossian, P. Germain, and N. Masmoudi, Dynamics near the subcritical transition of the 3D Couette flow II: Above threshold case. 2015, arXiv:1506.03721.

[6] J. Bedrossian, P. Germain, and N. Masmoudi, On the stability threshold for the 3D Couette flow in Sobolev regularity. *Ann. of Math.* **185** (2017), 541–608.

[7] J. Bedrossian, P. Germain, and N. Masmoudi, Stability of the Couette flow at high Reynolds number in 2D and 3D. *Bull. Amer. Math. Soc. (N.S.)* **56** (2019), 373–414.

[8] J. Bedrossian, P. Germain, and N. Masmoudi, Dynamics near the subcritical transition of the 3D Couette flow I: Below threshold case. *Mem. Amer. Math. Soc.* **266** (2020), no. 1294.

[9] J. Bedrossian and S. He, Suppression of blow-up in Patlak–Keller–Segel via shear flows. *SIAM J. Math. Anal.* **49** (2017), 4722–4766.

[10] J. Bedrossian and N. Masmoudi, Inviscid damping and the asymptotic stability of planar shear flows in the 2D Euler equations. *Publ. Math. Inst. Hautes Études Sci.* **122** (2015), 195–300.

[11] F. Bouchet and H. Morita, Large time behavior and asymptotic stability of the 2D Euler and linearized Euler equations. *Phys. D* **239** (2010), 948–966.

[12] K. Case, Stability of inviscid plane Couette flow. *Phys. Fluids* **3** (1960), 143–148.

[13] S. J. Chapman, Subcritical transition in channel flows. *J. Fluid Mech.* **451** (2002), 35–97.

[14] Q. Chen, T. Li, D. Wei, and Z. Zhang, Transition threshold for the 2D Couette flow in a finite channel. *Arch. Ration. Mech. Anal.* **238** (2020), 125–183.

[15] Q. Chen, D. Wei, and Z. Zhang, Linear stability of pipe Poiseuille flow at high Reynolds number. 2019, arXiv:1910.14245. *Comm. Pure Appl. Math.*, in press.

[16] Q. Chen, D. Wei, and Z. Zhang, Transition threshold for the 3D Couette flow in a finite channel. 2020, arXiv:2006.00721. *Mem. Amer. Math. Soc.*, in press.

[17] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatos, Diffusion and mixing in fluid flow. *Ann. of Math.* **168** (2008), 643–674.

[18] M. Coti Zelati, M. G. Delgadino, and T. M. Elgindi, On the relation between enhanced dissipation timescales and mixing rates. *Comm. Pure Appl. Math.* **73** (2020), 1205–1244.

[19] Y. Deng and N. Masmoudi, Long time instability of the Couette flow in low Gevrey spaces. 2018, arXiv:1803.01246.

[20] P. Drazin and W. Reid, *Hydrodynamic stability*. Cambridge Monogr. Mech. Appl. Math., Cambridge Univ. Press, New York, 1981.

[21] T. Gallay, Enhanced dissipation and axisymmetrization of two-dimensional viscous vortices. *Arch. Ration. Mech. Anal.* **230** (2018), 939–975.

[22] T. Gebhardt and S. Grossmann, Chaos transition despite linear stability. *Phys. Rev. E* **50** (1994), 3705–3711.

[23] E. Grenier, T. Nguyen, F. Rousset, and A. Soffer, Linear inviscid damping and enhanced viscous dissipation of shear flows by using the conjugate operator method. *J. Funct. Anal.* **278** (2020), 108339, 27 pp.

[24] S. He, Suppression of blow-up in parabolic-parabolic Patlak–Keller–Segel via strictly monotone shear flows. *Nonlinearity* **31** (2018), 3651–3688.

[25] B. Hof, A. Juel, and T. Mullin, Scaling of the turbulence transition threshold in a pipe. *Phys. Rev. Lett.* **91** (2004), 244502.

[26] S. Ibrahim, Y. Maekawa, and N. Masmoudi, On pseudospectral bound for non-selfadjoint operators and its application to stability of Kolmogorov flows. *Ann. PDE* **5** (2019), Paper No. 14, 84 pp.

[27] A. Ionescu and H. Jia, Nonlinear inviscid damping near monotonic shear flows. 2018, arXiv:2001.03087.

[28] A. Ionescu and H. Jia, Axi-symmetrization near point vortex solutions for the 2D Euler equation. 2019, arXiv:1904.09170.

[29] A. Ionescu and H. Jia, Inviscid damping near the Couette flow in a channel. *Comm. Math. Phys.* **374** (2020), 2015–2096.

[30] L. Kelvin, Stability of fluid motion-rectilinear motion of viscous fluid between two parallel plates. *Philos. Mag.* **24** (1887), 188–196.

[31] A. Kiselev and X. Xu, Suppression of chemotactic explosion by mixing. *Arch. Ration. Mech. Anal.* **222** (2016), 1077–1112.

[32] L. Landau, On the vibration of the electronic plasma. *J. Phys. USSR* **10** (1946), 25.

[33] T. Li, D. Wei, and Z. Zhang, Pseudospectral and spectral bounds for the Oseen vortices operator. *Ann. Sci. Éc. Norm. Supér.* **53** (2020), 993–1035.

[34] T. Li, D. Wei, and Z. Zhang, Pseudospectral bound and transition threshold for the 3D Kolmogorov flow. *Comm. Pure Appl. Math.* **73** (2020), 465–557.

[35] Y. C. Li and Z. Lin, A resolution of the Sommerfeld paradox. *SIAM J. Math. Anal.* **43** (2011), 1923–1954.

[36] Z. Lin and M. Xu, Metastability of Kolmogorov flows and inviscid damping of shear flows. *Arch. Ration. Mech. Anal.* **231** (2019), 1811–1852.

[37] Z. Lin and C. Zeng, Inviscid dynamical structures near Couette flow. *Arch. Ration. Mech. Anal.* **200** (2011), 1075–1097.

[38] A. Lundbladh, D. Henningson, and S. Reddy, Threshold amplitudes for transition in channel flows. In *Transition, turbulence and combustion*, pp. 309–318, Springer, New York, 1994.

[39] N. Masmoudi and Z. Zhao, Nonlinear inviscid damping for a class of monotone shear flows in finite channel. 2020, arXiv:2001.08564.

[40] F. Mellibovskya and A. Meseguer, Pipe flow transition threshold following localized impulsive perturbations. *Phys. Fluids* **19** (2007), 044102.

[41] W. Orr, The stability or instability of steady motions of a perfect liquid and of a viscous liquid. Part I: A perfect liquid. *Proc. R. Irish Acad. Sec. A: Math. Phys. Sci.* **27** (1907), 9–68.

[42] S. Orszag and L. Kells, Transition to turbulence in plane Poiseuille and plane Couette flow. *J. Fluid Mech.* **96** (1980), 159–205.

[43] O. Reynolds, An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Proc. R. Soc. Lond.* **35** (1883), 84.

[44] V. A. Romanov, Stability of plane-parallel Couette flow. *Funktsional. Anal. i Prilozhen.* **7** (1973), 62–73.

[45] D. Ryutov, Landau damping: half a century with the great discovery. *Plasma Phys. Control Fusion* **41** (1999), A1–A12.

[46] P. Schmid and D. Henningson, *Stability and transition in shear flows*. Appl. Math. Sci. 142, Springer, New York, 2001.

[47] L. N. Trefethen, Pseudospectra of linear operators. *SIAM Rev.* **39** (1997), 383–406.

[48] L. Trefethen, A. Trefethen, S. Reddy, and T. Driscoll, Hydrodynamic stability without eigenvalues. *Science* **261** (1993), 578–584.

[49] C. Villani, Hypocoercivity. *Mem. Amer. Math. Soc.* **202** (2009), no. 950, iv+141.

[50] D. Wei, Diffusion and mixing in fluid flow via the resolvent estimate. *Sci. China Math.* **64** (2021), 507–518.

[51] D. Wei and Z. Zhang, Enhanced dissipation for the Kolmogorov flow via the hypocoercivity method. *Sci. China Math.* **62** (2019), 1219–1232.

[52] D. Wei and Z. Zhang, Transition threshold for the 3D Couette flow in Sobolev space. *Comm. Pure Appl. Math.* DOI 10.1002/cpa.21948.

[53] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping for a class of monotone shear flow in Sobolev spaces. *Comm. Pure Appl. Math.* **71** (2018), 617–687.

[54] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping and vorticity depletion for shear flows. *Ann. PDE* **5** (2019), Paper No. 3, 101 pp.

[55] D. Wei, Z. Zhang, and W. Zhao, Linear inviscid damping and enhanced dissipation for the Kolmogorov flow. *Adv. Math.* **362** (2020), 106963.

[56] D. Wei, Z. Zhang, and H. Zhu, Linear inviscid damping for the β-plane equation. *Comm. Math. Phys.* **375** (2020), 127–174.

[57] A. Yaglom, *Hydrodynamic instability and transition to turbulence*. Fluid Mech. Appl. 100, Springer, New York, 2012.

[58] C. Zillinger, Linear inviscid damping for monotone shear flows in a finite periodic channel, boundary effects, blow-up and critical Sobolev regularity. *Arch. Ration. Mech. Anal.* **221** (2016), 1449–1509.

[59] C. Zillinger, Linear inviscid damping for monotone shear flows. *Trans. Amer. Math. Soc.* **369** (2017), 8799–8855.

DONGYI WEI

School of Mathematical Science, Peking University, 100871, Beijing, P. R. China, jnwdyi@pku.edu.cn

ZHIFEI ZHANG

School of Mathematical Science, Peking University, 100871, Beijing, P. R. China, zfzhang@math.pku.edu.cn

11. MATHEMATICAL PHYSICS

SPECIAL LECTURE

RECENT PROGRESS IN GENERAL RELATIVITY

PETER HINTZ AND GUSTAV HOLZEGEL

ABSTRACT

We review recent progress in general relativity. After a brief introduction to some of the key analytical and geometric features of the Einstein equations, we focus on two main developments: the stability of black hole solutions, and the formation, structure, and dynamical stability of singularities.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 83C05; Secondary 83C57, 35L05, 35B40, 35L81

KEYWORDS

Einstein's field equations, black hole stability, Strong Cosmic Censorship

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3924–3984
DOI 10.4171/ICM2022/128

1. INTRODUCTION

The objective of this article is to describe recent progress in the mathematical analysis of the Einstein equations of general relativity. General relativity is the theory of gravitation postulated by Einstein in 1915. It superseded the Newtonian theory and rose to one of the best experimentally tested physical theories we have. The Einstein equations can accurately describe violent astrophysical processes happening in our universe (such as the merger of black holes), they provide a model for the evolution and dynamics of our entire universe in cosmology, and they are also central for the functioning of the ubiquitously used GPS system on Earth. Recent experimental breakthroughs such as the detection of gravitational waves, a prediction of Einstein's theory, have inspired new developments in mathematics and physics some of which we shall discuss below.

This article consists of three main parts, which are aimed at readers with potentially different levels of expertise. The first part (Section 1) contains an introduction to some of the basic geometric and analytic principles governing the study of general relativity (with emphasis on the issue of diffeomorphism invariance of the governing equations). It also includes examples of black hole solutions and a discussion of their geometry. This part is intended for mathematicians who are perhaps familiar with the theory of partial differential equations but have otherwise little prior knowledge of general relativity.

The second and main part (Section 2) contains a review of recent mathematical results that have been obtained in the study of the stability of black holes. This part is aimed mostly at readers who have had some previous experience with the study of wave equations on curved backgrounds; familiarity with the Einstein equations will be useful for the results on nonlinear or linearized gravity. Research during the past 15–20 years pioneered the techniques that led in particular to the powerful nonlinear results which have been obtained in the past five years and that we shall describe in more detail. We have kept the discussion rather informal: theorems are often stated rather loosely in order to avoid introducing too much notation; additional details are provided in the text. We hope that our descriptions of the main ideas of the proofs can serve as an invitation to delve into the papers in this area in some more detail.

The third part (Section 3) is concerned with the issue of singularities. We shall discuss the structure of singularities appearing inside black holes as well as their stability, which is related to Penrose's famous *Strong Cosmic Censorship* conjecture. We also discuss results concerning *naked singularities*, which are related to *Weak Cosmic Censorship*.

Unfortunately, in this overview we cannot do justice to the wide range of developments in mathematical relativity. The choice of topics is influenced both by our personal expertise and taste. In particular, we almost exclusively focus on the vacuum equations. There are many exciting developments that we will not be able to discuss: these include recent progress on the weak limit of the Einstein equations (e.g., Burnett's conjecture [32,123], also [155]) and the structure and dynamics of cosmological singularities (stable big bang formation [82,83,181]), among others.

1.1. The Einstein equations

From a mathematical point of view, the Einstein equations constitute a set of nonlinear partial differential equations formulated in the language of differential geometry with the dynamical variable being a Lorentzian manifold $(\mathcal{M}, g)$:

$$\mathrm{Ric}(g) - \frac{1}{2} R g + \Lambda g = 8\pi \mathbb{T}. \tag{1.1}$$

Here $\mathrm{Ric}(g)$ and $R = \mathrm{tr}_g \mathrm{Ric}$ denote the Ricci tensor and scalar curvature of $(\mathcal{M}, g)$, Λ is the cosmological constant, and $\mathbb{T}$ is the stress–energy–momentum tensor of the matter present in the spacetime. Below we shall mostly restrict attention to the vacuum case $\mathbb{T} = 0$ for which the equations (1.1), unlike their Newtonian analogue, already exhibit extremely rich dynamics, although comments will be made about the analysis of spacetimes with matter. Furthermore, we will focus on the physical case $\dim \mathcal{M} = 4$.

Applying the trace reversal operator $\mathsf{G}_g : T \mapsto T - \frac{1}{2} g \,\mathrm{tr}_g T$ to equation (1.1) yields the following equivalent form of the Einstein equations:

$$\mathrm{Ric}(g) - \Lambda g = 8\pi \left(\mathbb{T} - \frac{1}{2} g \,\mathrm{tr}_g \mathbb{T} \right) \quad (= 0 \text{ in vacuum}). \tag{1.2}$$

1.2. The Cauchy problem and generalised harmonic gauges

While not immediate from the coordinate independent formulation (1.1), from a PDE point of view (1.1) should be viewed as a hyperbolic system of equations, i.e., as a system admitting an appropriate initial value problem. The geometric notion of initial data is as follows:

Definition 1. A triple $(\Sigma, \overline{g}, K)$ consisting of a smooth Riemannian manifold $(\Sigma, \overline{g})$ and a smooth symmetric 2-tensor K on Σ is called a (smooth) vacuum initial data set for (1.1) (i.e., with $\mathbb{T} = 0$) if it satisfies the constraint equations

$$\overline{R} + (\overline{\mathrm{tr}}\, K)^2 - |K|_{\overline{g}}^2 = 2\Lambda, \quad \overline{\mathrm{div}}\, K - \mathrm{d}\,\overline{\mathrm{tr}}\, K = 0. \tag{1.3}$$

Theorem 1 ([39, 40, 180, 187]). *Let $(\Sigma, \overline{g}, K)$ be a smooth vacuum initial data set. Then there exists a unique smooth maximum Cauchy development, i.e., an $(\mathcal{M}, g)$ with the property that*

(1) *$(\mathcal{M}, g)$ solves (1.1) with $\mathbb{T} = 0$.*

(2) *There exists an embedding $i : \Sigma \to \mathcal{M}$ such that $i(\Sigma)$ is a Cauchy hypersurface in $(\mathcal{M}, g)$ and such that the induced metric and second fundamental form of the embedding agree with $\overline{g}$ and K.*

(3) *If $(\tilde{\mathcal{M}}, \tilde{g})$ also satisfies (1) and (2), then there exists an isometric embedding $(\tilde{\mathcal{M}}, \tilde{g}) \to (\mathcal{M}, g)$ commuting with the embeddings of Σ.*

We remark that the constraint equations (1.3) are the Gauss and Gauss–Codazzi equations induced by (1.1) on Σ and thus necessary conditions. We also note that for simplicity we have stated the theorem in the smooth category although one typically proves Sobolev versions of the result. Finally, the manifold $\mathcal{M}$ is diffeomorphic to $\mathbb{R} \times \Sigma$ [21].

Theorem 1 allows one to talk sensibly about the dynamics of solutions to the Einstein equations. This will allow us to formulate the problem of stability in Section 2.

Proving Theorem 1 requires fixing a gauge. This is a mechanism to eliminate the diffeomorphism covariance of (1.1), i.e., the fact that for any diffeomorphism $\phi : \mathcal{M} \to \mathcal{M}$, the pullback $\phi^* g$ is a solution of (1.1) whenever g is. Consider first a local problem in which $\Sigma = B(0, 1) \subset \mathbb{R}^3$ is the unit ball, and we aim to construct, in a neighborhood of $i(\Sigma) = \{0\} \times \Sigma \subset \mathcal{M}' := \mathbb{R} \times B(0, 1)$, a solution g of (1.1) which induces the data $(\overline{g}, K)$ at the hypersurface $\{0\} \times B(0, 1)$. In local coordinates $(t, x) = (z^0, z^1, z^2, z^3)$ on $\mathcal{M}'$, equation (1.2) for the metric $g = (g_{ij})_{1\leq i,j\leq 4}$ takes the form

$$\begin{aligned}\mathrm{Ric}(g)_{ij} - \Lambda g_{ij} &= -\frac{1}{2} g^{k\ell}\partial_k\partial_\ell g_{ij} + \frac{1}{2}\big(\partial_i W_j(z, g, \partial g) + \partial_j W_i(z, g, \partial g)\big) \\ &\quad + N_{ij}(z, g, \partial g) = 0 \end{aligned} \tag{1.4}$$

(with summation over repeated indices), where (g^{ij}) is the matrix inverse of (g_{ij}), furthermore $N_{ij}(z, g, \partial g)$ is a nonlinear expression in the coefficients of g and its first coordinate derivatives, and finally

$$W_i(z, g, \partial g) = g_{i\ell} g^{jk} \Gamma^\ell_{jk},$$

with $\Gamma^i_{jk} = \Gamma^i_{jk}(g)$ denoting the Christoffel symbols of g. Given any *gauge source functions* $F_i = F_i(z, g)$, we then aim to solve equation (1.4) in the *generalized harmonic gauge* $W_i(z, g, \partial g) = F_i(z, g)$. (The special case $F_i = 0$ is the *wave coordinate gauge*.[1]) Inserting this gauge condition into (1.4), one obtains a system of quasilinear wave equations for the metric coefficients g_{ij}, with principal part given by $-\frac{1}{2}\Box_g g_{ij}$. We can write this system in the compact form

$$P(g) := \mathrm{Ric}(g) - \Lambda g - \delta^*_g (W - F) = 0, \tag{1.5}$$

where $W = W_i \mathrm{d}z^i$ and $F = F_i \mathrm{d}z^i$, and δ^*_g is the symmetric gradient defined by $(\delta^*_g \omega)_{ij} = \frac{1}{2}(\omega_{i;j} + \omega_{j;i})$ where $\omega_{i;j} := (\nabla_{\partial_j}\omega)(\partial_i)$. For future purposes, note that $P(g) = 0$ is a quasilinear wave equation when g is Lorentzian, whether $W - F = 0$ or not. Equation (1.5) is called the *gauge-fixed Einstein (vacuum) equation*.

One now solves the initial value problem in the gauge $W - F = 0$ as follows:

(1) One constructs (algebraically, i.e., without having to solve any differential equation) smooth Cauchy data (g^0, g^1), where $g^\mu = (g^\mu_{ij}(x))$ is a spacetime symmetric 2-tensor (symmetric 4×4 matrix) on Σ, in such a way that $\overline{g}$, resp. K are the induced metric, resp. second fundamental form of $\{0\} \times \Sigma$ induced by any spacetime metric g on $\mathcal{M}'$ with $(g^0, g^1) = (g|_{t=0}, \partial_t g|_{t=0})$. Moreover, the flexibility in the choice of (g^0, g^1) is used to ensure that the gauge condition $W - F = 0$ is satisfied at $t = 0$; the verification of this condition indeed only requires knowledge of the Cauchy data (g^0, g^1).

[1] The terminology arises from the fact that $W^i = \Box_g z^i$ where $\Box_g = |g|^{-1/2}\partial_i |g|^{1/2} g^{ij}\partial_j$ is the scalar wave operator. That is, in the wave coordinate gauge, the coordinate functions z^i satisfy the homogeneous wave equation.

(2) One solves the initial value problem $P(g)=0$, $(g,\partial_t g)|_{t=0}=(g_0,g_1)$ for g. Since this is a quasilinear wave equation, one has local existence and uniqueness of solutions (but solutions may develop singularities in finite time).

(3) The constraint equations (1.3) together with $P(g)=0$ and $W-F=0$ at $i(\Sigma)$ can be shown to imply, via a direct computation, that also $\partial_t(W-F)=0$ at Σ.

(4) The second Bianchi identity is equivalent to the statement that for any metric g, one has $\operatorname{div}_g \mathsf{G}_g \mathrm{Ric}(g)=0$. Applying $\operatorname{div}_g \mathsf{G}_g$ to equation (1.5) thus produces a decoupled equation for $W-F$,

$$\operatorname{div}_g \mathsf{G}_g \delta_g^*(W-F)=0.$$

This is a homogeneous wave equation with principal part $-\frac{1}{2}\Box_g(W-F)$. Since $W-F$ has trivial Cauchy data at $i(\Sigma)$, we conclude that $W-F\equiv 0$ in the domain of dependence of $i(\Sigma)$ with respect to the metric g.

(5) Plugging $W-F=0$ into (1.5) shows that also $\mathrm{Ric}(g)-\Lambda g=0$ in the same domain.

This argument also shows that solutions of the Einstein vacuum equations obey finite speed of propagation. Thus, passing from local solutions to the maximal Cauchy development can be accomplished by carefully gluing together local solutions.

We point out that different choices of gauge source functions F_i may cause singularities to form at different subsets of spacetime. For controlling the *global* evolution of spacetimes, it is thus of central importance to make a well-informed choice of F_i; there is no known method to make an optimal or even a good choice in general. A particularly geometric choice can be made when $\mathcal{M}=\mathbb{R}\times\Sigma$ is already equipped with a "background metric" g^0: in this case one can take $F_i=g_{i\ell}g^{jk}\Gamma^{\ell}_{jk}(g^0)$, and the gauge condition $W-F=0$ is equivalent to the requirement that the pointwise identity map $(\mathcal{M},g)\to(\mathcal{M},g^0)$ be a wave map [93].

An attractive feature of (1.5) is that it highlights the principally scalar (albeit tensorial) character of the gauge-fixed Einstein equation. Thus, many aspects of its analysis are no more difficult than for the linear scalar wave equation, while one retains the flexibility to work in particular coordinates, or splittings of the tangent or cotangent bundles, wherever needed. Further aspects of (variants of) (1.5) and of generalized harmonic gauges will be discussed in Section 2.4.2.

1.3. Double null gauge and the characteristic initial value problem

A particularly geometrically adapted gauge to write the Einstein equation in is the *double null gauge*. The idea is to foliate the spacetime by ingoing and outgoing null hypersurfaces which intersect in (spacelike) 2-manifolds. From a physical perspective, one may expect that this will reveal important structure in the equations as gravitational waves propagate along null hypersurfaces. Indeed, the double null gauge has been successfully employed in several seemingly unrelated contexts, for instance, the formation of black holes [44], the stability of black holes (see Section 2.4.1), the theory of impulsive gravitational waves [153,154] and the construction of naked singularities (see Section 3.2).

Let $\mathcal{W} \subset \mathbb{R}^2$ be a nonempty open subset. A double null gauge on a manifold $\mathcal{M} = \mathcal{W} \times \mathbb{S}^2$ is a coordinate system $(u, v, \theta^1, \theta^2)$ such that the metric takes the form

$$g = -4\Omega^2 \mathrm{d}u\, \mathrm{d}v + \slashed{g}_{AB}(\mathrm{d}\theta^A - b^A\, \mathrm{d}v)(\mathrm{d}\theta^B - b^B\, \mathrm{d}v), \tag{1.6}$$

where Ω is a spacetime function, $b : \mathcal{M} \to T\mathcal{M}$ an $\mathbb{S}^2_{u,v}$-vector and $\slashed{g}$ the induced metric on the spheres $\mathbb{S}^2_{u,v} = \{(u,v)\} \times \mathbb{S}^2$.[2] We remark that any Lorentzian metric can be locally put into this form by solving the eikonal equation associated with the metric g.

Associated with a double null foliation is a local double null frame given by

$$e_3 = \frac{1}{\Omega}\partial_u, \quad e_4 = \frac{1}{\Omega}(\partial_v + b^A \partial_{\theta^A}), \quad e_A = \partial_{\theta^A}. \tag{1.7}$$

The vectors e_3 and e_4 are null, and the frame satisfies the normalization conditions $g(e_3, e_4) = -2$, $g(e_3, e_A) = 0 = g(e_4, e_A)$, and $g(e_A, e_B) = \slashed{g}_{AB}$.

Given a double null gauge, one can define Ricci coefficients and curvature components with respect to the above frame (all these are the coefficients of $\mathbb{S}^2_{u,v}$-tensors)

$$\begin{aligned}
\chi_{AB} &= g(\nabla_A e_4, e_B), & \underline{\chi}_{AB} &= g(\nabla_A e_3, e_B), & \eta &= -\frac{1}{2} g(\nabla_{e_3} e_A, e_4),\\
\hat{\omega} &= \frac{1}{2} g(\nabla_{e_4} e_3, e_4), & \underline{\hat{\omega}} &= \frac{1}{2} g(\nabla_{e_3} e_4, e_3), & \underline{\eta} &= -\frac{1}{2} g(\nabla_{e_4} e_A, e_3),\\
\alpha_{AB} &= R(e_A, e_4, e_B, e_4), & \underline{\alpha}_{AB} &= R(e_A, e_3, e_B, e_3), & \beta &= \frac{1}{2} R(e_A, e_4, e_3, e_4),\\
\rho &= \frac{1}{4} R(e_4, e_3, e_4, e_3), & \sigma &= \frac{1}{4} {}^\star R(e_3, e_4, e_4, e_3), & \underline{\beta} &= \frac{1}{2} R(e_A, e_3, e_3, e_4),
\end{aligned}$$

and write out the null-structure equations (which relate the intrinsic and extrinsic geometry of the spacetime foliation) in terms of $\mathbb{S}^2_{u,v}$-tensors; this leads to a system of transport equations along the null cones, and elliptic equations on the spheres. An example of a transport equation is (note that $\alpha, \underline{\alpha}$ are $\slashed{g}$-traceless as a consequence of the Einstein equations)

$$\slashed{\nabla}_4 \hat{\chi} + (\mathrm{tr}\, \chi)\hat{\chi} - \hat{\omega}\hat{\chi} = -\alpha, \tag{1.8}$$

where $\hat{\chi}$ denotes the $\slashed{g}$-traceless part of χ and $\slashed{\nabla}_4$ denotes the projected covariant derivative in the e_4 direction. An example of an elliptic equation is

$$\slashed{\mathrm{d}}\mathrm{iv}\, \hat{\chi} = -\frac{1}{2}\hat{\chi} \cdot (\eta - \underline{\eta}) + \frac{1}{4}\mathrm{tr}\, \chi(\eta - \underline{\eta}) + \frac{1}{2}\slashed{\nabla}\, \mathrm{tr}\, \chi - \beta, \tag{1.9}$$

where $\slashed{\mathrm{d}}\mathrm{iv}$ denotes the $\slashed{g}$-divergence on $\mathbb{S}^2_{u,v}$. The analytical content of the Einstein equations (1.1) is then captured by these structure equations in conjunction with the Bianchi equations, which capture the essential hyperbolicity of (1.1). An example of two such null-decomposed equations is

$$\slashed{\nabla}_3 \alpha + \frac{1}{2}(\mathrm{tr}\underline{\chi})\alpha + 2\underline{\hat{\omega}}\alpha = -2\slashed{\mathcal{D}}_2^\star \beta - 3\rho\hat{\chi} - 3{}^\star\hat{\chi}\sigma + \frac{1}{2}(9\eta - 2\underline{\eta})\,\hat{\otimes}\,\beta, \tag{1.10}$$

$$\slashed{\nabla}_4 \beta + 2(\mathrm{tr}\, \chi)\beta - \hat{\omega}\beta = \slashed{\mathrm{d}}\mathrm{iv}\, \alpha + \eta \cdot \alpha, \tag{1.11}$$

where $\slashed{\mathcal{D}}_2^\star$ denotes the symmetric traceless part of the covariant derivative $\slashed{\nabla}$ on $(\mathbb{S}^2_{u,v}, \slashed{g})$, $\hat{\otimes}$ the symmetric traceless tensor product, and * is the Hodge-star operator.

[2] Tensors $\mathbb{S}^2_{u,v}$ can be canonically identified with spacetime tensors having the property that any contraction with the null directions in (1.7) is identically zero.

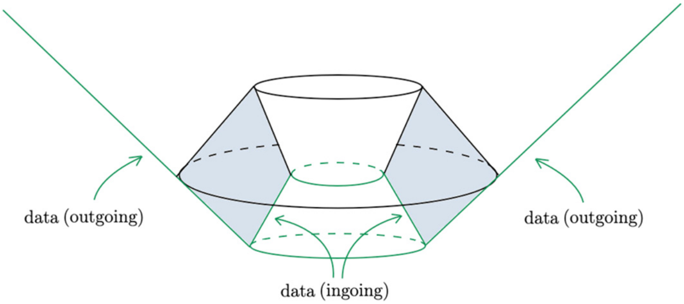


FIGURE 1
The characteristic initial value problem and the double null foliation. Data is specified on the green ingoing and outgoing cones and the solution exists in the grey shaded region.

As mentioned above, given a solution of the Einstein equations, we can locally put the metric into a double null gauge. Conversely, one can *construct* local solutions to the Einstein equations in a double null gauge by solving a characteristic initial value problem, where initial data are prescribed on two intersecting null cones (see Figure 1).

Theorem 2 ([130, 149, 178]). *Consider suitable smooth vacuum initial data prescribed on two (what will be) null hypersurfaces intersecting transversally on a spacelike* 2*-sphere* $S_0 = N_1 \cap N_2$*. Then there exists a nonempty maximum development* $(\mathcal{M}, g)$ *which is bounded in the past by a neighborhood of* S_0 *in* $N_1 \cup N_2$*.*

The proof of the theorem reduces the problem to the situation of Theorem 1 (the hypersurface S_0 can in fact be any two-dimensional spacelike surface).

As we shall see, the relevance of the double null gauge is most apparent in (semi)-global problems through the way it allows to estimate the solution. We mention explicitly already Luk's result [149] which provides estimates on the size of the corresponding maximum development in Theorem 2 by means of exploiting the null structure in the equations.

We have not given a precise notion of a vacuum initial data set prescribed on intersecting null hypersurfaces in Theorem 2. The definition and the procedure to construct such data can be found in [44]. Roughly speaking, in canonical coordinates (1.6) and using stereographic coordinates on the sphere, the free data correspond to prescribing, in a smooth fashion, a symmetric traceless 2×2-matrix along each of the initial cones as well as the mean curvatures and the torsion at the sphere of intersection. All remaining geometric quantities are then determined by solving ordinary differential equations along the initial cones.

1.4. Explicit solutions

1.4.1. Maximally symmetric solutions

The simplest solution to (1.1) for $\Lambda = 0$ is *Minkowski space* $(\mathbb{R}^4, \eta)$, where $\eta = -\mathrm{d}t^2 + \sum_{j=1}^{3} (\mathrm{d}x^j)^2$ in standard coordinates (t, x^1, x^2, x^3) on $\mathbb{R}^4$. It is geodesically complete and maximally symmetric in that the spacetime admits the maximum number of

Killing vectors, namely 10. These correspond to the infinitesimal generators of the Poincaré group of special relativity (spacetime translations, spatial rotations, Lorentz boosts). Passing to polar coordinates (r, θ, ϕ) on $\mathbb{R}^3$ and denoting by $\mathring{g} = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ the standard metric on $\mathbb{S}^2$, we have

$$\eta = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\mathring{g} = -\mathrm{d}U\,\mathrm{d}V + r^2(U,V)\mathring{g}, \quad U = t + r,\ V = t - r,$$

with $r(U,V) = \frac{1}{2}(U - V)$, and $(U,V) \in (-\infty,\infty) \times (-\infty,\infty)$ is restricted to the subset where $r(U,V) \geq 0$. Since η is spherically symmetric, we can give a simple description of the causal geometry by depicting the (U,V) plane, compactified at $U = \infty$ and at $V = -\infty$. See Figure 2. The ideal boundary at $U = \infty$, resp. $V = -\infty$, is called *future*, resp. *past null infinity* ($\mathscr{I}^+$, resp. $\mathscr{I}^-$).

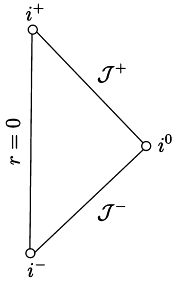


FIGURE 2
Penrose diagram of the Minkowski spacetime.

Certain approaches [116] to the analysis of wave equations on the Minkowski spacetime (or suitable perturbations thereof [18, 19]) instead focus first on the fact that η is homogeneous of degree -2 with respect to scaling in (t, x). Thus, one attaches an ideal boundary at $|(t,x)| = \infty$ by passing to the radial (or projective) compactification $\overline{\mathbb{R}^4}$ of $\mathbb{R}^4$, which is a closed 4-ball. All forward light cones intersect the ideal boundary in the same 2-sphere (analogously for backward light cones). Resolving these by means of real blow-up produces, as front faces, future and past null infinity. The resulting manifold with corners can also be regarded as a blow-up of the Penrose diagram at i^0 and $i^\pm$, see Figure 3.

The maximally symmetric analogue of Minkowski space for $\Lambda > 0$ is called *de Sitter space*. This can be defined as the cylinder

$$\mathcal{M} = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)_s \times \mathbb{S}^3, \quad g = \Omega^{-2}\overline{g}, \quad \Omega^2 = \frac{\Lambda}{3}\cos^2 s, \quad \overline{g} = -\mathrm{d}s^2 + g_{\mathbb{S}^3}. \tag{1.12}$$

The boundary at $s = \pm\frac{\pi}{2}$ is called the *future/past conformal boundary*. Since null-geodesics of conformally related metrics are the same up to reparametrization, the causal structure of $(\mathcal{M}, g)$ is the same as that of $\mathcal{M}$ equipped with the smooth (down to $s = \pm\frac{\pi}{2}$) metric $\overline{g}$. See Figure 4.

Due to the finite speed of propagation for solutions of wave equations, one can consider wave equations in a *static patch of de Sitter space*, which is the intersection of the

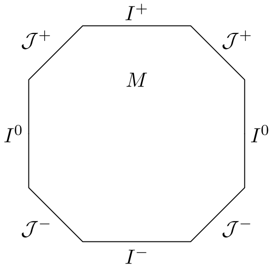


FIGURE 3
Resolution (blow-up) of the radial compactification $\overline{\mathbb{R}^4}$.

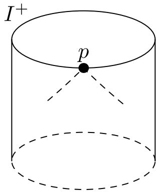


FIGURE 4
(Global) de Sitter space. Also shown is part of the backwards light cone from a point p on the future conformal boundary I^+.

timelike past of a point $p = (\frac{\pi}{2}, q)$ and the timelike future of $(-\frac{\pi}{2}, q)$. One can introduce coordinates in such a domain in which the de Sitter metric is static,

$$g = -\left(1 - \frac{\Lambda}{3}r^2\right)\mathrm{d}t^2 + \left(1 - \frac{\Lambda}{3}r^2\right)^{-1}\mathrm{d}r^2 + r^2\mathring{\not{g}}. \tag{1.13}$$

The singularity of this expression at the *cosmological horizon* $r^{-1}(\sqrt{3/\Lambda})$ (which in $s > 0$ is the backwards light cone with vertex p) is a coordinate singularity.

Finally, the maximally symmetric spacetime with $\Lambda < 0$ is called anti-de Sitter space (AdS). This is the manifold $\mathbb{R}^4$ equipped with the metric (1.13) where now $\Lambda < 0$. Using the transformation $r = \tan\psi$ (with $\psi \in (0, \pi/2)$) the metric (1.13) can be written as $g = \frac{1}{\cos^2\psi}(-dt^2 + d\psi^2 + \sin^2\psi\,\mathring{\not{g}})$ from which it becomes apparent that AdS is conformal to $\mathbb{R} \times \mathbb{S}^3_h \subset \mathbb{R} \times \mathbb{S}^3$ ($\mathbb{S}^3_h$ denoting a hemisphere of $\mathbb{S}^3$), equipped with the metric $-dt^2 + g_{\mathbb{S}^3}$, i.e., conformal to one “half” of the Einstein static cylinder. The timelike boundary $\psi = \frac{\pi}{2}$ in the conformal picture corresponds to the timelike *conformal infinity* ($r = \infty$) of anti-de Sitter space. The spacetime is not globally hyperbolic and boundary conditions will have to be imposed to get a well-posed evolution for hyperbolic equations on or near these backgrounds. See Figure 5.

1.4.2. The Schwarzschild manifold

A nontrivial solution to the Einstein equations was found by Schwarzschild in 1915. The Schwarzschild solution describes what we (today) call a black hole solution. We follow a

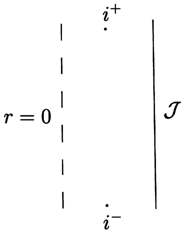


FIGURE 5
The Penrose diagram of anti-de Sitter space with its timelike conformal boundary $\mathcal{J}$.

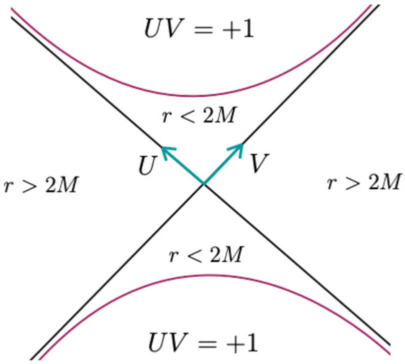


FIGURE 6
The causal geometry of the maximally extended Schwarzschild manifold.

somewhat revisionist approach in presenting the metric which however emphasizes directly its geometric nature and its connection to the double null gauge introduced in Section 1.3.

Given $M>0$, equip $\mathcal{M}=(-\infty,\infty)_U\times(-\infty,\infty)_V\times\mathbb{S}^2\cap\{UV<1\}$ with the metric

$$g_M=-4\Omega_K^2 \mathrm{d}U\,\mathrm{d}V+r^2(U,V)\mathring{\not{g}},\quad \Omega_K^2=\frac{8M^3}{r}\exp\left(-\frac{r}{2M}\right),$$

where $r:(-\infty,\infty)\times(-\infty,\infty)\to\mathbb{R}^+$ is defined implicitly by

$$\left(\frac{r(U,V)}{2M}-1\right)\exp\left(\frac{r(U,V)}{2M}\right)=-UV.$$

We time-orient $(\mathcal{M},g_M)$ by declaring $\partial_U+\partial_V$ to be future directed. The metric is spherically symmetric, and we can give a simple depiction of the causal geometry by depicting the (U,V)-plane below. We observe that for the region $U<0$, $V>0$ we must have $r<2M$; moreover, any future directed curve causal curve emanating from this region remains in this region, has finite affine length, and terminates on the asymptotic boundary $UV=1$, where $r\to 0$ and the Kretschmann scalar $R_{\mu\nu\sigma\nu}R^{\mu\nu\sigma\nu}$ blows up like r^{-6}. It follows that the spacetime is geodesically incomplete and $\mathcal{C}^2$-inextendible (in fact, $\mathcal{C}^0$-inextendible [188]) as a Lorentzian manifold. One may compactify the U and the V coordinates to produce the well-known Penrose-diagram of the Schwarzschild metric, see Figures 6 and 7.

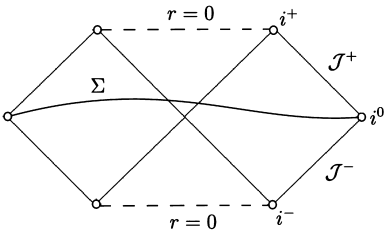


FIGURE 7
Penrose diagram of the Schwarzschild manifold.

The set $r = \infty$ is now realized as a (null) boundary of the spacetime, and we can define the black hole region as $\mathcal{M} \setminus J^-(\mathscr{I}^+)$, i.e., as the set of observers that cannot communicate with asymptotic observers in the far away region of spacetime. The black hole region is bounded by the set $r^{-1}(2M)$, which is a union of null hypersurfaces.

We finally note that if we restrict to the black hole exterior region $U > 0$, $V > 0$, then the sequence of coordinate transformations $U = -e^{-\frac{u}{2M}}$, $V = e^{\frac{v}{2M}}$, $u = \frac{t-r^\star}{2}$, $v = \frac{t+r^\star}{2}$, where $\frac{dr^\star}{dr} = \frac{1}{1-\frac{2M}{r}}$, brings the metric into the standard (static) form where the area radius r is used as a coordinate:

$$g = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathring{g}. \tag{1.14}$$

This coordinate system breaks down when r equals the Schwarzschild radius $r = 2M$; coordinates valid across $r = 2M$ (besides the Kruskal coordinates U, V above) are discussed in Section 1.4.3. Expression (1.14) shows that the Schwarzschild metric is stationary (∂_t is a Killing vector field and timelike for large r, or indeed for all $r > 2M$).

An important feature of the Schwarzschild metric and other black hole spacetimes discussed below is the existence of *trapped null-geodesics* in the exterior region $r > 2M$, i.e., future and past inextendible null-geodesics which when quotienting out by time translations (i.e., projecting to the (r, θ, ϕ) variables) remain in a compact subset of $\{r > 2M\}$. The trapped set is defined as the subset of phase space $T^*\mathcal{M}$ consisting of all (z, ζ) so that the null-geodesic with initial position z and initial momentum $\zeta \in T^*\mathcal{M}$, $\zeta \neq 0$, is trapped. Writing

$$\zeta = \sigma\,\mathrm{d}t + \xi\,\mathrm{d}r + \eta, \quad \eta \in T^*\mathbb{S}^2,$$

the trapped set of the Schwarzschild spacetime is the conic set

$$\Gamma = \left\{\zeta \in T^*\mathcal{M} \setminus o : r = 3M, \xi = 0, |\eta|^2_{\mathring{g}^{-1}} = 27M^2\sigma^2\right\}. \tag{1.15}$$

Its projection to the base manifold $\mathcal{M}$ is the hypersurface $r = 3M$. The trapped set is unstable, and indeed ν-normally hyperbolic for all ν [118], as will be discussed in more detail in Section 2.3.4.

We finally mention the important *red-shift effect* [171], which is in fact a general feature of nondegenerate black hole horizons. In the geometric optics approximation, it manifests itself by the frequency of waves (measured with respect to an appropriate notion of time) being shifted towards longer (i.e., less energetic) frequencies as they propagate near the event horizon. For hyperbolic equations, the red-shift effect can be captured by a physical space energy identity with good coercive properties near the horizon [63]. From the viewpoint of microlocal analysis these are the radial estimates of [201] (see Section 2.3.4).

1.4.3. Further spherically symmetric spacetimes

The Schwarzschild solution generalizes to the Reissner–Nordström–((anti)-de Sitter) solution of the Einstein–Maxwell equations with cosmological constant (1.1) (we omit the explicit formulas for the electromagnetic field here). The line element in so-called static coordinates is

$$g = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \frac{\Lambda}{3}r^2\right)dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \frac{\Lambda}{3}r^2\right)^{-1} dr^2 + r^2\mathring{\slashed{g}}. \quad (1.16)$$

Near the zeros of $F(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \frac{\Lambda}{3}r^2$, one needs to pass to other coordinate systems to unravel the maximally extended spacetimes shown in the Penrose diagrams below. For instance, near the event horizon $r = r_+$ where F changes sign from $-$ to $+$, one can introduce ingoing Eddington–Finkelstein coordinates, $v = t + \int F^{-1}\,\mathrm{d}r$, in which

$$g = -F(r)\,\mathrm{d}v^2 + 2\,\mathrm{d}v\,\mathrm{d}r + r^2\mathring{\slashed{g}}.$$

For $\Lambda = 0$ but nonzero subextremal charge $0 \neq |Q| < M$, the Penrose diagram of the Reissner–Nordström spacetime differs dramatically from that of the Schwarzschild spacetime in the black hole region: while there still is an event horizon at $r = r_+ := M + \sqrt{M^2 - Q^2}$, there is now also a future/past *inner horizon* (or *Cauchy horizon*) at $r = r_- := M - \sqrt{M^2 - Q^2}$ across which the metric extends analytically. The Cauchy horizon is the boundary of the maximal Cauchy development of the initial data at the hypersurface Σ indicated in Figure 8. For $\Lambda > 0$ and $Q = 0$, the metric (1.16) is a vacuum solution of (1.1) and called the Schwarzschild–de Sitter (SdS) metric; we consider only the subextremal case $0 < 9M\Lambda^2 < 1$. Its geometry near the black hole and near the event horizon $r = r_-$ (the smaller positive root of $1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2 = 0$) is then the same as for the

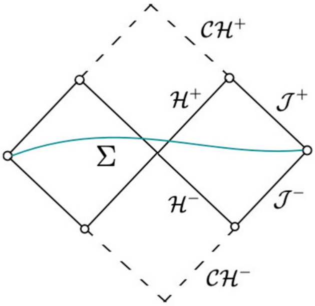


FIGURE 8
A piece of the maximal analytic extension of the Reissner–Nordström spacetime.

Schwarzschild metric, but now there is also a second horizon, called *cosmological horizon*, at the larger positive root $r = r_+$ of $1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2 = 0$; this is analogous to the cosmological horizon of the static patch of de Sitter space. The metric extends analytically past this horizon and asymptotes to the de Sitter metric as $r \to \infty$. See Figure 9.

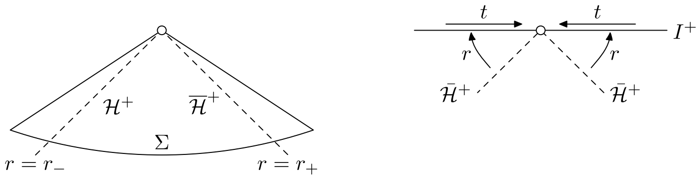


FIGURE 9

(Left) Penrose diagram of a neighborhood $r_- - \varepsilon < r < r_+ + \varepsilon$ of the domain of outer communications of a subextremal Schwarzschild–de Sitter (SdS) spacetime near the causal future of a hyperboloidal spacelike slice Σ. (Right) An illustration of a SdS black hole with a focus on its asymptotically de Sitter geometry far from the black hole; $\overline{\mathcal{H}^+}$ denotes the cosmological horizon.

The (subextremal) Reissner–Nordström–de Sitter spacetime has an event horizon and a cosmological horizon just like the Schwarzschild–de Sitter spacetime. Only the structure of the black hole interior depends on whether $Q = 0$ (in which case there is a terminal singularity as in the Schwarzschild case) or $Q \neq 0$ (in which case there is a Cauchy horizon across which the metric extends analytically).

For $\Lambda < 0$ and $Q = 0$, the metric (1.16) is a vacuum solution of (1.1) and called the Schwarzschild–anti-de Sitter metric. Its crucial geometric features are the timelike conformal boundary at infinity (which is future complete) and the future complete event horizon located at the unique real zero of $F(r)$.

All these spherically symmetric black hole spacetimes have trapped sets of the same form as (1.15), with $3M$ and $27M^2$ replaced by appropriate constants.

1.4.4. The Kerr metric and related metrics

In 1963 Roy Kerr found a generalization of the Schwarzschild family of metrics to a family of vacuum solutions of (1.1) (with $\Lambda = 0$) which incorporates also angular momentum. For parameters $M > 0$ and $a \in [-M, M]$, and setting $r_+ = M + \sqrt{M^2 - a^2}$, the *Kerr family of metrics*, in Boyer–Lindquist coordinates $t \in \mathbb{R}$, $r \in (r_+, \infty)$, $\theta \in (0, \pi)$, $\phi \in (0, 2\pi)$, takes the form

$$\begin{aligned} g_{M,a} &= -\frac{\Delta}{\varrho^2}(\mathrm{d}t - a\sin^2\theta\,\mathrm{d}\phi)^2 + \varrho^2\left(\frac{\mathrm{d}r^2}{\Delta} + \mathrm{d}\theta^2\right) + \frac{\sin^2\theta}{\varrho^2}\big(a\,\mathrm{d}t - (r^2 + a^2)\,\mathrm{d}\phi\big)^2, \\ \Delta &= r^2 - 2Mr + a^2, \quad \varrho^2 = r^2 + a^2\cos^2\theta. \end{aligned} \tag{1.17}$$

For $a = 0$, this reduces to (1.14).

For now, we focus on the *subextremal range* $a \in (-M, M)$. For $a \neq 0$, the Penrose diagram of suitable two-dimensional timelike slices of the maximal analytic extension of the Kerr spacetime has the same form as that of the Reissner–Nordström spacetime, see Figure 8. In particular, there is an event horizon at $r = r_+$ and a Cauchy horizon at $r = r_- := M - \sqrt{M^2 - a^2}$.

Furthermore, there is a trapped set Γ, which as in the Schwarzschild case is a smooth (in fact, analytic) conic submanifold $\Gamma \subset T^*\mathcal{M} \setminus o$ of phase space over the black hole exterior $r > r_+$; it is an ν-normally hyperbolic (for every ν) invariant submanifold for the lift of the null-geodesic flow to $T^*\mathcal{M}$, as first noted by Wunsch–Zworski [210] and proved in the full subextremal range by Dyatlov [76]. The projection of Γ to the base $\mathcal{M}$ is however no longer a smooth submanifold, but rather a full-dimensional closed set (with compact intersection with any t-level set) with non-empty interior.

Another novel feature of rotating Kerr metrics is the presence of *superradiance*. This means that the energy $-g_{M,a}(\dot{\gamma}, \partial_t)$ of a future lightlike geodesic γ with respect to the generator ∂_t of time translations may be *negative*; here ∂_t is the unique (up to scaling) Killing vector field which for sufficiently large r/M is future timelike. This is the basis of the Penrose effect for energy extraction from rotating black holes. On the level of analysis, this problem is overcome by means of so-called red-shift or radial point estimates.

We mention a geometric and an algebraic fact about the Kerr metric. Firstly, there exists a global double null foliation on the Kerr manifold (constructed in [174]). Secondly, there exists a (nonintegrable) null-frame, called the algebraically special frame, on the Kerr manifold with respect to which all but the curvature components ρ and σ (defined as in Section 1.3 but for the null frame being the algebraically special frame) vanish.

Finally, we note that for *extremal Kerr black holes*, with $|a| = M$, the event horizon at $r = M$ degenerates (the function Δ in (1.17) has a double zero). Furthermore, the trapped set now extends down to the horizon and ceases to be normally hyperbolic [76].

The generalization of (1.17) allowing for the presence of a cosmological constant Λ and an electric charge Q was found by Carter [36], following the discovery [170] of the charged analogue in the case $\Lambda = 0$. It is called the Kerr–Newman–((anti)-de Sitter) metric,

$$\begin{aligned} g_{M,a,\Lambda,Q} &= -\frac{\Delta}{(1+\lambda)\varrho^2}(\mathrm{d}t - a\sin^2\theta\,\mathrm{d}\phi)^2 + \varrho^2\left(\frac{\mathrm{d}r^2}{\Delta} + \frac{\mathrm{d}\theta^2}{\kappa}\right) \\ &\quad + \frac{\kappa\sin^2\theta}{(1+\lambda)^2\varrho^2}\big(a\,\mathrm{d}t - (r^2+a^2)\,\mathrm{d}\phi\big)^2, \\ \lambda &= \frac{\Lambda}{3}a^2, \quad \kappa = 1 + \lambda\cos^2\theta, \quad \varrho^2 = r^2 + a^2\cos^2\theta, \\ \Delta &= (r^2+a^2)\left(1 - \frac{\Lambda}{3}r^2\right) - 2Mr + (1+\lambda^2)Q^2. \end{aligned} \tag{1.18}$$

(We again omit the explicit expression for the electromagnetic field.) For $Q = 0$ and $\Lambda > 0$ ($\Lambda < 0$), this is called the Kerr–(anti-)de Sitter metric. For simplicity, in these notes we restrict attention to the case of small angular momenta a and small charges Q; in this case, the Penrose diagram of suitable two-dimensional slices of a neighborhood of the black hole exterior region of Kerr–Newman–de Sitter spacetimes is the same as the one of a SdS spacetime, as shown in Figure 9. Again, there is a trapped set with the same (phase space and physical space) structure as in the subextremal Kerr case.

1.5. Matter models

In the notation of equation (1.1), we have so far restricted ourselves to the vacuum case $\mathbb{T} = 0$. However, real world physical systems typically involve matter. We briefly discuss the most common matter models studied in connection with the Einstein equations. In each of these cases, the stated expression for $\mathbb{T}$ arises by direct calculation from the Euler–Lagrange equation for a suitable Lagrangian (the Einstein–Hilbert action plus additional terms describing the matter).

For real-valued *scalar fields* ϕ with mass m, one takes

$$\mathbb{T}_{\mu\nu} = (\partial_\mu \phi)(\partial_\nu \phi) + m^2\phi^2 - \frac{1}{2} g_{\mu\nu} |\nabla\phi|_g^2. \tag{1.19}$$

The second Bianchi identity implies that for a solution of (1.1) with this energy–momentum tensor, ϕ necessarily solves the Klein–Gordon equation (for $m = 0$ the wave equation)

$$(\Box_g - m^2)\phi = 0.$$

For *electromagnetic fields* $F = F_{\mu\nu} \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$, one takes

$$\mathbb{T}_{\mu\nu} = g^{\alpha\beta} F_{\alpha\mu} F_{\beta\mu} - \frac{1}{4} F^{\alpha\beta} F_{\alpha\beta} g_{\mu\nu}. \tag{1.20}$$

The second Bianchi identity gives as the equations of motion the Maxwell equations

$$\mathrm{d}F = 0, \quad \mathrm{div}_g\, F = 0.$$

On spacetimes with nonzero electromagnetic fields $F = \mathrm{d}A$, one can also consider *charged scalar fields*; they are sections of a complex line bundle satisfying a wave equation defined with respect to the connection $\mathrm{d} - iA$.

Finally, uncharged collisionless (“Vlasov”) matter with mass $m \geq 0$ is described by a density distribution $f : T\mathcal{M} \to [0, \infty)$ with support in the set of future causal v with $g(v, v) = -m^2$; the energy–momentum tensor at the point $p \in \mathcal{M}$ is

$$\mathbb{T}_{\mu\nu}(p) = \int_{T_p\mathcal{M}} f(p, v) v_\mu v_\nu, \quad v_\mu = g_{\mu\nu} v^\nu. \tag{1.21}$$

The equation of motion for the density f is the transport equation $Xf = 0$, where X is the geodesic vector field on $T\mathcal{M}$.

2. THE STABILITY OF BLACK HOLE SOLUTIONS

Before we turn to the discussion of the stability of the black hole solutions (described in Section 1.4) in Section 2.2, we record what is known about the stability of the maximally symmetric solutions.

2.1. Prelude: stability of maximally symmetric solutions

The sign of the cosmological constant has a dramatic effect on the global structure of the maximally symmetric solutions, and thus we discuss the three cases separately.

2.1.1. $\Lambda = 0$

For $\Lambda = 0$, we have the following seminal result:

Theorem 3 ([45]). *Minkowski spacetime $(\mathbb{R}^4, \eta)$ is nonlinearly asymptotically stable.*

We note earlier work of Friedrich [87] proving a version of the above theorem for initial data which are exactly Schwarzschildean outside a compact set (such data were later constructed in [47, 48]) or prescribed on a hyperboloidal slice ending at null infinity.

While the original proof of Theorem 3 is closer in spirit to the analysis of the equations in double null form (in particular, [45] estimates curvature and Ricci coefficients instead of metric components), a simplified proof of the theorem (with weaker conclusions regarding the asymptotics) was later given in harmonic gauge by Lindblad–Rodnianski [146].

Studying the stability of flat space is still an active area of research with many new developments regarding regularity [23], optimal asymptotic decay rates [116, 145], and coupling to various matter models. A particularly interesting direction is to consider flat space as a solution to the (massive or massless) Einstein–Vlasov system. Unlike for the scalar field or electromagnetic radiation, the matter does not satisfy wave-type equations but instead transport equations (see Section 1.5). This requires a several new ideas including the construction of various lifts of geometrically adapted vector fields to the mass-shell to identify a suitable version of the null condition in the nonlinearities. In summary we have:

Theorem 4 ([24, 79, 147, 199]). *Minkowski spacetime $(\mathbb{R}^4, \eta)$ is nonlinearly asymptotically stable as a solution of the coupled Einstein–Vlasov system.*

2.1.2. $\Lambda > 0$

The first general nonlinear stability result for the Einstein vacuum equations was obtained for perturbations of de Sitter space $(-\frac{\pi}{2}, \frac{\pi}{2})_s \times \mathbb{S}^3$ by Friedrich [87]: the metric evolving from small and sufficiently regular perturbations of de Sitter initial data at $\Sigma = \{s = 0\}$ can be written as $\Omega^{-2}\overline{g}$ where Ω is positive near Σ and vanishes simply at what becomes the future and past conformal boundary, cf. (1.12). (Thus, the spacetime is *not* asymptotic to de Sitter spacetime as $\Omega \searrow 0$.) Moreover, such asymptotically de Sitter spacetimes can be characterized via suitable asymptotic initial data (two scalar functions and a symmetric 2-tensor K on a Riemannian 3-manifold $(\mathbb{S}^3, h)$) at the future conformal boundary which satisfy *linear* constraint equations (K must be trace- and divergence-free). See Section 4.1 for a recent result making use of this fact on a conceptual level. Extensions of [87] to general dimensions were proved in [4, 179]. A different perspective on the stability of small neighborhoods of the static patch of de Sitter space (in generalized harmonic gauges) was given in [115], see also Section 2.5.2 below.

2.1.3. $\Lambda < 0$

The least understood case is $\Lambda < 0$. Here the Einstein equations become a nonlinear initial boundary value problem for which well-posedness was established in [88]. The question of global stability or instability depends on the boundary conditions imposed at the conformal boundary. The most interesting case are reflective boundary conditions as they preclude any mechanism for energy to be radiated away. (In the case of dissipative boundary conditions, [119] established strong decay for the linearized problem.) The fact that linear fields do not decay lead [55] to conjecture the nonlinear instability of AdS. The problem was first investigated heuristically and numerically for the spherically symmetric scalar field in the influential [25] which proposed a mechanism of energy transfer from low to high frequencies based on resonant interactions. After a large body of works in the theoretical physics literature (see [17, 53, 69] and also the discussion and references in [169]) trying to extend the range of validity of non-linear perturbation theory, Moschidis succeeded in proving the following result:

Theorem 5 ([169]). *Anti-de Sitter spacetime is dynamically unstable as a solution to the spherically symmetric Einstein–Vlasov system.*

The proof proceeds by constructing a one-parameter family of initial data $\mathscr{D}^{\varepsilon}$, consisting of a collection of carefully arranged (both in physical and in momentum space) Vlasov beams, which converges in a suitable topology to the trivial (anti-de Sitter) data as $\varepsilon \to 0$ and is such that for all $\varepsilon > 0$ the maximum development contains a black hole region. Hence, remarkably, (the proof of) Theorem 5 controls the dynamics all the way to the formation of a black hole!

The proof discovers and exploits a nonlinear growth mechanism in physical space (which has no linear analogue and is quite different from the heuristic mechanisms based on resonances for nonlinear perturbations) which relies on the observation that the beams transfer energy to one another when they pass through each other and that this transfer depends on where in spacetime the interaction happens. This observation is at the root of constructing the initial configuration of the beams.

The next natural step is to generalize Theorem 5 to the spherically symmetric Einstein scalar field system. A proof of singularity formation for the Einstein vacuum equation without symmetry assumptions may then be well within reach; this would complete the picture of the vacuum (in)stability of the maximally symmetric solutions.

We finally remark that it is not clear whether instability holds for all (small) data. The existence of geons and "islands of stability" has been widely discussed in the physics literature [94, 122, 184]. For the problem of constructing small data time-periodic solutions in this setting mathematically rigorous progress has recently been made (for semilinear toy problems) in [1].

2.2. The formulation of the stability problem and overview of the results

To formulate the exterior stability problem for black holes it will be useful to distinguish informally the following concepts:[3]

[3] These concepts can be modified in a straightforward manner so that they apply for $\Lambda > 0$ or $\Lambda < 0$ as well.

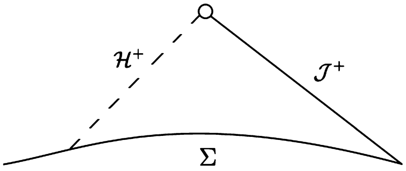


FIGURE 10
Penrose diagram of a (dynamical) black hole spacetime.

(1) *Nonlinear stability:* Given suitable (i.e., characteristic or spacelike and of sufficient regularity) initial data near those of a member of the Kerr family, the associated maximum development $(\mathcal{M}, g)$ has the following properties:

(i) It contains a subset of the form given in Figure 10. In particular, future null-infinity $\mathcal{J}^+$ is complete and $J^-(\mathcal{J}^+)$ is bounded to the future by a regular future complete event horizon $\mathcal{H}^+$.

(ii) Orbital stability (g remains close to $g_{M,a}$ on $\mathcal{M}'$).

(iii) Asymptotic stability ($g \to g_{\tilde{M},\tilde{a}}$ with $|M - \tilde{M}| < \varepsilon$, $|a - \tilde{a}| < \varepsilon$ as an appropriate notion of time goes to infinity).

(2) *Linear stability:* Linearize equations (1.1) in the metric g around a fixed member of the Kerr family; this produces the equations of linearized gravity. Given suitable initial data for this linearized system, prove that, in a suitable gauge, solutions remain bounded (orbital stability) and indeed decay in time to a linearized (in the parameters (M, a)) Kerr metric (asymptotic stability) on the black hole exterior.

(3) *Toy stability:* Given suitable initial data for the toy problem $\Box_{g_{M,a}} \psi = 0$, prove that solutions remain bounded (orbital stability) and decay in time (asymptotic stability) on the black hole exterior.

While the seminal works in the physics literature, starting with [177], concern aspects of the linear stability problem, the first rigorous theorems are due to Wald and Kay [133,207] on toy stability. The results on toy stability have reached a rather complete state in the past decade; this is the content of Section 2.3. With the conceptual and technical insights thus gained, linear and nonlinear black hole stability problems have become accessible, at least in the case $\Lambda \geq 0$, in the past five years. We discuss the current state of knowledge regarding linear stability in Section 2.4, and regarding nonlinear stability in Section 2.5.

2.3. Toy stability

The following theorem summarizes the picture that has been obtained for the analysis of the toy problem in the various black hole geometries.

Theorem 6. *Consider a solution to the scalar wave equation*

$$\Box_{g_{M,a,\Lambda}}\psi = 0 \tag{2.1}$$

on the black hole exterior of a Kerr–((A)dS) spacetime arising from suitable initial data (and, in the Kerr–AdS case, with Dirichlet boundary conditions at the conformal boundary). Then

(1) *If* $\Lambda > 0$ *and* $|a| \ll M$ *then* ψ *decays exponentially in time to a constant* [73].

(2) *If* $\Lambda = 0$ *and* $|a| < M$ *then* ψ *decays inverse polynomially in time* [66]. *(See Theorem 7 below for the extremal case* $|a| = M$*.)*

(3) *If* $\Lambda < 0$ *and the parameters* (M, a, Λ) *satisfy the Hawking–Reall bound then* ψ *decays logarithmically in time* [120].

There are three main geometric phenomena associated with black holes that are directly relevant for understanding the global behavior of hyperbolic equations on black hole backgrounds and which play a crucial role in the proof of Theorem 6. These are (see the discussion in Section 1.4.4):

(1) the red-shift effect;

(2) the presence of trapped null geodesics;

(3) superradiance.

While the above phenomena are present for all black hole geometries discussed in Section 1.4.4, their strength, coupling, and the large scale geometry of the underlying spacetime lead to the quite different dynamical behaviors exhibited by Theorem 6. We provide a short discussion of these phenomena and how they enter the proof of Theorem 6.

2.3.1. $\Lambda = 0$. The classical vector field approach

In the simplest case $\Lambda = 0$, $a = 0$, the phenomenon of superradiance is absent and the problem can be entirely understood in physical space. As mentioned at the end of Section 1.4.2, the trapped geodesics are all concentrated at $r = 3M$, and one may prove (using appropriate multipliers) the following two estimates [26,27,63]:

$$\mathbb{E}[\psi](\tau) \le \mathbb{E}[\psi](0) \quad \text{for all } \tau \ge 0 \qquad \text{(boundedness)}, \tag{2.2}$$

$$\mathbb{I}_{\mathrm{deg}}[\psi](\tau_1,\tau_2) \le \mathbb{E}[\psi](0) \quad \text{for all } \tau_2 \ge \tau_1 \ge 0 \qquad \text{(local integrated energy decay)}, \tag{2.3}$$

where

$$\mathbb{E}[\psi](\tau) = \int_{\Sigma_{t^\star}} (\partial_{t^\star}\psi)^2 + \left(1 - \frac{2M}{r}\right)(\partial_r\psi)^2 + |\nabla\!\!\!/\,\psi|^2$$

$$\mathbb{I}_{\mathrm{deg}}[\psi](\tau_1,\tau_2) = \int_{\tau_1}^{\tau_2} d\tau \int_{\Sigma_\tau} \frac{1}{r^3}\Bigg[(R^\star\psi)^2 + \left(1 - \frac{3M}{r}\right)^2 \times \left((\partial_{t^\star}\psi)^2 + \left(1 - \frac{2M}{r}\right)(\partial_r\psi)^2 + |\nabla\!\!\!/\,\psi|^2\right)\Bigg].$$

These energies are defined in $(t^\star, r, \theta, \phi)$ coordinates in which the Schwarzschild metric takes the regular form $g = -(1 - \frac{2M}{r})(\mathrm{d}t^\star)^2 + \frac{4M}{r}\mathrm{d}t^\star \mathrm{d}r + (1 + \frac{2M}{r})\mathrm{d}r^2 + r^2 \mathring{\not{g}}$. Furthermore, Σ_τ denotes a slice (which intersects the horizon) of constant τ, and $R^\star = \frac{2M}{r}\partial_t + (1 - \frac{2M}{r})\partial_r$.

The degeneration at $r = 3M$ in the estimate (2.3) is necessary (although it can be weakened to logarithmic degeneration using microlocal techniques) and a manifestation of the trapped null geodesics. The red-shift effect allows one to eliminate the degeneration at $r = 2M$ for the transversal (∂_r in these coordinates) derivatives in (2.2) and (2.3) and can be realized as a physical space multiplier. From the resulting nondegenerate version of the estimates (2.2) and (2.3) one can prove, using a very general physical space method that merely uses the asymptotically flatness of the spacetimes (introduced in [65], see also [168]) inverse polynomial decay rates for the solutions which are in particular sufficiently strong for nonlinear applications.

For $\Lambda = 0$, $|a| \ll M$, superradiance is present and the Killing field ∂_t will not produce a coercive energy on spacelike slices. The naive estimate (2.2) fails as energy associated with the vector field ∂_t on a later slice can be larger than that of the initial slice. Moreover, trapped null geodesics now exist on a set of full measure (near $r = 3M$) in spacetime. The estimate (2.3) fails and it is not clear how to prove the required analogue. One approach—which was also the one that was later generalized to the full subextremal case—was to use the separability of equation (2.1) on Kerr and to exploit the fact that when one looks at pieces of the solution supported on certain (angular and time) frequencies, then good uniform estimates can be proven from the ordinary differential equations governing the behavior of the frequency localized components. It is a *tour de force* to construct these frequency localized multipliers which typically exploit a smallness parameter arising from the definition of the frequency regimes. A key insight is that superradiance can be controlled by the red-shift effect. Summing the estimates and the fact that the solution is a priori not L^2 in time provide further technical challenges. See [64]. Decay in the case $|a| \ll M$ was also proved in [198] by means of pseudodifferential multipliers near the trapped set and in [6] by exploiting the second order Carter symmetry operator (related to a hidden symmetry of the spacetime not related to Killing vector fields).

Two additional insights lead to a treatment of the full subextremal case $\Lambda = 0$, $|a| < M$ in [66]. The first was that, in the frequency decomposition of the solution outlined above, frequency triples that are affected by superradiance are nontrapped. Thus these two obstacles for decay happen to be disjoint when viewed in frequency space (this breaks down precisely in the extremal case $|a| = M$). The second was a quantitative version of mode stability for the wave equation established in [193] which allowed one to treat the range of bounded frequencies (where roughly speaking no smallness factor is available). This also allowed one to estimate precisely the amount of amplification of the solution through the mechanism of superradiance, see also [67].

2.3.2. $\Lambda = 0$: the extremal case

We have

Theorem 7 ([15]). *Consider a solution to the scalar wave equation*

$$\Box_{g_{M,a=M}} \psi = 0 \tag{2.4}$$

on the black hole exterior of an extremal Kerr spacetime. Then axisymmetric ψ decay inverse polynomially in time. However, along the event horizon $\mathcal{H}^+$ higher transversal derivatives of ψ generically grow in time (Aretakis instability).

Theorem 7 was first proved for the spherically symmetric extremal Reissner–Nordström metric in [13, 14] (without the restriction on axisymmetric solutions). The main difficulty is that the aforementioned red-shift effect degenerates and one cannot remove the degeneration at the horizon in the estimates. In fact, there are conservation laws on the event horizon $\mathcal{H}^+$ (discovered by Aretakis) which constitute direct obstructions to decay.

In the case of extremal Kerr, the problems of degenerate red-shift, trapping and superradiance are now fully coupled and cannot be studied separately even at the frequency decomposed level. This is the reason why the global behavior of solutions is only understood for axisymmetric solutions (which are not subject to superradiance). The general case is an open problem that has received a lot of attention from both theoretical physics (see, for instance, [37]) and mathematics recently and is expected to exhibit additional instabilities. See also [1, GAJIC] for recent work in this direction.

2.3.3. $\Lambda = 0$: sharp asymptotics

It is a natural question to ask about the precise decay rates in Theorem 6(2). This problem has a long tradition in the physics literature going back to work of Price [175, 176], with refinements given in [99]. While this question is interesting in its own right, lower bounds on the decay rate directly inform the behavior of solutions in the black hole interior (see Section 3.1 below). The following result is the current state-of-the-art.

Theorem 8 ([11, 106]). *Consider a solution to the scalar wave equation*

$$\Box_{g_{M,a}} \psi = 0 \tag{2.5}$$

on the black hole exterior of a subextremal ($|a| < M$) Kerr spacetime. Then the following uniform pointwise estimates hold for some $\eta \in (0,1)$:

$$\left|\psi - \frac{Q_0(\tau + r)}{\tau^2(\tau+2r)^2}\right| \leq \frac{E_0}{(\tau+2r)\tau^{2+\eta}}, \tag{2.6}$$

$$\left|r^{-\ell}\psi_{\ell=1} - \frac{Q_\ell(r,\theta,\phi)(\tau + r)}{\tau^3(\tau+2r)^3}\right| \leq \frac{E_0}{(\tau+2r)^2\tau^{3+\eta}}, \tag{2.7}$$

$$\left|r^{-\ell}\psi_{\geq\ell} - \frac{Q_\ell(r,\theta,\phi)(\tau + r)}{\tau^{2+\ell}(\tau+2r)^{2+\ell}}\right| \leq \frac{E_0}{(\tau+2r)^{1+\ell}\tau^{2+\ell+\eta}} \quad \textit{when } a = 0. \tag{2.8}$$

Here τ is a coordinate corresponding to a hyperboloidal slicing of the exterior with the slices ending at the horizon and future null infinity, and Q_ℓ is a bounded function in r

tending as $r \to \infty$ to an explicitly computable initial data quantity related to the Newman–Penrose charges. Finally, $\psi_{\geq \ell}$ denotes the projection of ψ to the standard spherical harmonics defined with respect to Boyer–Lindquist (θ, ϕ) coordinates and E_0 is a constant determined from a weighted initial data Sobolev norm.

The asymptotics (2.8) were first proved for $\ell = 0$, on a class of asymptotically flat spherically symmetric spacetimes including Schwarzschild and Reissner–Nordström, by Angelopoulos–Aretakis–Gajic [8]. Their subsequent work [9] extracts also a logarithmic subleading term in the large τ expansion of the radiation field (defined as the limit $\lim_{r\to\infty} r\psi(\tau, r, \theta, \phi)$) for spherically symmetric waves. On a class of asymptotically flat spacetimes which include subextremal Kerr spacetimes, Hintz [106] gave the first proof of (2.6). The proof is based on a careful spectral analysis near zero energy, see Section 2.3.4, with direct antecedents in the work of Donninger–Schlag–Soffer [71] and Tataru [197] which proved upper bounds of $|\psi|$ consistent with (but for $\ell \geq 1$ weaker than) the asymptotics stated above. The paper [106] also proves the estimate (2.8) in spatially compact sets and identifies $r^\ell Q_\ell(r, \theta, \phi)$ as a *generalized zero mode* of the wave equation, namely the unique stationary solution of $\Box_{g_{\Lambda=0,M,a=0}}(r^\ell Q_\ell) = 0$ with the property that $r^\ell Q_\ell(r, \theta, \phi) = q_\ell(r)Y$, $q_\ell = r^\ell + \mathcal{O}(r^{\ell-1+\varepsilon})$ as $r \to \infty$, where Y is a suitable degree ℓ spherical harmonic (depending on the initial data).

Angelopoulos–Aretakis–Gajic [11, 12] gave a physical space proof of Theorem 8. This interpolates a refinement (introducing carefully constructed higher order commutators adapted to the angular modes) of the r^p-method [65], which gets one close to the optimal rates, with a clever way to exploit the conservation of the Newman–Penrose charges along null infinity. In fact, the Newman–Penrose charges in Theorem 8 are not the ones associated with ψ itself but that of a "time-inverted" ψ and generically nonvanishing, even for data of compact support. Estimates analogous to (2.7) have been derived for higher modes but take a more complicated form, which we do not present here. We merely remark that obtaining the rates for higher modes in the case $a \neq 0$ is very delicate due to the coupling of angular modes (Kerr being only axisymmetric).

In the extremal spherically symmetric case (Section 2.3.2), the asymptotics are quite different [10]. In fact, the extremality of the horizon can be seen in the expansion on null infinity giving rise to speculations about the experimental detection of extremal black holes in the universe [7].

2.3.4. $\Lambda = 0$: spectral theoretic approach

Another approach to the proof of Theorem 6(2) is based entirely on spectral theory and phase space analysis. The starting point is a foliation of the spacetime by level sets of a time function t_* which are transversal to the future event horizon and asymptote to t-level sets as $r \to \infty$. (In practice, it is more convenient to work instead with t_* whose level sets are transversal to future null infinity.) Since $|\psi| < Ce^{Ct_*}$ for some $C > 0$, one can write ψ

as the inverse Fourier transform[4]

$$\psi(t_*) = \int_{\Im\sigma=C+1} e^{-i\sigma t_*}\widehat{\Box}(\sigma)^{-1}\hat{f}(\sigma)\,\mathrm{d}\sigma, \quad \Box = \Box_{g_{M,a}}, \tag{2.9}$$

where $\hat{f}(\sigma)$ is an explicit expression involving the Cauchy data of ψ, and the *spectral family* $\widehat{\Box}(\sigma)$ is obtained from $\Box$ by replacing ∂_{t_*} by $-i\sigma$. The inverse $\widehat{\Box}(\sigma)^{-1}$ in (2.9) is the *outgoing resolvent*, with range comprised of functions which decay as $r \to \infty$. The strategy is now to shift the contour of integration down to the real axis $\Im\sigma = 0$. Executing this relies on several ingredients.

The first ingredient is the analyticity of the resolvent $\widehat{\Box}(\sigma)^{-1}$ in $\Im\sigma > 0$ as well as the existence the limiting resolvent as $\Im\sigma \searrow 0$. This is established in two steps. The first step is that for $\Im\sigma \geq 0$, one can realize $\widehat{\Box}(\sigma)$ as a Fredholm operator between suitable function spaces (based on weighted L^2-Sobolev spaces), with locally uniform estimates; we only discuss this in the case $\sigma \in \mathbb{R}$. The operator $\widehat{\Box}(\sigma)$ satisfies elliptic estimates except in the region where $\partial_{t_*} = \partial_t$ is not timelike, which happens precisely in the ergoregion and the black hole interior. But microlocally, i.e., in phase space, the flow of the Hamiltonian vector field associated to the principal symbol of $\widehat{\Box}(\sigma)$—which here means the null-geodesic flow lifted to phase space restricted to the annihilator of ∂_{t_*}, and projecting out the t_*-coordinate—has useful structure: there is a source at $N^*\{r = r_+\} \setminus o$ (the conormal bundle of the event horizon); this is related to the classical red-shift effect. There, one gets free microlocal estimates for u solving

$$\widehat{\Box}(\sigma)u = f \tag{2.10}$$

in terms of f, called *radial point estimates* **[201, §2.4]**. These take the form

$$\|Au\|_{H^s} \leq C\left(\left\|G\widehat{\Box}(\sigma)u\right\|_{H^{s-1}} + \|\chi u\|_{H^{-N}}\right)$$

where $A, G \in \Psi^0$ are pseudodifferential operators localizing to suitable conic neighborhoods of $N^*\{r = r_+\}$, and χ localizes near $r = r_+$ in the base $\mathcal{M}$. The classical Duistermaat–Hörmander theorem on the propagation of regularity **[72]** allows one to propagate this control on u along the null-geodesic flow, which in the case $a \neq 0$ enters the black hole exterior, but which in any case ultimately enters the black hole interior.

Another source of nonellipticity of $\widehat{\Box}(\sigma)$ for real $\sigma \neq 0$ is due to the presence of an asymptotically flat end of the spatial slice $t_*^{-1}(0)$, and concerns the lack of arbitrary *decay* rather than regularity; indeed one has to allow for u in (2.10) to have outgoing asymptotics $u \sim r^{-1}e^{i\sigma r}$. This can be captured microlocally in Melrose's scattering calculus **[162]** and indeed historically was the first instance of a microlocal radial point estimate.

Altogether, one obtains locally uniform estimates in the punctured upper half-plane

$$\|u\|_{H^{s,\ell}} \leq C\left(\left\|\widehat{\Box}(\sigma)u\right\|_{H^{s-1,\ell+1}} + \|u\|_{H^{-N,-N}}\right), \quad \Im\sigma \geq 0,\ \sigma \neq 0, \tag{2.11}$$

[4] The choice of sign of σ in this formula (and thus also in the corresponding formula for the Fourier transform) is conventional.

where $H^{s,\ell} = r^{-\ell} H^s$ is a weighted Sobolev space. Analogous estimates on dual function spaces for the adjoint $\widehat{\Box}(\sigma)^*$ give the claimed Fredholm property of $\widehat{\Box}(\sigma)$. The invertibility of $\widehat{\Box}(\sigma)$, together with sharp mapping properties of the inverse, follows from the triviality of the kernel. This is the place, finally, where the mode stability results [193, 209] enter the analysis. Direct differentiation in σ then gives high regularity of $\widehat{\Box}(\sigma)^{-1}$ in $\sigma \neq 0$.

Uniform analysis near $\sigma = 0$ is delicate due to the degeneration of $\widehat{\Box}(\sigma)$ at spatial infinity when $\sigma \searrow 0$. Sharp Fredholm estimates were obtained by Vasy [202, 204] using a second microlocal combination of the scattering calculus (for $\sigma \neq 0$) and the b-calculus (for $\sigma = 0$) [161], following direct resolvent estimates [29] and (in a restricted geometric setting) direct constructions of the resolvent kernel [95–97, 101]. While bounds and mild (conormal) regularity of the resolvent near zero energy are sufficient to obtain some decay, Hintz [106] developed a method to obtain the first few terms of a polyhomogeneous (generalised Taylor) expansion of $\widehat{\Box}(\sigma)^{-1} \hat{f}(\sigma)$ at $\sigma = 0$. This uses resolvent identities and, in turns, the inversion of $\widehat{\Box}(0)$ and a rescaled model problem[5] capturing the transition from zero to non-zero spectral parameters. The expansion, upon restriction to bounded spatial subsets, takes the schematic form

$$\widehat{\Box}(\sigma)^{-1} \hat{f}(\sigma) = [\text{holomorphic}] + \sigma^2 \log(\sigma + i0)c + [\text{more regular error terms}] \tag{2.12}$$

for some constant $c \in \mathbb{C}$. Upon taking the inverse Fourier transform, the strongest singular term will give rise to the leading order long time asymptotics as $t_* \to \infty$, given by $2ct_*^{-3}$; the regularity of the error terms determines the decay rate of the remainder.

The second ingredient required to execute the contour shifting in the integral (2.9) concerns *high energy estimates*, i.e., quantitative bounds on $\widehat{\Box}(\sigma)^{-1}$ as $\Re\sigma \to \infty$ (locally uniformly in $\Im\sigma \geq 0$). It is in this high frequency regime that the structure of the trapping becomes relevant. To explain this in rough terms, consider the *semiclassically rescaled equation*

$$P_{h,z} u := h^2 \widehat{\Box}(h^{-1} z) u = f, \quad h = |\sigma|^{-1}, \quad z = \frac{\sigma}{|\sigma|} = 1 + \mathcal{O}(h). \tag{2.13}$$

As a guiding example, consider briefly the Minkowskian wave operator $\Box = -D_{t_*}^2 + \sum D_{x^j}^2$; then $P_{h,z} = \sum (hD_{x^j})^2 - 1 + \mathcal{O}(h)$, and thus according to geometric optics, high frequency ($\sim h^{-1}$) oscillations have momenta in $\{\sum \xi_j^2 - 1 = 0\}$ and propagate along lifted geodesics, which are the projections to spatial coordinates and momenta of Minkowskian null-geodesics. Generalizing to the Kerr case, high frequency oscillations of u solving (2.13) are localized in the *characteristic set* of spatial momenta ξ so that $-\mathrm{d}t_* + \xi$ is lightlike, and propagate along projections (to the spatial phase space) of lifted null-geodesics. The dynamics of this projected lifted null-geodesic flow is more ornate than in the bounded frequency regime: there is now a trapped set (in the Schwarzschild case: the restriction of Γ in (1.15) to $t = 0$ and $\sigma = -1$ if we take $t_* = t$ near $r = 3M$) at which the flow is ν-normally hyperbolic for all ν. However, *purely based on the dynamical nature of the trapping* and its interplay with the symplectic structure of phase space, one can apply black box results [77, 210] (see

[5] It is obtained by taking $\hat{r} = \sigma r$ and considering the limit $\sigma \to 0$ with fixed $\hat{r}$.

also [108]) on the propagation of semiclassical regularity (i.e., bounds on amplitudes of high frequency oscillations) for microlocal control of u there. Combining this with semiclassical radial point estimates at the event horizon [201, §2.8] [78, APPENDIX E] and at spatial infinity [203, 205], one ultimately obtains estimates in semiclassical function spaces (with each derivative weighted by a factor of h) analogous to (2.11),

$$\|u\|_{H_h^{s,\ell}} \leq C\big(h^{-1-\varepsilon}\|P_{h,z}u\|_{H_h^{s-1,\ell+1}} + h^N\|u\|_{H_h^{-N,-N}}\big),$$

where the $\varepsilon > 0$ loss can be sharpened to a logarithmic loss when $\Im z \geq 0$; this loss comes from the trapping estimate. For small $h > 0$, the second, error, term on the right can be absorbed into the left-hand side, and one obtains the invertibility (with quantitative bounds) of $P_{h,z}$ and thus of $\widehat{\Box}(\sigma)$.

Equipped with these high energy estimates, one can justify the contour shifting down to the real axis; the loss of powers of $h^{-1} = |\sigma|$ corresponds to a necessary loss [186] of regularity of the solution (when estimated in decaying function spaces) relative to the initial data.

Important precursors of the low frequency analysis of [106, 204] are the works by Donninger–Schlag–Soffer [71] (based on direct resolvent kernel constructions in spherical symmetry) and Tataru [197] on Price's law (based on resolvent estimates and weak versions of the expansion (2.12)). In Tataru's approach, uniform resolvent control down to the real axis is deduced from the assumption of a suitable form of *local energy decay*; this assumption needs to be verified separately. (Thus, [197] upgrades weak to sharp decay.) In the full subextremal range, local energy decay was first proved in the aforementioned [66], following earlier work for slow angular momenta [6, 64]. For more on the relationship between mode stability and local energy decay, see [165, 167].

2.3.5. $\Lambda > 0$: exponential decay and quasinormal mode expansions

For linear scalar waves on slowly rotating Kerr–de Sitter black hole spacetimes $(M, g_{\Lambda,M,a})$, Dyatlov [73–75] proved a full asymptotic expansion

$$\psi(t_*, x) = \sum_{\Im\sigma_j \geq -\alpha} e^{-i\sigma_j t_*} t_*^k a_{jk}(x) + \mathcal{O}(e^{-\alpha t_*}), \quad x = (r, \theta, \phi), \tag{2.14}$$

for all $\alpha \in \mathbb{R}$ avoiding the discrete set of accumulation points of $\{-\Im\sigma_j\}$. Here, the σ_j are the *resonances* or *quasinormal modes* (QNMs);[6] they are the poles of the *meromorphic* continuation of the resolvent $\widehat{\Box_{g_{\Lambda,M,a}}}(\sigma)^{-1}$ from $\Im\sigma \gg 1$ to the complex plane in σ. The structure of the set $\{\sigma_j\}$ was analyzed in great detail in [75]; here we only record that the only resonance with $\Im\sigma_j \geq 0$ is $\sigma_0 := 0$ (with multiplicity 1 and a_{00} a constant), and thus ψ decays exponentially fast to a constant. The existence of a meromorphic continuation of the resolvent (as opposed to the nonalgebraic singularity (2.12) in the case $\Lambda = 0$) is due to the presence of the cosmological horizon and the related fact that one work with a *compact* spatial manifold;

[6] In the absence of multiplicities ($k = 0$), a_{j0} is a corresponding mode solution (or *resonant state*).

the general microlocal framework for the relevant spectral theory (both for bounded and high frequencies) was provided in Vasy's seminal work [201]. (The analyticity of the quasinormal mode solutions, for analytic choices of time functions t_*, was proved in [148], based on [89].)

In the Schwarzschild–de Sitter case, (2.14) was proved by Bony–Häfner [28] in the black hole exterior, with uniformity down to the horizons provided in [164] using the relationship of SdS and asymptotically hyperbolic spaces [160, 163]. For results on the set of resonances and mode solutions in the small black hole limit $M\Lambda^2 \searrow 0$, see [117]. Results on quasinormal modes on charged black hole spacetimes were proved in [22,127]. We remark that energy methods [62] have thus far been successful in proving superpolynomial decay to constants.

2.3.6. $\Lambda < 0$: stable trapping and logarithmic decay

The existence of the conformal boundary (where null geodesics are reflected) in conjunction with the existence of trapped geodesics leads to the phenomenon of *stable* trapping, which gives rise to an inverse logarithmic rate for solutions provided the Hawking–Reall bound is satisfied. (The Hawking–Reall bound ensures the existence of a globally causal Killing field on the exterior and hence eliminates the difficulty of superradiance.) This rate was established as an upper bound in [120] and is in fact optimal for general solutions, as follows either from quasimode constructions [121] or the existence of quasinormal modes exponentially close to the real axis [90, 91]; see [208] for the development of a general theory of quasinormal modes in this setting. (Inverse logarithmic decay rates are familiar from the obstacle problem in Minkowski space [33].) Outside the Hawking–Reall bound, Dold [70] constructed exponentially growing solutions using techniques from [192].

Finally, the behavior is expected to be radically different if dissipative boundary conditions are imposed on the scalar field, in which case strong decay is likely to hold.

2.4. Linear stability

Here, we only discuss the case $\Lambda = 0$. The reason is that there are currently no rigorous results for $\Lambda < 0$ in the case of reflecting boundary conditions, whereas for $\Lambda > 0$ nonlinear stability was proved directly (see Section 2.5) without prior work on linear stability.

Theorem 9 ([5, 100]). *Linear stability holds for slowly rotating Kerr spacetimes.*

A natural approach to Theorem 9 is to try to reduce it to the toy problem. However, both the generalized harmonic gauge (Section 1.2) as well as the double null gauge (Section 1.3) lead to a highly coupled *system* of linearized equations. In the following, we describe several different approaches to address this problem.

2.4.1. The double null approach

The first linear stability result was proved by Dafermos–Holzegel–Rodnianski in [58] and concerns the linear stability of the Schwarzschild metric. The approach of [58] is based on expressing (1.1) in a double null gauge and linearizing the resulting system with respect to Schwarzschild. For clarity, we describe the main ideas more generally for the linearization around Kerr. Let us fix the differentiable structure of the Pretorius–Israel double null coordinates (u, v, θ) on the Kerr manifold with parameters (M, a) and consider a one-parameter family of metrics expressed in these coordinates

$$\boldsymbol{g}(\varepsilon) = -4\boldsymbol{\Omega}^2(\varepsilon)\mathrm{d}u\,\mathrm{d}v + \boldsymbol{\not g}_{AB}(\varepsilon)\big(\mathrm{d}\theta^A - \boldsymbol{b}^A(\varepsilon)\mathrm{d}v\big)\big(\mathrm{d}\theta^B - \boldsymbol{b}^B(\varepsilon)\mathrm{d}v\big) \tag{2.15}$$

such that $\varepsilon = 0$ corresponds to the Kerr metric of mass M and specific angular momentum a. In other words, we identify the ingoing and outgoing null cones of each member of the family with the respective cones of the Kerr exterior.

If $\boldsymbol{\xi}$ is an $\mathbb{S}^2_{u,v}$-tensor denoting an arbitrary Ricci coefficient or curvature component associated with $\boldsymbol{g}(\varepsilon)$ and ξ denotes the analogous component for $\boldsymbol{g}(0)$, then $\boldsymbol{\xi} - \xi$ is a map from $\mathbb{R}$ into the bundle of $\mathbb{S}^2_{u,v}$-tensors on $\mathcal{M}$ and we can hence define

$$\overset{(1)}{\xi} := \frac{\mathrm{d}}{\mathrm{d}\varepsilon}(\boldsymbol{\xi} - \xi)|_{\varepsilon=0},$$

which we call a linearized Ricci coefficient or curvature component, respectively. Note that we can indeed consider the difference of the two tensors as we have identified the notion of $\mathbb{S}^2_{u,v}$-tensors for the family (2.15), i.e., we have fixed the tensor bundle of $\mathbb{S}^2_{u,v}$-tensors on the manifold independently of ε.

To produce the linearized Bianchi and null structure equations, one writes down the null structure and Bianchi equations once for general ε (in bold font) and once for $\varepsilon = 0$ (in standard font) and then subtracts the two equations ignoring terms of order ε^2. This linearization process is entirely straightforward and particularly simple in the Schwarzschild case where all $\mathbb{S}^2_{u,v}$-one-forms and symmetric traceless tensors vanish identically for the (spherically symmetric) background. It produces a system of equations for $\mathbb{S}^2_{u,v}$-tensors representing linearized curvature components and Ricci coefficients on the Kerr manifold with all differential operators being defined with respect to the Kerr background metric. For instance, in the (algebraically simpler) Schwarzschild case $a = 0$, where the Pretorius–Israel double null coordinates become the familiar Eddington–Finkelstein double null coordinates, the linearization of (1.10), (1.11) reads

$$\Omega\not\nabla_3\big(r\Omega^2\overset{(1)}{\alpha}\big) = -2\Omega^2 r\not{\mathcal{D}}_2^\star\big(\Omega\overset{(1)}{\beta}\big) - \frac{3M}{r^2}\Omega^3\overset{(1)}{\hat{\chi}}, \tag{2.16}$$

$$\Omega\not\nabla_4\big(r^4\Omega^{-1}\overset{(1)}{\beta}\big) = -r^3\not{\mathrm{d}}\mathrm{iv}\big(r\overset{(1)}{\alpha}\big), \tag{2.17}$$

and the structure equation (1.8) becomes

$$\Omega\not\nabla_4\big(r^2\overset{(1)}{\hat{\chi}}\Omega\big) + \frac{2M}{r^2}\big(r^2\overset{(1)}{\hat{\chi}}\Omega\big) = -r^2\Omega^2\overset{(1)}{\alpha}. \tag{2.18}$$

We now collect an important result, which is due to Teukolsky [200]. For this we recall from Section 1.4.4 the algebraically special frame of Kerr $(e_3^{as}, e_4^{as}, e_1^{as}, e_2^{as})$. We define an ε-dependent family of frames $(\boldsymbol{e}_3^{as}, \boldsymbol{e}_4^{as}, \boldsymbol{e}_1^{as}, \boldsymbol{e}_2^{as})$ which is null with respect to $\boldsymbol{g}(\varepsilon)$ and reduces for $\varepsilon = 0$ to $(e_3^{as}, e_4^{as}, e_1^{as}, e_2^{as})$. We then define the linearized quantities

$$\overset{(1)}{\alpha}_{as}(e_A^{as}, e_B^{as}) := \lim_{\varepsilon\to 0}\frac{\mathbf{Riem}(\boldsymbol{e}_4^{as}, \boldsymbol{e}_A^{as}, \boldsymbol{e}_4^{as}, \boldsymbol{e}_B^{as})}{\varepsilon}, \tag{2.19}$$

$$\overset{(1)}{\underline{\alpha}}_{as}(e^{as}_A, e^{as}_B) := \lim_{\varepsilon\to 0} \frac{\mathbf{Riem}(\boldsymbol{e}^{as}_3, \boldsymbol{e}^{as}_A, \boldsymbol{e}^{as}_3, \boldsymbol{e}^{as}_B)}{\varepsilon}. \tag{2.20}$$

Note that these quantities are generally *not* $\mathbb{S}^2_{u,v}$-tensors unless $a = 0$ but can be interpreted as horizontal tensors by identifying the horizontal structures (i.e., the spaces $g(\varepsilon)$-orthogonal to the distribution $(\boldsymbol{e}^{as}_3, \boldsymbol{e}^{as}_4)$) for different ε just as we did for $\mathbb{S}^2_{u,v}$-tensors earlier. They can also be viewed as elements of a complex line bundle of spin-± 2-weighted functions, see **[200]** and **[57, §2.2]**. Note that for $a = 0$ we have $\overset{(1)}{\alpha}_{as}(e^{as}_A, e^{as}_B) = \overset{(1)}{\alpha}(e_A, e_B)$ and $\overset{(1)}{\underline{\alpha}}_{as}(e^{as}_A, e^{as}_B) = \overset{(1)}{\underline{\alpha}}(e_A, e_B)$ as the double null frame agrees with the algebraically special frame. Also one may check that (2.19) and (2.20) do not depend on the particular choice of frame $(\boldsymbol{e}^{as}_3, \boldsymbol{e}^{as}_4, \boldsymbol{e}^{as}_1, \boldsymbol{e}^{as}_2)$ described above. In other words, there is a gauge invariance to order ε under $\boldsymbol{g}(\varepsilon)$-frame rotations.

Proposition (**[200]**). The quantities $\overset{(1)}{\alpha}_{as}$ and $\overset{(1)}{\underline{\alpha}}_{as}$ satisfy (individually) decoupled wave equations, called the spin ± 2 Teukolsky equations.

In the case $a = 0$, the Teukolsky equation takes the simple form (as easily checked from (2.16)–(2.18))

$$\begin{aligned} &\Omega\nabla\!\!\!/_4\Omega\nabla\!\!\!/_3\big(r\Omega^2\overset{(1)}{\alpha}\big) + \frac{2\Omega^2}{r^2} r^2 \mathcal{D}\!\!\!/^{\star}_2 \mathrm{d}\!\!\!/\mathrm{iv}\big(r\Omega^2\overset{(1)}{\alpha}\big) + \frac{4}{r}\left(1 - \frac{3M}{r}\right)\Omega\nabla\!\!\!/_3\big(r\Omega^2\overset{(1)}{\alpha}\big) \\ &\quad + \frac{6M\Omega^2}{r^3}\big(r\Omega^2\overset{(1)}{\alpha}\big) = 0, \end{aligned} \tag{2.21}$$

which we will focus on to convey some of the main ideas that follow. The problem with equation (2.21) is that because of the first order term, the standard physical space techniques for the toy problem do not apply: there is no natural conserved energy and the standard approach to prove (2.3) fails. Nevertheless, we have the following result:

Theorem 10. **[57,58,157]** *Solutions to the spin ± 2 Teukolsky equations arising from suitably weighted initial data on a Kerr spacetime with $|a| \ll M$ decay inverse polynomially in time on the black hole exterior.*

Proof. For $a = 0$, one may apply the physical space transformations

$$r^5\overset{(1)}{P} = \frac{r^3}{\Omega}\nabla\!\!\!/_3\big(\overset{(1)}{\psi} r^3\Omega\big), \quad r^3\Omega\overset{(1)}{\psi} = -\frac{1}{2}\frac{r^2}{\Omega}\nabla\!\!\!/_3\big(r\Omega^2\overset{(1)}{\alpha}\big) \tag{2.22}$$

introduced in **[58]**. These transformations are physical space versions of transformation introduced by Chandrasekhar in **[38]** at the mode-decomposed level. The point is that the quantity $\overset{(1)}{P}$ satisfies the *Regge–Wheeler* equation

$$\Omega\nabla\!\!\!/_3\Omega\nabla\!\!\!/_4\big(r^5\overset{(1)}{P}\big) + \frac{2\Omega^2}{r^2} r^2 \mathcal{D}\!\!\!/^{\star}_2 \mathrm{d}\!\!\!/\mathrm{iv}\big(r^5\overset{(1)}{P}\big) + V\big(r^5\overset{(1)}{P}\big) = 0, \tag{2.23}$$

where V is a potential with favorable properties. Equation (2.23) turns out to be an equation for which the estimates (2.2) and (2.3) can be proven, i.e., the toy model theory applies. Once (2.3) is proven for $\overset{(1)}{P}$ one may derive from (2.22) the identity

$$\frac{1}{2}\Omega\nabla\!\!\!/_3\big(r|r^3\Omega\overset{(1)}{\psi}|^2\big) + \frac{1}{2}\Omega^2|r^3\Omega\overset{(1)}{\psi}|^2 = r^4\overset{(1)}{P}\Omega^2 \cdot \overset{(1)}{\psi} r^3\Omega. \tag{2.24}$$

Applying Cauchy–Schwarz inequality on the right and using the integrated decay estimate for $\overset{(1)}{P}$, this can be integrated forwards to produce boundedness and integrated decay estimates for $\overset{(1)}{\psi}$. Repeating the same procedure for the pair $(\overset{(1)}{\psi}, \overset{(1)}{\alpha})$, one obtains the desired estimates for $\overset{(1)}{\alpha}$. Obviously, this process loses derivatives but these can be recovered by studying the wave equations for $\overset{(1)}{\psi}$ and $\overset{(1)}{\alpha}$ with the just obtained a priori estimates on the lower order terms.

For $|a| \ll M$, a straightforward modification of the transformations (2.22) produces an analogue of (2.23), which is now a coupled wave equation schematically of the form

$$\Box_{RW,g_{M,a}} \overset{(1)}{P} = a\mathcal{F}\big(\overset{(1)}{\psi}, \overset{(1)}{\alpha}\big) \tag{2.25}$$

where $\Box_{RW,g_{M,a}}$ is the Regge–Wheeler operator associated with the Kerr metric $g_{M,a}$ to which again the techniques from the toy problem apply and $\mathcal{F}(\overset{(1)}{\psi}, \overset{(1)}{\alpha})$ denotes an explicit expression involving up to first derivatives of $\overset{(1)}{\psi}$ and $\overset{(1)}{\alpha}$. However, just as for the toy problem, proving estimates for $\Box_{RW,g_{M,a}}$ now requires frequency decomposition and a form of separability of the equations, which makes the problem technically more involved. Moreover, the transport estimates for the lower order quantities are now directly coupled to (2.25) so all estimates have to be proven at the same time. Key are the smallness of $|a|$ in the coupling as well as a special structure in the right-hand side of (2.25), which needs to be identified and exploited. ■

In the full subextremal case $|a| < M$, we have the following recent milestone:

Theorem 11 ([194]). *Theorem 10 holds for the full subextremal range $|a| < M$ provided solutions are a priori assumed to be future-integrable.*

We remark that the assumption of future-integrability in Theorem 11 ensures that one can take the Fourier transform in time and hence prove estimates at the level of the radial ODE governing the dynamics of the frequency decomposed pieces of the solution. For the wave equation, proving these estimates (i.e., Theorem 11) is the main difficulty in the proof of Theorem 6. Removing the assumption of future integrability is expected to follow along the lines of [66] and would lead to a proof of Theorem 10 for the full subextremal range and complete our picture of the dynamics of the Teukolsky equation.

The proof of Theorem 11 adds several new ideas to the proof of Theorem 10. It requires a much more subtle construction of the multipliers and various applications of the Teukolsky–Starobinski identities since smallness of $|a|$ cannot be exploited. We finally note also the papers [80, 81] for related results on the Teukolsky equation.

We pause for a moment to recap what we have achieved in proving Theorem 9. We have obtained a linearized system of equations in double null gauge and we have shown that certain quantities within this system satisfy decay estimates. In the $a = 0$ case, these are precisely $\overset{(1)}{\alpha}$ and $\overset{(1)}{\underline{\alpha}}$. The second and equally important step is to identify a hierarchical structure in the linearized system that allows proving boundedness and decay for all dynamical quantities, preferably without loss of derivatives (the latter with the nonlinear problem in mind) from the quantities that have been shown to decay. This was achieved for $a = 0$ in [58]. This part of the proof relies on a complete understanding of two important classes of special solutions, which we again discuss in the $a = 0$ case:

(A) An explicit 4-dimensional family of solutions to the system arising from the fact that the Schwarzschild solution sits as a one-parameter family inside the Kerr family. Since to specify a nearby Kerr from the point of view of Schwarzschild metric one needs to prescribe both a direction of rotation (i.e., a 3-vector) and a change of mass (a scalar), the corresponding family is indeed 4-dimensional.

(B) An infinite-dimensional family of pure gauge solutions, parametrized by a set of spacetime functions $f_i(u, v, \theta)$. These arise from the fact that certain infinitesimal coordinate transformations preserve the double null form (2.15) to order ε^2 while changing the dynamical quantities to order ε in an explicit fashion. For instance, the coordinate transformation

$$\tilde{u} = u, \quad \tilde{v} = v + \varepsilon f_2(v, \theta), \quad \tilde{\theta}^A = \theta^A + \varepsilon \frac{2}{r(u,v)} \not{g}^{AB} \partial_A f_2(v, \theta)$$

is easily seen to preserve the double null form (2.15) to order ε. These transformations are the infinitesimal versions of a change of double null foliation, i.e., perturbing the spheres and the foliations of the cones slightly. At the linearized level they generate a special class of solutions which are called pure gauge solutions. These solutions may be added to a given reference solution of the linearized system to achieve a specific normalization of the linearized Ricci coefficients on suitable chosen cones of the background. Finally, one observes that pure gauge solutions always have $\overset{(1)}{\alpha} = 0 = \overset{(1)}{\underline{\alpha}}$, i.e., the Teukolsky quantities are gauge invariant.

With these observations, we can state the following slightly more specific version of Theorem 9 for $a = 0$.

Theorem 12 ([58]). *All solutions to the linearized vacuum Einstein equations around Schwarzschild arising from regular asymptotically flat initial data*

- *remain uniformly bounded on the exterior and*
- *decay inverse polynomially (through a suitable foliation) to a standard linearized Kerr solution (a 4-dimensional space) after adding a pure gauge solution which can itself be estimated by the size of the data.*

The point here is that in a gauge that is normalized with respect to the cones where the initial data is prescribed ("initial data gauge") one will only be able to prove *boundedness*. In order to see *decay*, one needs to add a pure gauge solution that achieves $\Omega^{-1}\overset{(1)}{\Omega} = 0$ on $\mathcal{H}^+$ for the sum of the two solutions, i.e., one needs to normalize the solution with respect to the future event horizon to see decay. This is the future gauge or teleological normalization. While the pure gauge solution required to do this has to be determined dynamically by solving an ODE along the event horizon, it can still be bounded uniformly by initial data.

The reason that going to the teleological normalization is required to prove decay can be understood from the fact that while the (linearized) Bianchi equations capture the hyperbolic nature of the Einstein equations, the (linearized) null structure equations also involve *transport* equations. Solutions to transport equations do generally not decay to zero if integrated from initial data. On the other hand, integrating them backwards from the future (in this case, the event horizon) with zero data captures the decay (provided the right-hand side of the relevant transport equation has been shown to decay sufficiently fast, which is, for instance, the case if it involves a Teukolsky quantity). Geometrically, one may say that in the initial data normalized gauge the solution converges to a linearized Kerr but not in the standard double null coordinates.

We note that while we have described parts of the argument for the case $a = 0$ (which is the one treated in [58]), the double null approach can be pursued also in the case $|a| \ll M$ by treating the errors arising from non-vanishing a in the transport equation perturbatively and using Theorem 10. This is a problem that is likely to be solved in the near future and would provide a complete proof of Theorem 9 in double null gauge. Finally, generalizing carefully the transformations (2.22), linear stability of the Reissner–Nordström solution to the coupled Einstein–Maxwell system has been proven in [92].

2.4.2. Generalized harmonic gauge

Häfner–Hintz–Vasy [100] proved the linear stability of slowly rotating Kerr black holes for initial data with standard decay bounds on the initial data (roughly pointwise $o(r^{-1})$ and $o(r^{-2})$). Their proof is based on a precise analysis of the resolvent of a linearized gauge-fixed Einstein operator. More precisely, when studying the linear stability of $g_{M,a}$, $|a| \ll M$, [100] uses the linearization of the generalized harmonic gauge 1-form

$$W_\mu = g_{\mu\nu} g^{\kappa\lambda}\big(\Gamma(g)^\nu_{\kappa\lambda} - \Gamma(g_{M,a})^\nu_{\kappa\lambda}\big)$$

around $g = g_{M,a}$, which maps $h \mapsto \operatorname{div}_g \mathsf{G}_g h$. One then considers the linearization L of the quasilinear wave operator

$$P(g) := \operatorname{Ric}(g) - (\delta_g^* + q)W$$

around $g = g_{M,a}$. Here q is a suitable stationary bundle map (i.e., differential operator of order 0) from 1-forms to symmetric 2-tensors, used to implement *constraint damping* below. This linearization maps

$$L : h \mapsto D_g \operatorname{Ric}(h) - (\delta_g^* + q) \operatorname{div}_g \mathsf{G}_g h.$$

The principal part of L is $-\frac{1}{2}\Box_g$, and thus the Fredholm theory for the spectral analysis for the scalar wave operator sketched in Section 2.3.4 is available for the analysis of $\hat{L}(\sigma)$, $\Im\sigma \geq 0$, as well. High energy estimates require the verification of a sign condition on the subprincipal symbol at trapped set (required for application of the black box high energy estimates [77]), which was first verified in [104], and which can be shown to amount to polynomial bounds for the length of vectors that are parallel transported along trapped null-geodesics (the latter was proved in the general Kerr case in [159]).

Existence of the resolvent $\hat{L}(\sigma)^{-1}$ for nonzero σ then reduces to the problem of proving mode stability for metric perturbations; moreover, the behavior near $\sigma = 0$ is complicated due to the presence of stationary solutions (linearized Kerr metrics in a suitable gauge). Furthermore, it is (for robustness under perturbations, and also for eventual nonlinear purposes) important to *not* require the metric perturbations h to satisfy the linearized constraint equations at the Cauchy hypersurface $t_* = 0$. The analysis of $\hat{L}(\sigma)h = 0$ (i.e., L acting on the metric perturbation $e^{-i\sigma t_*}h$, with h outgoing) then starts off with the (linearized) second Bianchi identity, which gives the decoupled equation

$$\operatorname{div}_g \mathsf{G}_g(\delta_g^* + q)\eta = 0, \quad \eta = \operatorname{div}_g \mathsf{G}_g(e^{-i\sigma t_*}h).$$

This is a wave equation for the 1-form η, and for $q = 0$ it is indeed the tensor wave operator on 1-forms. The latter satisfies mode stability by direct calculation similarly to [114], except for the presence of a 0-mode corresponding to the Coulomb solution. The purpose of the stationary map q is to perturb this stationary solution away, thus giving mode stability for $\operatorname{div}_g \mathsf{G}_g(\delta_g^* + q)$ in the full closed upper half plane.[7] Thus, any mode solution h of $\hat{L}(\sigma)$ with $\Im\sigma \geq 0$ *automatically* verifies the linearized gauge condition

$$\eta = \operatorname{div}_g \mathsf{G}_g(e^{-i\sigma t_*}h) = 0. \tag{2.26}$$

Therefore, we also have a solution of the linearized Einstein equation *without gauge condition*,

$$D_g\mathrm{Ric}(e^{-i\sigma t_*}h) = 0. \tag{2.27}$$

Mode stability for this equation is well-known in the Schwarzschild case $(M, a) = (M_0, 0)$ [141,166,177,206,211], and [100] proceeds perturbatively off this case. Concretely, for nonzero σ, metric mode stability is the statement that $e^{-i\sigma t_*}h = \delta_g^*(e^{-i\sigma t_*}\omega)$ is a symmetric gradient (i.e., a Lie derivative when using vectors instead of 1-forms). Plugging this into the gauge condition (2.26) gives another wave equation for the gauge potential ω itself,

$$\operatorname{div}_g \mathsf{G}_g\delta_g^*(e^{-i\sigma t_*}\omega) = 0. \tag{2.28}$$

Mode stability for this equation implies that $\omega = 0$ and hence $h = 0$. This proves mode stability for L in the punctured upper half-plane.

For stationary perturbations ($\sigma = 0$), the mode stability for (2.27) in the Schwarzschild case implies that h is the sum of a linearized Kerr metric and a pure gauge term (i.e., a symmetric gradient). The gauge condition (2.28) then further restricts the pure gauge term $\delta_g^*\omega$ to a finite-dimensional space. Particular instances of such pure gauge terms, constructed

[7] In fact, one can then show that solutions of $\operatorname{div}_g \mathsf{G}_g(\delta_g^* + q)\eta = 0$ with sufficiently smooth and decaying initial data decay in time to 0. Therefore, initial violations of the linearized gauge condition $\eta = 0$ decay in time; equivalently, initial violations of the linearized constraint equations for h decay in time. Hence, the addition of q implements *constraint damping*, the roots of which go back to the numerics literature [31,98].

in [100, §7], are symmetric gradients of asymptotic (as $r \to \infty$) translations $\omega = \mathrm{d}x^i + o(1)$ (note that $\mathrm{d}x^i$ is a Killing 1-form for the Minkowski metric) and asymptotic rotations. There are also generalized zero modes which are first-order polynomials in time, arising from gauge potentials ω which are asymptotic Lorentz boosts.

We stress that it is *only* at this point—i.e., where one needs to know the structure (pure gauge, or linearized Kerr) of mode solutions of the linearized Einstein vacuum equation for individual values of $\sigma \in \mathbb{C}$—that the highly delicate reductions of the linearized Einstein equations to scalar "master equations" are used (here the Regge–Wheeler [177] and Zerilli [211] equations). The a priori information, obtained by microlocal means which only rely on *qualitative* features of the null-geodesic flow, of uniform Fredholm properties of $\hat{L}(\sigma)$ acting between suitable function spaces is then easily upgraded to invertibility, regularity in σ, and the precise structure at $\sigma = 0$.

Altogether, one can then show that in the Schwarzschild case $\hat{L}(\sigma)^{-1}$ has a second-order pole at $\sigma = 0$ with explicit singular terms, plus a Hölder-regular remainder. This structure persists in the slowly rotating Kerr case, essentially since one can construct a space of generalized zero modes for small $|a|$ of the same dimension as for $a = 0$. (We remark that this construction requires, besides knowledge of the Kerr family of solutions, only soft perturbative arguments, and does not involve the Teukolsky equation.) Altogether, we have

Theorem 13 ([100]). *Let t_* be a time function with level sets transversal to the future event horizon and equal to the level sets of the Boyer–Lindquist time function t for large r. Consider a neighborhood $\mathcal{M} = [0,\infty)_{t_*} \times \Sigma$, $\Sigma = [2M_0 - \varepsilon, \infty) \times \mathbb{S}^2$, of the domain of outer communications of the mass M_0 Schwarzschild black hole restricted to the causal future of the Cauchy surface $t_* = 0$. Let $\alpha \in (0,1)$, and let $h_0, h_1 \in \mathcal{C}^\infty(\Sigma; S^2T^*_\Sigma\mathcal{M})$ be Cauchy data with*

$$|h_0| \lesssim r^{-1-\alpha}, \quad |h_1| \lesssim r^{-2-\alpha},$$

and similar bounds for 8 derivatives along $r\partial_r$ and spherical vector fields. For (M,a) close to $(M_0, 0)$, let h denote the solution of the initial value problem

$$L_{M,a}h = 0, \quad (h, \mathcal{L}_{\partial_{t_*}}h)|_{t_*=0} = (h_0, h_1).$$

Then there exist (M', a'), and a vector field V lying in a 7-dimensional space $\mathcal{V}_{M,a}$ of vector fields on $\mathcal{M}$ so that

$$h = \frac{\mathrm{d}}{\mathrm{d}s} g_{M+sM',a+sa'}|_{s=0} + \mathcal{L}_V g_{M,a} + \tilde{h},$$

where $|\tilde{h}| \lesssim t_*^{-1-\alpha}$ in spatially compact regions. The space $\mathcal{V}_{M,a}$ is spanned by asymptotic translations and boosts, and an additional explicit vector field. (The latter vector field can be eliminated by a small change of the gauge condition. Asymptotic rotations are the same as infinitesimal changes of the black hole rotation axis.)

If the data (h_0, h_1) arise from initial data for the linearized Einstein equations (i.e., satisfying the linearized constraint equations), this implies the linear stability of slowly rotating Kerr black holes. However, Theorem 13 is significantly more general, as it applies to *general* data (h_0, h_1); this type of generality was crucial in the non-linear stability proof of Kerr–de Sitter black holes, see Section 2.5.2.

Note that, unlike in the double null gauge, the linearized metric h typically grows linearly in time due to the existence of asymptotic Lorentz boosts in the space $\mathcal{V}_{M,a}$. Decay to a linearized Kerr metric can only be seen after subtracting a suitable element of the 7-dimensional space $\mathcal{L}_V g_{M,a}$, $V \in \mathcal{V}_{M,a}$, of pure gauge solutions.

There have also been a number of works which employ vector field methods to prove the linear stability of the Schwarzschild metric in generalised harmonic gauges (for initial data satisfying the linearized constraints), see [124–126,131].

2.4.3. Other approaches: outgoing radiation gauge

Andersson–Bäckdahl–Blue–Ma [5] gave the first proof of linear stability, assuming strong decay on the initial data. Their strategy is to assume suitable decay for solutions $\overset{(1)}{\alpha}_{as}$, $\overset{(1)}{\underline{\alpha}}_{as}$ (called $\psi_{\pm 2}$ in [5] in accord with classical Newman–Penrose notation) of the Teukolsky equations and recover the full metric perturbation via successive integrations in a suitable hierarchy. In order to accomplish this, [5] employs the *outgoing radiation gauge*.[8] Decay for the metric perturbation is proved via weighted Hardy inequalities. Special care has to be taken near null infinity, where the Teukolsky–Starobinsky identities (fourth-order differential identities relating ψ_{+2} and ψ_{-2}) play a key role for ensuring integrability at various stages in the hierarchy. In order to obtain decay for the metric coefficients, the initial data for the metric perturbation are required to have strong decay (roughly pointwise $o(r^{-7/2})$ and $o(r^{-9/2})$ decay for the linearized metric and second fundamental form). This in particular forces the linearized mass and angular momentum of the final linearized Kerr solution to *vanish* [2], and hence a key feature of the (nonlinear) stability problem is suppressed.

Given that decay results [57,157] for the Teukolsky equation on slowly rotating Kerr backgrounds are known (cf. Theorem 10 above), [5] gives an unconditional proof of the linear stability (for strongly decaying data) in this regime.

2.5. Nonlinear stability

The first nonlinear stability result for any family of black hole spacetimes without symmetry assumptions was proved for slowly rotating Kerr–de Sitter black holes ($\Lambda > 0$) by Hintz–Vasy [115]. The results in the asymptotically flat Kerr setting are not quite yet complete, though stability under special symmetries [139] as well as the full codimensional stability of the Schwarzschild family [59] are known. See also [137,138,140] for progress in the slowly rotating case. Finally, we refer the readers to [59, §IV.2] for a discussion of (necessarily codimension restricted) nonlinear stability statements that one could attempt to prove in the *extremal* case.

[8] The metric perturbation is trace-free (with respect to the background Kerr metric) and has vanishing contraction with the ingoing principal null vector field.

2.5.1. The case $\Lambda = 0$

With the linear problem resolved, the road is open to address the non-linear problem, that is to prove that *the subextremal Kerr family is nonlinearly stable*. Note that perturbations of Schwarzschild initial data are generally expected to converge to a Kerr solution with small angular momentum, so stability of the Schwarzschild family cannot hold without further restrictions on the data. However, we have the following result, which proves the nonlinear stability of Schwarzschild for the subset of data for which it actually holds:

Theorem 14 ([59]). *Nonlinear stability holds for the Schwarzschild spacetime provided the initial data lie on a codimension 3 submanifold of the moduli space of initial data.*

We emphasize again that the codimension 3 assumption is necessary because the Schwarzschild family is contained as the $a = 0$ subcase of the Kerr family. Outside the codimension 3 submanifold, one expects solutions to necessarily asymptote to a Kerr solution with $a \neq 0$, since the dimension of linearized Kerr solutions fixing the mass is equal to 3 in our parametrization. It is in this sense that Theorem 14 encompasses all data near Schwarzschild that converge back to a member of the Schwarzschild family. We note that Theorem 14 had been proved previously for polarized axisymmetric initial data in [139]. That work already contains some of the difficulties of the full problem. See also [140] for further discussion of their approach to the problem.

While capturing many of the nonlinear difficulties such as identifying the mass of the final solution, constructing teleological gauges and identifying a version of the null condition in them, the fact that the final state in Theorem 14 is Schwarzschild simplifies considerably both the algebra and the analysis. In particular, Theorem 14 can be (and is) proven using entirely physical space based techniques.

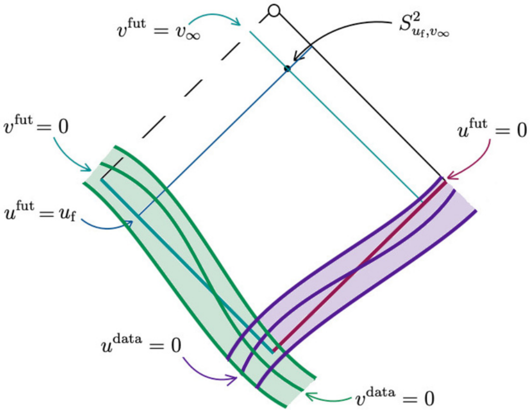
Before we provide a brief overview of the main ideas in the proof, we recall items (i)–(iii) from the characterization of nonlinear stability in Section 2.2. We remark that in the proof of Theorem 14, the global closeness of statement (ii) can be expressed at the top order energy level with respect to the same quantity that measures a suitable "initial" energy quantity, i.e., without loss of derivatives. In this sense, Theorem 14 contains a true orbital stability statement. Note also that Theorem 14 is indeed the nonlinear analogue of Theorem 12 as the latter can be viewed as the statement of linear asymptotic stability of Schwarzschild up to a three-dimensional space of initial data, which, in the linear problem, can be directly identified at the level of initial data.

Theorem 14 is proven by expressing the equations in a double null gauge and hence the natural setup is to prescribe characteristic initial data intersecting in a topological 2-sphere. There is a well-established procedure, indicated at the end of Section 1.3, to prescribe initial data in this setting. We now decompose the space of initial data into disjoint 3-parameter families $D = D_0 + \sum_{m=-1}^{1} \lambda_m D_m^{\mathrm{Kerr}}$, where D_0 varies over a suitable space and is of size ε_0. Here D_m^{Kerr} essentially prescribes the three $\ell = 1$ modes of the torsion on the sphere of intersection and the vector $(\lambda_{-1}, \lambda_0, \lambda_1)$ is a measure of the size of the angular momentum of the data. We prove that given any $d \in D_0$ we can find a $(\lambda_{-1}^{\star}, \lambda_0^{\star}, \lambda_1^{\star})$ such that the corresponding data set D converges to Schwarzschild. We emphasize that the vector $(\lambda^\star_{-1}, \lambda^\star_0, \lambda^\star_1)$ (as well as the final mass of the solution) has to be determined teleologically, i.e., from the entire dynamics of the solution.

We now come to the main ideas of the proof. One crucial ingredient, which we already saw in the linear problem, is the use of a double null gauge which is normalized from the future, i.e., certain Ricci coefficients have to take their Schwarzschild (with mass to be determined!) values on the asymptotic hypersurfaces. The construction of such future gauges, which geometrically corresponds to finding a nearby sphere and foliating the ingoing and outgoing light cones in a prescribed fashion, is based on a implicit function theorem type argument. This uses that in the linear case we can prescribe the desired values by adding a pure gauge solution. However, the nonlinear argument is considerably complicated by the fact that the $\ell = 0$ and $\ell = 1$ modes (mass and angular momentum) require special treatment and couple nonlinearly into the iteration.

Having proven the existence of the future gauges, we can consider a solution in the initial data gauge with coordinates $(u^{\text{data}}, v^{\text{data}}, \theta^{\text{data}})$ or in the future gauge with coordinates $(u^{\text{fut}}, v^{\text{fut}}, \theta^{\text{fut}})$ with relations of the form $u^{\text{fut}} = u^{\text{data}} + f_1(u^{\text{data}}, v^{\text{data}}, \theta^{\text{data}})$, etc., in the region where both gauges are defined. As in the linear problem, we can estimate the f's provided estimates on the Ricci and curvature components are available in both of the gauges.

The proof proceeds by a large scale bootstrap argument along the following lines. Given our 3-parameter family, we consider the largest $u_f > 0$ such that in the future gauge normalized at the sphere $\mathbb{S}^2_{u_f, v_\infty}$ the following bootstrap assumptions hold in the corresponding bootstrap region $\mathcal{M}'(u_f)$ indicated in the picture below:



(I) We have $|f_i| \leq \varepsilon$ in the shaded ("near initial data") region of u-width ≈ 1. In particular, the $u^{\text{fut}} = 0$ cone is ε-close to the cone $u^{\text{data}} = 0$.

(II) We have decay estimates for Ricci coefficients and null curvature components minus their Schwarzschild value (which is determined from the spherical average of the curvature component ρ on $\mathbb{S}^2_{u_f,v_\infty}$) *in the future gauge*, for instance,

$$\left|r^2\left(\Omega\,\mathrm{tr}\,\chi - r^{-1}\left(1-\frac{2M(u_f)}{r}\right)\right)\right| \le \frac{\varepsilon}{u}, \quad \left|r^4\alpha\right| \le \frac{\varepsilon}{u}, \quad \text{etc.,} \tag{2.29}$$

as well as a hierarchy of higher order estimates (in L^2 on spheres, null-cones and spacetime regions) for all $\vec{\lambda} \in \mathcal{R}(u_f)$ such that $|(r^5 \mathrm{c}\not{\mathrm{u}}\mathrm{rl}\,\beta)_{\ell=1}(u_f, v_\infty)| \le \frac{\varepsilon}{u_f}$ with equality on $\partial\mathcal{R}(u_f)$. (That is, for every u_f we define a corresponding set $\mathcal{R}(u_f)$ of admissible $\vec{\lambda}$ which satisfy this.)

The main task then is to show that $u_f = \infty$ by improving the above bootstrap assumptions. This proceeds along the following lines:

(1) One shows that in the shaded region the curvature components $\alpha, \underline{\alpha}$ defined with respect to the future double null gauge agree with the components defined in the initial data double null gauge up to quadratic error terms. This is a manifestation of the fact that α is gauge invariant in linear theory, that is, comparing the α's in the two gauges only produces terms quadratic in f which by bootstrap assumption (I) are indeed $O(\varepsilon^2)$.

(2) One estimates α and $\underline{\alpha}$ in the future gauge from their (now nonlinear) Teukolsky equations. The initial data are $\varepsilon_0 + O(\varepsilon^2)$ by the previous step, with ε_0 denoting the size of the initial data. Here the main challenge beyond linear theory is to estimate the nonlinear error terms, which can be shown to exhibit a version of the null condition. This improves the bounds on α and $\underline{\alpha}$ from $\frac{\varepsilon}{u}$ to $\frac{\varepsilon_0+\varepsilon^2}{u}$.

(3) One estimates all Ricci and curvature coefficients in the future gauge from the improved bounds on α and $\underline{\alpha}$ and the gauge conditions on the asymptotic hypersurfaces. This improves all estimates from $\frac{\varepsilon}{u}$ to $\frac{\varepsilon_0+\varepsilon^2}{u}$. This step involves a number of technical difficulties most of which are, however, present (in a milder form) in the linear theory.

(4) One improves $|f| \le \varepsilon$ from the fact that by the previous step one now has, in the initial data region, bounds on the Ricci coefficients in the future gauge ($\le \varepsilon_0 + \varepsilon^2$) and bounds on these coefficients in the initial data gauge (by Cauchy stability). Since f can be estimated from these, we deduce $|f| \le \varepsilon_0 + \varepsilon^2$. This improves (II). Note that the fact that the initial data remains close to the old initial data is a nonlinear version of the fact that the pure gauge solution one needed to add in Theorem 12 was uniformly bounded by initial data.

(5) Having improved all the estimates, we can extend the spacetime slightly to retarded time $u_f + \delta$ and construct a new future gauge from a new future sphere. The key now is to show that $\mathcal{R}(u_f + \delta) \subsetneq \mathcal{R}(u_f)$, i.e., that the set of admissible λ of our three parameter family shrinks. This strict monotonicity can be established by carefully examining the evolution of angular momentum between the

"old" sphere and the "new" sphere. Finally, it is the topological degree of the map from the space of λ's (i.e., $\mathcal{R}(u_f)$) to the space of angular momenta in the future that is bootstrapped and ensures that for every u_f the set $\mathcal{R}(u_f)$ contains a tuple $(\lambda_{-1}, \lambda_0, \lambda_1)$ that gets mapped to zero angular momentum. This in turn implies that we can construct a sequence $(u_f)_i \to \infty$ with corresponding $\lambda_i \to \lambda^\star \in \bigcap \mathcal{R}((u_f)_i)$.

2.5.2. The case $\Lambda > 0$

The proof of the nonlinear stability of slowly rotating Kerr–de Sitter (KdS) black holes by Hintz–Vasy [115] applies spectral theoretic and microlocal methods to the analysis of a variant of the quasilinear wave equation (1.5). Consider a neighborhood

$$\mathcal{M} := [0,\infty)_{t_*} \times \Sigma, \quad \Sigma = [r_1, r_2] \times \mathbb{S}^2 \quad (r_1 = r_- - \varepsilon,\ r_2 = r_+ + \varepsilon)$$

of the black hole exterior for a subextremal Schwarzschild–de Sitter metric $g_{\Lambda,M_0,0}$ in the causal future of a spacelike hypersurface $t_*^{-1}(0) \cong \Sigma$; here, r_1, r_2 are the radii of the event and cosmological horizon, respectively, and t_* is a time function whose level sets are transversal to the future event and cosmological horizons. (See also Figure 9.) For (M, a) near (M_0, a), one can consider the KdS metric $g_{\Lambda,M,a}$ as a stationary metric on $\mathcal{M}$ with smooth dependence on (M, a); in particular, future affine complete pieces of the event and cosmological horizons of these nearby KdS black holes are contained in $\mathcal{M}$ still.

The desired asymptotic stability statement suggests writing the spacetime metric g as $g = g_{\Lambda,M,a} + \tilde{g}$ and regarding the final black hole parameters (M, a) as unknowns; the gravitational wave tail $\tilde{g}$ is an unknown as well and required to be exponentially decaying. The starting point for the gauge is the generalized harmonic gauge 1-form with coefficients $W(g)_\mu = g_{\mu\nu} g^{\kappa\lambda}(\Gamma(g)^\nu_{\kappa\lambda} - \Gamma(g^0)^\nu_{\kappa\lambda})$ measuring the failure of $(\mathcal{M}, g) \to (\mathcal{M}, g^0)$ to be a wave map; here we take

$$g^0 = g_{\Lambda,M_0,0}.$$

Since a KdS metric $g_{\Lambda,M,a}$ for $(M, a) \neq (M_0, 0)$ has no reason to satisfy the gauge condition $W(g_{\Lambda,M,a}) = 0$, one should really use the gauge condition $W(g) - W(g'_{\Lambda,M,a}) = 0$ depending on the unknown final parameters (M, a); here

$$g'_{\Lambda,M,a} = \chi g_{\Lambda,M_0,0} + (1-\chi) g_{\Lambda,M,a}$$

interpolates between $g_{\Lambda,M_0,0}$ for $t_* \le 1$ and $g_{\Lambda,M,a}$ for $t_* \ge 2$. For further flexibility, one allows for a gauge source 1-form ϑ with compact support (or appropriate decay) in time; ϑ will lie in a suitable finite-dimensional space of 1-forms.

The nonlinear equation solved in [115] is then

$$\begin{aligned} P(M, a, \tilde{g}, \vartheta) &:= \mathrm{Ric}(g) - \Lambda g - (\delta^*_{g^0} + q)\big(W(g) - W(g'_{\Lambda,M,a}) - \vartheta\big) = 0, \\ g &= g'_{\Lambda,M,a} + \tilde{g}. \end{aligned} \tag{2.30}$$

Here, the stationary bundle map q is chosen so as to implement constraint damping, i.e., so that homogeneous solutions of the wave operator $\mathrm{div}_{g^0}\, \mathsf{G}_{g^0}(\delta^*_{g^0} + q)$ decay exponentially in time (i.e., it satisfies mode stability in an upper half plane $\Im\sigma \geq -\alpha$ for some $\alpha > 0$). Considerable effort is required to show the existence of such a q; see [115, §8], where q is defined using a large parameter, and the required mode stability is proved using asymptotic analysis in the large parameter.

In order to analyze (2.30), consider first the linearization of the right-hand side of (2.30) in g around $g = g^0 = g_{\Lambda,M_0,0}$: it maps

$$L : h \mapsto D_{g^0}\mathrm{Ric}(h) - \Lambda h - (\delta^*_{g^0} + q)\,\mathrm{div}_{g^0}\,\mathsf{G}_{g^0}.$$

The spectral and mode analysis of this equation is in parts very close to that in the Kerr case discussed previously, namely one combines constraint damping with the metric mode stability for linearized perturbations of the Schwarzschild–de Sitter metric [141]. Unlike in the Kerr case, however, mode stability for the 1-form wave equation

$$\mathrm{div}_{g^0}\,\mathsf{G}_{g^0}(\delta^*_{g^0}\omega) = 0 \tag{2.31}$$

(governing those gauge potentials ω whose symmetric gradients satisfy the linearized gauge condition) is not known, and indeed can be shown to *fail* in the de Sitter case (i.e., without the presence of a black hole) due to the presence of a finite number of resonances in the upper half-plane.[9]

Thus, solutions of $Lh = 0$ have a partial resonance expansion (ignoring multiplicities) of the schematic form

$$h(t_*, x) = \left(\frac{\mathrm{d}}{\mathrm{d}s} g_{\Lambda,M_0+sM',sa'}|_{s=0} + [\text{gauge correction}]\right) + \sum_{j=1}^{N} c_j \delta^*_{g^0}(e^{-i\sigma_j t_*}\omega_j)$$
$$+ \mathcal{O}(e^{-\alpha t_*}),$$

where the σ_j are the finitely many modes with $\Im\sigma_j \geq 0$ corresponding to the resonances of the wave operator in (2.31); the ω_j are the corresponding mode solutions, and the c_j are suitable complex scalars depending on the initial conditions of h. The terms are then handled as follows:

(1) The gauge correction is a pure gauge term ensuring that the first summand on the right satisfies the linearized gauge condition $D_{g^0}W = 0$; it would be absent if we had $W(g_{\Lambda,M,a}) = 0$ for all (M, a) near $(M_0, 0)$. Changing the final black hole parameters from $(M_0, 0)$ to $(M_0 + M', a')$, and correspondingly changing the final gauge condition in (2.30) gives rise to essentially the same term.

(2) The nondecaying pure gauge contributions from the terms $\delta^*_{g^0}(e^{-i\sigma_j t_*}\omega_j)$ can be eliminated for late times t_* by an (explicit) change of the gauge condition, i.e., by solving

$$Lh - (\delta^*_{g^0} + q)\vartheta = 0$$

[9] This is another reason why the choice of q in (2.30) requires large parameter techniques: q *cannot* be small if it is to shift these resonances all the way into the lower half-plane.

with a suitable (explicit, depending on the c_j and ω_j) gauge source 1-form ϑ; there is one such 1-form for each of the N pure gauge mode solutions.

(3) The $\mathcal{O}(e^{-\alpha t_*})$ tail contributes to the exponentially decaying tail $\tilde{g}$.

This can be rephrased as follows: the linearization of $P(M,a,\tilde{g},\vartheta)$ at $(M_0,0,0,0)$ in the argument $\tilde{g}$ is surjective if one supplements the range by a finite-dimensional space consisting of

(1) linearized black hole parameter (and associated gauge) changes—i.e., the range of the linearization of P in (M,a); and

(2) gauge modifications—i.e., the range of the linearization of P in ϑ acting on an N-dimensional space of 1-forms ϑ.

Perturbative arguments prove this surjectivity for $(M,a,\tilde{g},\vartheta)$ near $(M_0,0,0,0)$. Thus, small data initial value problems for $P(M,a,\tilde{g},\vartheta)=0$ can be solved using a Newton iteration scheme; due to a loss of derivatives due to trapping, [115] really uses a Nash–Moser iteration in the simple form given by Saint-Raymond [185]. Once one has a solution of the Cauchy problem for $P(M,a,\tilde{g},\vartheta)=0$, the standard arguments sketched in Section 1.2 imply that $g=g'_{\Lambda,M,a}+\tilde{g}$ is a solution of the initial value problem for $\mathrm{Ric}(g)-\Lambda g=0$.

Theorem 15 ([115]). *Let $\Lambda>0$ and $M_i>0$, $a_i\in\mathbb{R}$, $|a_i|\ll M_i$. Let t_* (given by $t_*=t-F(r)$ in Boyer–Lindquist coordinates for suitable F) be a time function whose level sets are transversal to the future event and cosmological horizons. Let $\Sigma\subset t_*^{-1}(0)$ denote a spacelike hypersurface which extends a bit beyond the event and cosmological horizons. Suppose $\overline{g}$, K are solutions of the constraint equations (1.3) which are close (in the norm of $H^{21}(\Sigma;S^2T^*\Sigma)$) to the initial data at Σ of the metric g_{Λ,M_i,a_i}. Then on $\mathcal{M}=[0,\infty)_{t_*}\times\Sigma$ there exists a solution g of the initial value problem for* $\mathrm{Ric}(g)-\Lambda g=0$ *which decays exponentially fast to a Kerr–de Sitter metric: there exist $M_f>0$ and $a_f\in\mathbb{R}$, with (M_f,a_f) close to (M_i,a_i), and $\alpha>0$ so that*

$$g=g_{\Lambda,M_f,a_f}+\tilde{g},\quad \tilde{g}=\mathcal{O}(e^{-\alpha t_*}).$$

In particular, by the stable manifold theorem, the event and cosmological horizons of the perturbed spacetime $(\mathcal{M},g)$ are exponentially decaying (as $t_*\to\infty$) perturbations of the horizons of g_{Λ,M_f,a_f}; in other words, $(\mathcal{M},g)$ contains these two future affine complete horizons. For partial results on the stability of the cosmological (asymptotically de Sitter) part of Kerr–de Sitter spacetimes, see [190,191].

To complete the discussion of the proof of Theorem 15, we explain a few aspects of the (non-)linear analysis on asymptotically Kerr–de Sitter spaces. The nonlinear iteration scheme used in the proof of Theorem 15 involves the global solution, at each step, of a linear wave equation

$$L_{M,a,\tilde{g},\vartheta}h=[\text{nonlinear error term}]. \tag{2.32}$$

Here $L_{M,a,\tilde{g},\vartheta}$ is a wave operator (acting on symmetric 2-tensors) on the spacetime $\mathcal{M}$ equipped with a metric $g_{\Lambda,M,a}+\tilde{g}$ that settles down exponentially fast to a KdS metric.

Thus, the spectral methods which are effective for providing sharp asymptotics of stationary problems need to be supplemented by estimates *on the nonstationary spacetime* which are effective for proving the sharp regularity of linear waves. Roughly speaking, given a solution h of the linear wave equation (2.32) which, together with a large number of derivatives, obeys a weak exponential bound $h = \mathcal{O}(e^{Ct_*})$ for some fixed C (such estimates are discussed below), one can rewrite this equation as

$$L_0 h = -\tilde{L}h,$$

where $L_0 = L_{M,a,0,0}$ is the stationary part and $\tilde{L} = L_{M,a,\tilde{g},\vartheta} - L_{M,a,0,0}$ the (second order) remainder *with exponentially decaying* $\mathcal{O}(e^{-\alpha t_*})$ *coefficients*. Spectral methods for L_0 thus give precise asymptotics for h up to errors with an extra $e^{-\alpha t_*}$ amount of decay relative to the a priori information on h, i.e., $\tilde{L}h = \mathcal{O}(e^{(C-\alpha)t_*})$. Full asymptotics for h can then be obtained by iteration.

It thus remains to show (arbitrarily) high regularity of h in a space allowing for a fixed amount of exponential growth. Energy estimates give a simple exponential bound in H^1. Higher regularity of h (in the *same* exponentially weighted space in time) is then proved by microlocal means: regularity of the initial data is propagated using the Duistermaat–Hörmander theorem for finite times; uniform control as $t_* \to \infty$ requires the use of radial point estimates (on exponentially weighted function spaces) near the horizons [111], and simple (since we are allowing for exponential growth of solutions) estimates at the trapped set [104, 110]. We remark that the use of Nash–Moser iteration requires the proof of *tame* versions of all these microlocal estimates; these were first given in [112].

Theorem 15 was subsequently extended by Hintz to the setting of charged black holes:

Theorem 16 ([105]). *The family of Kerr–Newman–de Sitter black holes with subextremal charge and small angular momenta is nonlinearly asymptotically stable. That is, the spacetime metric and electromagnetic 2-form evolving from a small perturbation of the initial data of such a Kerr–Newman–de Sitter black hole settle down exponentially fast to the metric and electromagnetic 2-form of a nearby Kerr–Newman–de Sitter metric in a suitable gauge (generalized harmonic gauge for the metric, generalized Lorenz gauge for the electromagnetic field).*

Previously, Hintz–Vasy [112] had shown the solvability of quasilinear wave equations on slowly rotating Kerr–de Sitter spacetimes (by combining microlocal and spectral methods as sketched above) assuming the absence of modes in the closed upper half-plane, following earlier work on asymptotically de Sitter spacetimes [102, 111].

2.5.3. The case $\Lambda < 0$

As in the case of the maximally symmetric solutions in Section 2.1.3, the $\Lambda < 0$ case (with reflective boundary conditions) may turn out to be the most difficult and constitutes a major outstanding problem in the field. The slow logarithmic rate obtained for the toy problem (see Section 2.3) lead [120] to conjecture nonlinear instability. However, there could be subtle cancelations entering the nonlinear dynamics that allow for some version of orbital stability. Note that the nature of the slow decay (which disappears if only finitely many modes are taken into account) makes detecting an instability in numerical simulations difficult. Finding an appropriate nonlinear toy problem where some of the difficulties can be understood seems to be a first step to attack this problem.

3. SINGULARITIES

We discuss results concerning singularities occurring in the interior of black holes in Section 3.1. Recent progress on naked singularities is discussed in Section 3.2.

3.1. The interior of black holes

Whereas the exterior (or, indeed, suitable neighborhoods of the exterior) of subextremal Kerr black holes is conjectured to be stable, the situation is different for the black hole *interior*. Note in particular that Kerr black holes with nonzero angular momentum have a nonempty Cauchy horizon, whereas Schwarzschild black holes have an entirely different interior structure, namely they have a terminal spacelike singularity across which the metric cannot be extended even as a continuous Lorentzian metric [188]. Regarding thus the interior structure rotating Kerr black holes as a reference point, a heuristic due to Simpson–Penrose [195] suggests that the Kerr Cauchy horizon is unstable, in the sense that for generic perturbations of the initial data the spacetime metric becomes singular at the Cauchy horizon. This is the content of Penrose's *Strong Cosmic Censorship* conjecture. The basic idea is that linear waves falling into the black hole are more and more blue-shifted as one approaches the Cauchy horizon, which when upgraded to the nonlinear setting is suggestive of the formation of a singularity. This heuristic also suggests a direct relationship between decay of perturbations in the black hole exterior and the regularity of the metric near the Cauchy horizon.

The precise notion of singularity to be used depends on the context. One notion [44] asks for the nonexistence of an extension of the spacetime with square integrable Christoffel symbols since square integrability is sufficient to make sense of the Einstein equations in a weak sense; other often used notions are $\mathscr{C}^2$-inextendibility, as it relates directly to the blow-up of curvature invariants such as the Kretschmann scalar, or $\mathscr{C}^0$-inextendibility of the metric. The current state-of-the-art in the vacuum case is the following *regularity* theorem; its proof uses the double gauge (which is particularly convenient also for locating the Cauchy horizon):

Theorem 17 ([60]). *Assume quantitative decay rates*[10] *of the metric and second fundamental form, along a spacelike hypersurface Σ_0 just beyond the event horizon, to the data of a subex-*

[10] These assumptions are compatible with the conjectured nonlinear stability of the Kerr family.

tremal Kerr metric with nonzero angular momentum. Then the future development of these data has a nontrivial Cauchy horizon across which the metric is continuously extendible.

Related regularity results for the linear scalar wave equation were proved by Franzen [85,86] and Hintz [103]. Reading Theorem 17 as a statement about the stability of the interior structure of Kerr black holes, one may consider the analogous problem for the stability of the Schwarzschild singularity (which necessarily requires working in a restricted symmetry class to disallow Kerr behavior). This was tackled by Alexakis–Fournodavlos [3] in polarized axial symmetry; see also [84] on the behavior of linear waves in the Schwarzschild interior.

For de Sitter black holes, waves decay in the exterior at an exponential rate, and correspondingly the regularity of metrics or linear scalar waves is expected to be higher at the Cauchy horizon. Quantitatively, linear waves with energy decaying like $\mathcal{O}(e^{-\alpha t_*})$ have almost $H^{1/2+\alpha/\kappa}$-regularity across the Cauchy horizon (and arbitrary regularity in the angular variables), where κ is the surface gravity of the Cauchy horizon [113] (see also [49]). This regularity can exceed H^1 for certain black hole parameters [34]; heuristically this corresponds to the expectation that the analogue of Theorem 17 in such settings yields spacetimes which, even upon perturbation, can be extended with square integrable Christoffel symbols. For rigorous results in this direction, see [50–52].

The first result on the existence of *singularities* was obtained for the linear scalar wave equation on Reissner–Nordström spacetimes with nonzero charge by Luk–Oh [150]; the key is the identification of a conserved quantity along null infinity, the nonvanishing of which allows for the propagation of suitable lower bounds into the black hole interior which imply blow-up of energy at the Cauchy horizon. The result by Luk–Sbierski [156] proves blow-up under the assumption of pointwise *lower* bounds for the linear wave along the event horizon of rotating Kerr black holes; these lower bounds were proved in [11,106], as discussed in Section 2.3.3.

Singularity formation at the Cauchy horizon for solutions of the Einstein equation is thus far only known in spherical symmetry for suitable matter models. Christodoulou [41] proved the $\mathcal{C}^0$-formulation of the Strong Cosmic Censorship conjecture for the Einstein–real scalar field system in spherical symmetry. In the presence of charge, Dafermos [54] with Rodnianski [61], on the other hand, proved that $\mathcal{C}^0$-regularity *does* hold for the Einstein–*Maxwell*–real scalar field system; this was complemented by the following result on the genericity of $\mathcal{C}^2$-singularities (improved to $\mathcal{C}^{0,1}$ in [189]):

Theorem 18 ([151,152]). *The $\mathcal{C}^2$ formulation of the Strong Cosmic Censorship conjecture for the Einstein–Maxwell–real scalar field system in spherical symmetry (with 2-ended asymptotically flat initial data on $\mathbb{R} \times \mathbb{S}^2$) is true.*

On charged Reissner–Nordström–AdS black hole spacetimes (thus $\Lambda < 0$), the behavior of linear waves near the Cauchy horizon was understood only recently in a series of works by Kehle who proved $\mathcal{C}^0$ bounds [135] and generic energy blow-up [134,136] depending on the validity of a diophantine condition on the quasinormal modes. Part of the difficulty here is the particularly slow (logarithmic) decay on the exterior.

3.2. Naked singularities

The singularities discussed in Section 3.1—be they spacelike or null—all share the feature of being "behind" an event horizon of a black hole exterior possessing a complete null infinity. Penrose's *Weak Cosmic Censorship* conjecture asserts that this is generically the case for solutions arising from asymptotically flat initial data: singularities always occur in the causal future of an event horizon and can thus not communicate with asymptotic observers at infinity. In [42,43] Christodoulou showed that the word "generic" is indeed necessary. He constructed solutions to the spherically symmetric Einstein scalar field system containing a naked singularity, i.e., spacetimes whose Penrose diagram looks as in Figure 11. In particular, the cone $\mathcal{N}$ is future null geodesically incomplete and does not extend to $\mathcal{O}$. This can be seen by the quantity $\frac{2m}{r}$ being bounded uniformly from below by a positive constant along $\mathcal{N}$, where m denotes the Hawking mass and r the area radius function of the spherically symmetric spacetime. In particular, one cannot make sense of the Einstein equations in any reasonable sense (in particular, not in the class of bounded variation) on or to the future of $\mathcal{O}$.

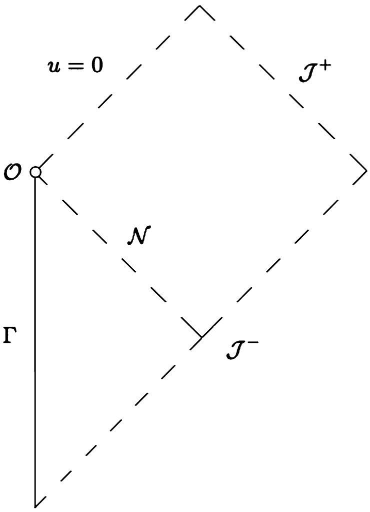


FIGURE 11
The naked singularity spacetimes of [42]. Here Γ denotes the center of the spherical symmetry. Null infinity $\mathcal{I}^+$ is future incomplete.

Christodoulou's construction relied on two fundamental ingredients. Let us denote a solution to the spherically symmetric Einstein scalar field system by $(g_{\mu\nu}, r, \phi)$. First, Christodoulou proved local well-posedness of the system in the (low regularity) class of BV solutions. Secondly, he introduced the notion of a k-self-similar solution of the system. This is a solution which admits a 1-parameter group of diffeomorphisms $(f_a)_{a\geq 0}$ such that $f_a^\star g = a^2 g$, $f_a^\star r = ar$ and $f_a^\star \phi = \phi - k \log a$. Imposing self-similarity and spherical symmetry reduces the Einstein equations to a two-dimensional autonomous dynamical system,

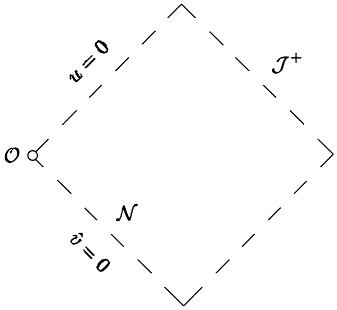


FIGURE 12
The naked singularity spacetimes of [183] arising from solving a characteristic initial value problem. Null infinity $\mathcal{J}^+$ is future incomplete.

whose dynamics Christodoulou analyzed. A suitable subset of its solutions could be interpreted as BV solutions (in fact they have higher regularity) and could be described by the above Penrose diagram. The key here is that the appearance of naked singularities fundamentally depends on k being *nonzero*.

In recent work, Rodnianski and Shlapentokh-Rothman transformed some of the above ideas to prove a result for the *vacuum equations without any symmetry assumptions*:

Theorem 19 ([183]). *There exists a large class of solutions to the vacuum Einstein equations containing naked singularities.*

The solutions of Theorem 19 are constructed directly by solving a characteristic initial value problem as discussed at the end of Section 1.3 leading to a Penrose diagram of the form shown in Figure 11.

The analogue of Christodoulou's well-posedness result for BV solutions in the proof of Theorem 19 is provided by the well-posedness theory that has been developed in recent years to construct low regularity solutions of the Einstein equations in a double null gauge, in particular the Luk–Rodnianski theory of impulsive gravitational waves [153,154][11] and the results of [182]. The idea is that one can construct solutions with very limited regularity in the null directions u and v but high regularity in the directions tangent to the double null spheres.

The analogue of Christodoulou's k-self-similar solutions can be described as follows. Assume momentarily that the solution has already been constructed in double null coordinates (note that the shift vector b has been put in the other null direction)

$$\begin{aligned} g &= -2\Omega^2(\mathrm{d}u \otimes \mathrm{d}v + \mathrm{d}v \otimes \mathrm{d}u) + \not{g}_{AB}(\mathrm{d}\theta^A - b^A\mathrm{d}u)(\mathrm{d}\theta^B - b^B\mathrm{d}u) \\ &= -2\Omega^2 v^{2\kappa}(1-2\kappa)^{-1}(\mathrm{d}u \otimes \mathrm{d}\hat{v} + \mathrm{d}\hat{v} \otimes \mathrm{d}u) + \not{g}_{AB}(\mathrm{d}\theta^A - b^A\mathrm{d}u)(\mathrm{d}\theta^B - b^B\mathrm{d}u), \end{aligned} \tag{3.1}$$

[11] In this theory, the components α and $\underline{\alpha}$ are not even in L^2_u and L^2_v. They can, however, be removed from the system of double null equations by a clever renormalization of the system.

with the two sets of double null coordinates related by $\hat{v} = v^{1-2\kappa}$ for a small and positive κ.

We say that a solution is self-similar if the scaling vector field $S = u\partial_u + v\partial_v$ satisfies $\mathcal{L}_S g = 2g$. This directly translates into constraints on the behavior of the metric functions, for instance $\Omega(u, v, \theta^A) = \check{\Omega}(\frac{v}{u}, \theta^A)$, etc. The κ-self similarity now enters by imposing the manner in which the metric and its derivatives extend to the cone $\mathcal{N}$, i.e., the hypersurface $\hat{v} = 0$. A κ-self similar solution with $\kappa \neq 0$ is defined in a way that the metric and its derivatives extend regularly ($C^{1,\gamma}$) to $\mathcal{N}$ in the $(u, \hat{v}, \theta)$ coordinates. In the (u, v, θ) coordinates, however, quantities will then be singular with specific powers of $\frac{v}{u}$; in other words, (u, v, θ) coordinates are not regular on $\mathcal{N}$.

To view this more geometrically, note the difference of the generator of the null cone $\hat{v} = 0$ given by $e_3|_{\hat{v}=0} = \partial_u + \tilde{b}^A\partial_{\theta^A}$ and the restriction of the scaling field $S|_{\hat{v}=0} = u\partial_u$. If we can construct solutions with $\tilde{b} \neq 0$ along $\mathcal{N}$, then there will be a twisting of the generators by the self-similar vector field along $\mathcal{N}$. The point now is that if $\kappa = 0$, then the constraint equations along $\mathcal{N}$ necessitate $\tilde{b} = 0$ on $\mathcal{N}$. If $\kappa \neq 0$, then extra terms appear in the constraint equations which allow for nontrivial $\tilde{b}$; this is the main mechanism for the naked singularity formation (and for proving the analogue of $\frac{2m}{r}$ being bounded below along $\mathcal{N}$).

The proof itself has two main steps. The constraint equations need to be solved along the ingoing and outgoing light cones in a way that is consistent with the self-similarity of the solution and in a way that the low regularity well-posedness results mentioned earlier still apply. This requires (as expected from the nongenericity of naked singularities!) fine tuning of the "free data" (Section 1.3) and a detailed analysis of the regularity at the intersection of the two light cones. Once the data is constructed and the local well-posedness theorem applied, the proof proceeds in a large scale bootstrap argument to complete the picture shown in the Penrose diagram of Figure 12. This uses the familiar scheme of energy estimates for the curvature components and transport equations for the connection coefficients; however, many intricate renormalizations (subtracting the singular self-similar part of any dynamical quantity) and careful choices of $\frac{v}{u}$-weights in the estimates are required.

4. FURTHER TOPICS

We briefly discuss two more topics of continued or recent interest: the construction of multi-black-hole spacetimes, and inverse problems for nonlinear wave equations on Lorentzian manifolds.

4.1. Black hole gluing

All spacetimes discussed so far contain at most a single black hole. There exist explicit solutions of the Einstein–Maxwell equations without or with cosmological constant, called Majumdar–Papapetrou [158, 172] and Kastor–Traschen [132] spacetimes. They are, however, very rigid, being based on a special algebraic ansatz for the metric and electromagnetic field, in which the functions controlling the ansatz solve a linear Laplace equation and indeed are shifted and scaled versions of $1/|x|$. These spacetimes can be regarded as containing black holes each of which have charge equal to their mass. A construction due to Brill–Lindquist [30] produces a many-black-hole vacuum solution by similar rigid means.

The first flexible construction of many-black-hole spacetimes with well-controlled asymptotic structure glues any (finite) number of Kerr–de Sitter black holes into neighborhoods of points on the future conformal boundary of de Sitter space.

Theorem 20 ([107]). *Let $\Lambda > 0$. Fix points $p_1, \dots, p_N \in \mathbb{S}^3$, masses $M_1, \dots, M_N$, and angular momenta $a_1, \dots, a_N$; assume that a certain balance condition holds. (When all a_j vanish, this balance condition reads $\sum_{j=1}^{N} M_j p_j = 0$, where one regards $\mathbb{S}^3 \subset \mathbb{R}^4$ as the unit sphere.) Then there exists a solution g of $\mathrm{Ric}(g) - \Lambda g = 0$ which near the point p_j on the future conformal boundary of de Sitter space is equal to g_{Λ, M_j, a_j}, and away from the points p_j converges to the de Sitter metric at an exponential rate.*

The gluing is accomplished via a backwards (or scattering) construction: a naively glued ansatz for the metric (which is an exact solution near the p_j) is corrected, *in the gluing region* (i.e., leaving neighborhoods of the p_j unaffected) [68], in Taylor series at the conformal boundary in a suitable generalised harmonic gauge, and the remaining error is solved away exactly by solving the gauge-fixed Einstein equation backwards from the conformal boundary. (See [56] for a loosely related scattering construction for asymptotically Kerr black holes.) The balance condition arises as an obstruction (cokernel) to the existence of a particular term in the Taylor expansion which needs to satisfy a linear divergence equation and yet have support away from the p_j.

Previous gluing constructions took place on the level of initial data sets, starting with Corvino's seminal work [47,48] on localized gluing and followed by many variants and generalizations (including wormholes, localized gluing in angular sectors) [35,46,128,129]. For recent gluing results for the characteristic initial value problem, see [16].

Chruściel–Mazzeo succeeded, using Friedrich's nonlinear stability result [87], in describing parts of the global structure of the spacetime evolving from the many-black-hole data of [46]; in particular, they show that the complement of the causal past of suitable observers at null infinity has several connected components, corresponding to several black hole regions. Analyzing the structure of compact subsets of these spacetimes is, however, entirely out of reach; the same is true for the spacetimes evolving from the initial data constructed numerically and used in numerical relativity for the study of black hole mergers [173].

The problem of constructing many-black-hole spacetimes with precise asymptotic control in the asymptotically flat setting (e.g., two Kerr black holes moving apart at a positive speed) is an interesting and challenging problem, (variants of) which may well be within reach in the near future.

4.2. Inverse problems

A rather different topic of investigation concerns the determination of a spacetime from measuring the propagation of waves inside the spacetime. This is typically phrased as the problem of reconstructing as large a spacetime region as possible from the Dirichlet-to-

Neumann map for boundary value problems for (nonlinear) wave equations on domains with timelike boundary, or from the source-to-solution map for forcing problems. For the linear wave equation on backgrounds *with time-independent metrics*, a complete solution of the first class of problems was obtained by Belishev–Kurylev [20] using a unique continuation result by Tataru [196].

A beautiful recent insight by Kurylev–Lassas–Uhlmann [143] is that nonlinearities can actually *simplify* the solution of the inverse problem, in particular in settings where the corresponding problem for the linear equation is not yet solved (e.g., for nonstationary metrics with nonanalytic time dependence).

The basic idea is that one can produce $\dim \mathcal{M} = 4$ small and mildly singular (distorted) plane wave solutions by imposing suitable Dirichlet boundary conditions or specifying suitable forcing terms. If these distorted plane waves interact nonlinearly at a spacetime point q, a new spherical wave is produced at q in the sense that a (very weak) singularity emanates from q; this singularity can be detected in the Neumann data or in some open set of the spacetime where one makes measurements. In this manner, the inverse problem is reduced to a geometric problem of reconstructing the Lorentzian structure (typically up to conformal diffeomorphisms) of a causal diamond $D \subset \mathcal{M}$ from the collection of the *light observation sets*—intersections of future light cones from points in q with the observation region. This geometric problem was solved in [143] for measurements in open subsets of $\mathcal{M}$, and in [109] for measurements on timelike boundaries under a convexity assumption.

There exists by now a large literature on similar inverse problems; here we only mention the results [142–144] on inverse problems concerned directly with the Einstein–scalar field and Einstein–Maxwell equations.

5. CONCLUSIONS AND OUTLOOK

The mathematical study of Einstein's theory of General Relativity is very natural: the main equation of the theory is "simple" despite being fundamental, i.e., not derived from a more general (classical) theory via any sort of approximation or averaging. And yet the structure of its solutions is fantastically rich, which thus provides a large arena for detailed investigations of various aspects of the theory.

It is a remarkable feature of general relativity that the simplest nontrivial vacuum spacetime—the Schwarzschild solution or its analogues in the presence of a nonzero cosmological constant—describes a *black hole*. (Moreover, even if the path of history was different, the study of (stationary) perturbations of the Schwarzschild solution could well have hinted at the existence of the Kerr family!) The study of perturbations of Schwarzschild or Kerr black holes in the context of the initial value problem can be regarded as a theoretical exploration of the question whether these solutions bear relevance as models for physical black holes. As discussed in Section 2, this line of investigation led to the discovery of fascinating geometric and analytic properties of Kerr spacetimes (such as the red-shift effect and superradiance, normally hyperbolic trapping, and mode stability), and inspired a vast amount of work, especially in the theory of partial differential equations, aimed at exploiting these properties (such as refined vector field methods, microlocal radial point and trapping estimates, flexible gauge-fixing methods). The confluence of a large variety of techniques from several areas of mathematics is the reason both for the recent success and for the excitement in the field. Given the progress discussed in Section 2, we anticipate a full resolution of the nonlinear stability problem for subextremal Kerr black holes in the near future. But even then will there be plenty of room for developments, e.g., coupling the Einstein equations with matter.

Analyzing the structure of singularities, whether cosmological, naked, or hidden behind event horizons of black holes, promises to continue being a fruitful area of research. In particular, controlling or constructing spacetimes with singularities requires the development of tools for the study of *large* data regimes (i.e., far from explicit model spacetimes), or calls for deep insights to find settings where large data regimes can still be regarded as perturbative in some sense.

The study of many-black-hole spacetimes has barely started. Motivated in particular by the recent experimental discoveries of black hole mergers, as well as by the advanced understanding of individual black holes, we anticipate this area of research to become prominent soon.

As the field advances, the technical demands will of course increase; however, we are confident that conceptual insights and the development of further elegant, yet powerful mathematical tools will act in a counterbalancing manner, thus keeping the field accessible and vibrant.

ACKNOWLEDGMENTS

The authors thank Gabriele Benomio, Mihalis Dafermos, Dejan Gajic, Yakov Shlapentokh-Rothman, and András Vasy for comments on the manuscript. G. Holzegel thanks Sakis Chatzikaleas for help with the pictures.

FUNDING

P. Hintz was partially supported by NSF Grant No. DMS-1955614 and a Sloan Research Fellowship. G. Holzegel acknowledges support by the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship endowed by the Federal Ministry of Education and Research and support by ERC Consolidator Grant 772249.

REFERENCES

[1] Oberwolfach report. 2021, preliminary version available at https://www.mfo.de/occasion/2135.

[2] S. Aksteiner and L. Andersson, Charges for linearized gravity. *Classical Quantum Gravity* **30** (2013), no. 15, 155016, 20 pp.

[3] S. Alexakis and G. Fournodavlos, Stable space-like singularity formation for axi-symmetric and polarized near-Schwarzschild black hole interiors. 2020, arXiv:2004.00692.

[4] M. T. Anderson, Existence and stability of even-dimensional asymptotically de Sitter spaces. *Ann. Henri Poincaré* **6** (2005), no. 5, 801–820.

[5] L. Andersson, T. Bäckdahl, P. Blue, and S. Ma, Stability for linearized gravity on the Kerr spacetime. 2019, arXiv:1903.03859.

[6] L. Andersson and P. Blue, Hidden symmetries and decay for the wave equation on the Kerr spacetime. *Ann. of Math.* **182** (2015), 787–853.

[7] Y. Angelopoulos, S. Aretakis, and D. Gajic, Horizon hair of extremal black holes and measurements at null infinity. *Phys. Rev. Lett.* **121** (2018), no. 13.

[8] Y. Angelopoulos, S. Aretakis, and D. Gajic, Late-time asymptotics for the wave equation on spherically symmetric, stationary spacetimes. *Adv. Math.* **323** (2018), 529–621.

[9] Y. Angelopoulos, S. Aretakis, and D. Gajic, Logarithmic corrections in the asymptotic expansion for the radiation field along null infinity. *J. Hyperbolic Differ. Equ.* **16** (2019), no. 01, 1–34.

[10] Y. Angelopoulos, S. Aretakis, and D. Gajic, Late-time asymptotics for the wave equation on extremal Reissner–Nordström backgrounds. *Adv. Math.* **375** (2020), 107363.

[11] Y. Angelopoulos, S. Aretakis, and D. Gajic, Late-time tails and mode coupling of linear waves on Kerr spacetimes. 2021, arXiv:2102.11884.

[12] Y. Angelopoulos, S. Aretakis, and D. Gajic, Price's law and precise late-time asymptotics for subextremal Reissner–Nordström black holes. 2021, arXiv:2102.11888.

[13] S. Aretakis, Stability and instability of extreme Reissner–Nordström black hole spacetimes for linear scalar perturbations I. *Comm. Math. Phys.* **307** (2011), no. 1, 17–63.

[14] S. Aretakis, Stability and instability of extreme Reissner–Nordström black hole spacetimes for linear scalar perturbations II. *Ann. Henri Poincaré* **12** (2011), no. 8, 1491–1538.

[15] S. Aretakis, Decay of axisymmetric solutions of the wave equation on extreme Kerr backgrounds. *J. Funct. Anal.* **263** (2012), no. 9, 2770–2831.

[16] S. Aretakis, S. Czimek, and I. Rodnianski, The characteristic gluing problem for the Einstein equations and applications. 2021, arXiv:2107.02441.

[17] V. Balasubramanian, A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, Holographic thermalization, stability of Anti-de Sitter space, and the Fermi–Pasta–Ulam paradox. *Phys. Rev. Lett.* **113** (2014), no. 7, 071601.

[18] D. Baskin, A. Vasy, and J. Wunsch, Asymptotics of radiation fields in asymptotically Minkowski space. *Amer. J. Math.* **137** (2015), no. 5, 1293–1364.

[19] D. Baskin, A. Vasy, and J. Wunsch, Asymptotics of scalar waves on long-range asymptotically Minkowski spaces. *Adv. Math.* **328** (2018), 160–216.

[20] M. Belishev and Y. Kurylev, To the reconstruction of a Riemannian manifold via its spectral data (BC–Method). *Comm. Partial Differential Equations* **17** (1992), no. 5–6, 767–804.

[21] A. N. Bernal and M. Sánchez, Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes. *Comm. Math. Phys.* **257** (2005), no. 1, 43–50.

[22] N. Besset and D. Häfner, Existence of exponentially growing finite energy solutions for the charged Klein–Gordon equation on the de Sitter–Kerr–Newman metric. *J. Hyperbolic Differ. Equ.* **18** (2021), no. 02, 293–310.

[23] L. Bieri and N. Zipser, *Extensions of the stability theorem of the Minkowski space in general relativity*. AMS/IP Stud. Adv. Math. 45, American Mathematical Society, 2009.

[24] L. Bigorgne, D. Fajman, J. Joudioux, J. Smulevici, and M. Thaller, Asymptotic stability of Minkowski space-time with non-compactly supported massless Vlasov matter. *Arch. Ration. Mech. Anal.* **242** (2021), no. 1, 1–147.

[25] P. Bizoń and A. Rostworowski, Weakly turbulent instability of anti-de Sitter spacetime. *Phys. Rev. Lett.* **107** (2011), no. 3, 031102.

[26] P. Blue and A. Soffer, Semilinear wave equations on the Schwarzschild manifold. I. Local decay estimates. *Adv. Differential Equations* **8** (2003), no. 5, 595–614.

[27] P. Blue and A. Soffer, Errata for "Global existence and scattering for the nonlinear Schrodinger equation on Schwarzschild manifolds", "Semilinear wave equations on the Schwarzschild manifold I: Local Decay Estimates", and "The wave equation on the Schwarzschild metric II: Local Decay for the spin 2 Regge Wheeler equation". 2006, arXiv:gr-qc/0608073.

[28] J.-F. Bony and D. Häfner, Decay and non-decay of the local energy for the wave equation on the de Sitter–Schwarzschild metric. *Comm. Math. Phys.* **282** (2008), no. 3, 697–719.

[29] J.-F. Bony and D. Häfner, Low frequency resolvent estimates for long range perturbations of the Euclidean Laplacian. *Math. Res. Lett.* **17** (2010), no. 2, 303–308.

[30] D. R. Brill and R. W. Lindquist, Interaction energy in geometrostatics. *Phys. Rev.* **131** (1963), no. 1, 471.

[31] O. Brodbeck, S. Frittelli, P. Hübner, and O. A. Reula, Einstein's equations with asymptotically stable constraint propagation. *J. Math. Phys.* **40** (1999), no. 2, 909–923.

[32] G. A. Burnett, The high-frequency limit in general relativity. *J. Math. Phys.* **30** (1989), no. 1, 90–96.

[33] N. Burq, Décroissance de l'énergie locale de l'équation des ondes pour le problème extérieur et absence de résonance au voisinage du réel. *Acta Math.* **180** (1998), no. 1, 1–29.

[34] V. Cardoso, J. L. Costa, K. Destounis, P. Hintz, and A. Jansen, Quasinormal modes and strong cosmic censorship. *Phys. Rev. Lett.* **120** (2018), no. 3, 031103.

[35] A. Carlotto and R. Schoen, Localizing solutions of the Einstein constraint equations. *Invent. Math.* **205** (2016), no. 3, 559–615.

[36] B. Carter, Hamilton–Jacobi and Schrödinger separable solutions of Einstein's equations. *Comm. Math. Phys.* **10** (1968), no. 4, 280–310.

[37] M. Casals and P. Zimmerman, Perturbations of an extremal Kerr spacetime: analytic framework and late-time tails. *Phys. Rev. D* **100** (2019), no. 12.

[38] S. Chandrasekhar, *The mathematical theory of black holes*. Internat. Ser. Monogr. Phys. 69, The Clarendon Press, Oxford University Press, New York, 1992. Revised reprint of the 1983 original, Oxford Science Publications.

[39] Y. Choquet-Bruhat, Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires. *Acta Math.* **88** (1952), no. 1, 141–225.

[40] Y. Choquet-Bruhat and R. Geroch, Global aspects of the Cauchy problem in general relativity. *Comm. Math. Phys.* **14** (1969), no. 4, 329–335.

[41] D. Christodoulou, The formation of black holes and singularities in spherically symmetric gravitational collapse. *Comm. Pure Appl. Math.* **44** (1991), no. 3, 339–373.

[42] D. Christodoulou, Examples of naked singularity formation in the gravitational collapse of a scalar field. *Ann. of Math.* (1994), 607–653.

[43] D. Christodoulou, The instability of naked singularities in the gravitational collapse of a scalar field. *Ann. of Math. (2)* **149** (1999), no. 1, 183–217.

[44] D. Christodoulou, *The formation of black holes in general relativity*. European Mathematical Society, 2009.

[45] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space*. Princeton Math. Ser. 41, Princeton University Press, Princeton, NJ, 1993.

[46] P. T. Chruściel and E. Delay, On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications. *Mém. Soc. Math. Fr. (N.S.)* **94** (2003), vi+103 pp.

[47] J. Corvino, Scalar curvature deformation and a gluing construction for the Einstein constraint equations. *Comm. Math. Phys.* **214** (2000), no. 1, 137–189.

[48] J. Corvino and R. M. Schoen, On the asymptotics for the vacuum Einstein constraint equations. *J. Differential Geom.* **73** (2006), no. 2, 185–217.

[49] J. L. Costa and A. T. Franzen, Bounded energy waves on the black hole interior of Reissner–Nordström–de Sitter. 2016, arXiv:1607.01018.

[50] J. L. Costa, P. M. Girão, J. Natário, and J. Drumond Silva, On the global uniqueness for the Einstein–Maxwell–scalar field system with a cosmological constant. II. Structure of the solutions and stability of the Cauchy horizon. *Comm. Math. Phys.* **339** (2015), no. 3, 903–947.

[51] J. L. Costa, P. M. Girão, J. Natário, and J. Drumond Silva, On the global uniqueness for the Einstein–Maxwell–scalar field system with a cosmological constant: I. Well posedness and breakdown criterion. *Classical Quantum Gravity* **32** (2015), no. 1, 015017, 33 pp. .

[52] J. L. Costa, P. M. Girão, J. Natário, and J. Drumond Silva, On the global uniqueness for the Einstein–Maxwell–scalar field system with a cosmological constant: III. Mass inflation and extendibility of the solutions. *Ann. PDE* **3** (2017), no. 1, 8.

[53] B. Craps, O. Evnin, and J. Vanhoof, Renormalization group, secular term resummation and AdS (in)stability. *J. High Energy Phys.* **10** (2014), 048.

[54] M. Dafermos, The interior of charged black holes and the problem of uniqueness in general relativity. *Comm. Pure Appl. Math.* **58** (2005), no. 4, 445–504.

[55] M. Dafermos and G. Holzegel, Dynamic instability of solitons in $4+1$-dimensional gravity with negative cosmological constant (unpublished). 2006, https://www.dpmms.cam.ac.uk/~md384/ADSinstability.pdf.

[56] M. Dafermos, G. Holzegel, and I. Rodnianski, A scattering theory construction of dynamical vacuum black holes. 2013, arXiv:1306.5364.

[57] M. Dafermos, G. Holzegel, and I. Rodnianski, Boundedness and decay for the Teukolsky equation on Kerr spacetimes I: the case $|a| \ll M$. *Ann. PDE* **5** (2019), no. 1, 2.

[58] M. Dafermos, G. Holzegel, and I. Rodnianski, The linear stability of the Schwarzschild solution to gravitational perturbations. *Acta Math.* **222** (2019), 1–214.

[59] M. Dafermos, G. Holzegel, I. Rodnianski, and M. Taylor, The nonlinear stability of the Schwarzschild solution to gravitational perturbations. 2021, arXiv:2104.08222.

[60] M. Dafermos and J. Luk, The interior of dynamical vacuum black holes I: The C^0-stability of the Kerr Cauchy horizon. 2017, arXiv:1710.01722.

[61] M. Dafermos and I. Rodnianski, A proof of Price's law for the collapse of a self-gravitating scalar field. *Invent. Math.* **162** (2005), no. 2, 381–457.

[62] M. Dafermos and I. Rodnianski, The wave equation on Schwarzschild–de Sitter spacetimes. 2007, arXiv:0709.2766.

[63] M. Dafermos and I. Rodnianski, The red-shift effect and radiation decay on black hole spacetimes. *Comm. Pure Appl. Math.* **62** (2009), no. 7, 859–919.

[64] M. Dafermos and I. Rodnianski, Decay for solutions of the wave equation on Kerr exterior spacetimes I–II: the cases $|a| \ll M$ or axisymmetry. 2010, arXiv:1010.5132.

[65] M. Dafermos and I. Rodnianski, A new physical-space approach to decay for the wave equation with applications to black hole spacetimes. In *XVIth international congress on mathematical physics*, pp. 421–432, World Scientific, 2010.

[66] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case $|a| < M$. *Ann. of Math. (2)* **183** (2016), no. 3, 787–913.

[67] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, A scattering theory for the wave equation on Kerr black hole exteriors. *Ann. Sci. Éc. Norm. Supér. (4)* **51** (2018), no. 2, 371–486.

[68] E. Delay, Smooth compactly supported solutions of some underdetermined elliptic PDE, with gluing applications. *Comm. Partial Differential Equations* **37** (2012), no. 10, 1689–1716.

[69] O. J. C. Dias, G. T. Horowitz, and J. E. Santos, Gravitational turbulent instability of Anti-de Sitter space. *Classical Quantum Gravity* **29** (2012), 194002.

[70] D. Dold, Unstable mode solutions to the Klein–Gordon equation in Kerr–anti-de Sitter spacetimes. *Comm. Math. Phys.* **350** (2017), no. 2, 639–697.

[71] R. Donninger, W. Schlag, and A. Soffer, A proof of Price's law on Schwarzschild black hole manifolds for all angular momenta. *Adv. Math.* **226** (2011), no. 1, 484–540.

[72] J. J. Duistermaat and L. Hörmander, Fourier integral operators. II. *Acta Math.* **128** (1972), no. 1, 183–269.

[73] S. Dyatlov, Exponential energy decay for Kerr–de Sitter black holes beyond event horizons. *Math. Res. Lett.* **18** (2011), no. 5, 1023–1035.

[74] S. Dyatlov, Quasi-normal modes and exponential energy decay for the Kerr–de Sitter black hole. *Comm. Math. Phys.* **306** (2011), no. 1, 119–163.

[75] S. Dyatlov, Asymptotic distribution of quasi-normal modes for Kerr–de Sitter black holes. *Ann. Henri Poincaré* **13** (2012), no. 5, 1101–1166.

[76] S. Dyatlov, Asymptotics of linear waves and resonances with applications to black holes. *Comm. Math. Phys.* **335** (2015), no. 3, 1445–1485.

[77] S. Dyatlov, Spectral gaps for normally hyperbolic trapping. *Ann. Inst. Fourier (Grenoble)* **66** (2016), no. 1, 55–82.

[78] S. Dyatlov and M. Zworski, *Mathematical theory of scattering resonances*. Grad. Stud. Math. 200, American Mathematical Society, 2019.

[79] D. Fajman, J. Joudioux, and J. Smulevici, The stability of the Minkowski space for the Einstein–Vlasov system. 2017, arXiv:1707.06141.

[80] F. Finster and J. Smoller, Decay of solutions of the Teukolsky equation for higher spin in the Schwarzschild geometry. *Adv. Theor. Math. Phys.* **13** (2009), no. 1, 71–110.

[81] F. Finster and J. Smoller, Linear stability of the non-extreme Kerr black hole. 2016, arXiv:1606.08005.

[82] G. Fournodavlos and J. Luk, Asymptotically Kasner-like singularities. 2020, arXiv:2003.13591.

[83] G. Fournodavlos, I. Rodnianski, and J. Speck, Stable Big Bang formation for Einstein's equations: the complete sub-critical regime. 2020, arXiv:2012.05888.

[84] G. Fournodavlos and J. Sbierski, Generic blow-up results for the wave equation in the interior of a Schwarzschild black hole. *Arch. Ration. Mech. Anal.* **235** (2020), no. 2, 927–971.

[85] A. T. Franzen, Boundedness of massless scalar waves on Reissner–Nordström interior backgrounds. *Comm. Math. Phys.* **343** (2016), no. 2, 601–650.

[86] A. T. Franzen, Boundedness of massless scalar waves on kerr interior backgrounds. *Ann. Henri Poincaré* **21** (2020), 1045–1111.

[87] H. Friedrich, On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure. *Comm. Math. Phys.* **107** (1986), no. 4, 587–609.

[88] H. Friedrich, Einstein equations and conformal structure: existence of anti-de Sitter-type space-times. *J. Geom. Phys.* **17** (1995), no. 2, 125–184.

[89] J. Galkowski and M. Zworski, Analytic hypoellipticity of Keldysh operators. 2020, arXiv:2003.08106.

[90] O. Gannot, Quasinormal modes for Schwarzschild–AdS black holes: exponential convergence to the real axis. *Comm. Math. Phys.* **330** (2014), no. 2, 771–799.

[91] O. Gannot, Existence of quasinormal modes for Kerr–AdS black holes. *Ann. Henri Poincaré* **18** (2017), 2757–2788.

[92] E. Giorgi, The linear stability of Reissner–Nordström spacetime for small charge. *Ann. PDE* **6** (2020), no. 2, 1–145.

[93] C. R. Graham and J. M. Lee, Einstein metrics with prescribed conformal infinity on the ball. *Adv. Math.* **87** (1991), no. 2, 186–225.

[94] S. R. Green, A. Maillard, L. Lehner, and S. L. Liebling, Islands of stability and recurrence times in AdS. *Phys. Rev. D* **92** (2015), no. 8, 084001.

[95] C. Guillarmou and A. Hassell, Resolvent at low energy and Riesz transform for Schrödinger operators on asymptotically conic manifolds. I. *Math. Ann.* **341** (2008), no. 4, 859–896.

[96] C. Guillarmou and A. Hassell, Resolvent at low energy and Riesz transform for Schrödinger operators on asymptotically conic manifolds. II. *Ann. Inst. Fourier (Grenoble)* **59** (2009), no. 4, 1553–1610.

[97] C. Guillarmou, A. Hassell, and A. Sikora, Resolvent at low energy III: the spectral measure. *Trans. Amer. Math. Soc.* **365** (2013), no. 11, 6103–6148.

[98] C. Gundlach, G. Calabrese, I. Hinder, and J. M. Martín-García, Constraint damping in the Z4 formulation and harmonic gauge. *Classical Quantum Gravity* **22** (2005), no. 17, 3767.

[99] C. Gundlach, R. H. Price, and J. Pullin, Late-time behavior of stellar collapse and explosions. I. Linearized perturbations. *Phys. Rev. D* **49** (1994), no. 2, 883.

[100] D. Häfner, P. Hintz, and A. Vasy, Linear stability of slowly rotating Kerr black holes. *Invent. Math.* **223** (2021), 1227–1406.

[101] A. Hassell and A. Vasy, The resolvent for Laplace-type operators on asymptotically conic spaces. *Ann. Inst. Fourier* **51** (2001), 1299–1346.

[102] P. Hintz, Global analysis of quasilinear wave equations on asymptotically de Sitter spaces. *Ann. Inst. Fourier* **66** (2016), no. 4, 1285–1408.

[103] P. Hintz, Boundedness and decay of scalar waves at the Cauchy horizon of the Kerr spacetime. *Comment. Math. Helv.* **92** (2017), no. 4, 801–837.

[104] P. Hintz, Resonance expansions for tensor-valued waves on asymptotically Kerr–de Sitter spaces. *J. Spectr. Theory* **7** (2017), 519–557.

[105] P. Hintz, Non-linear stability of the Kerr–Newman–de Sitter family of charged black holes. *Ann. PDE* **4** (2018), no. 1, 11.

[106] P. Hintz, A sharp version of Price’s law for wave decay on asymptotically flat spacetimes. 2020, arXiv:2004.01664.

[107] P. Hintz, Black hole gluing in de Sitter space. *Comm. Partial Differential Equations* **46** (2021), no. 7, 1280–1318.

[108] P. Hintz, Normally hyperbolic trapping on asymptotically stationary spacetimes. *Prob. Math. Phys.* **2** (2021), no. 1, 71–126.

[109] P. Hintz and G. Uhlmann, Reconstruction of Lorentzian manifolds from boundary light observation sets. *Int. Math. Res. Not.* (2018), rnx320.

[110] P. Hintz and A. Vasy, Non-trapping estimates near normally hyperbolic trapping. *Math. Res. Lett.* **21** (2014), no. 6, 1277–1304.

[111] P. Hintz and A. Vasy, Semilinear wave equations on asymptotically de Sitter, Kerr–de Sitter and Minkowski spacetimes. *Anal. PDE* **8** (2015), no. 8, 1807–1890.

[112] P. Hintz and A. Vasy, Global analysis of quasilinear wave equations on asymptotically Kerr–de Sitter spaces. *Int. Math. Res. Not.* **2016** (2016), no. 17, 5355–5426.

[113] P. Hintz and A. Vasy, Analysis of linear waves near the Cauchy horizon of cosmological black holes. *J. Math. Phys.* **58** (2017), no. 8, 081509.

[114] P. Hintz and A. Vasy, Asymptotics for the wave equation on differential forms on Kerr–de Sitter space. *J. Differential Geom.* **110** (2018), no. 2, 221–279.

[115] P. Hintz and A. Vasy, The global non-linear stability of the Kerr–de Sitter family of black holes. *Acta Math.* **220** (2018), 1–206.

[116] P. Hintz and A. Vasy, Stability of Minkowski space and polyhomogeneity of the metric. *Ann. PDE* **6** (2020), no. 2.

[117] P. Hintz and Y. Xie, Quasinormal modes of small Schwarzschild–de Sitter black holes. Preprint, 2020.

[118] M. W. Hirsch, C. C. Pugh, and M. Shub, *Invariant manifolds*. Lecture Notes in Math. 583, Springer, Berlin–New York, 1977.

[119] G. Holzegel, J. Luk, J. Smulevici, and C. Warnick, Asymptotic properties of linear field equations in anti-de Sitter space. *Comm. Math. Phys.* **374** (2020), no. 2, 1125–1178.

[120] G. Holzegel and J. Smulevici, Decay properties of Klein–Gordon fields on Kerr–AdS spacetimes. *Comm. Pure Appl. Math.* **66** (2013), no. 11, 1751–1802.

[121] G. Holzegel and J. Smulevici, Quasimodes and a lower bound on the uniform energy decay rate for Kerr–AdS spacetimes. *Anal. PDE* **7** (2014), no. 5, 1057–1090.

[122] G. T. Horowitz and J. E. Santos, Geons and the instability of Anti-de Sitter spacetime. *Surv. Differ. Geom.* **20** (2015), 321–335.

[123] C. Huneau and J. Luk, Trilinear compensated compactness and Burnett's conjecture in general relativity. 2019, arXiv:1907.10743.

[124] P.-K. Hung, The linear stability of the Schwarzschild spacetime in the harmonic gauge: odd part. 2018, arXiv:1803.03881.

[125] P.-K. Hung, The linear stability of the Schwarzschild spacetime in the harmonic gauge: even part. 2019, arXiv:1909.06733.

[126] P.-K. Hung, J. Keller, and M.-T. Wang, Linear stability of Schwarzschild spacetime: decay of metric coefficients. 2017, arXiv:1702.02843v3.

[127] A. Iantchenko, Quasi-normal modes for massless Dirac fields in Kerr–Newman–de Sitter black holes. 2015, arXiv:1511.09233.

[128] J. Isenberg, D. Maxwell, and D. Pollack, A gluing construction for non-vacuum solutions of the einstein-constraint equations. *Adv. Theor. Math. Phys.* **9** (2005), no. 1, 129–172.

[129] J. Isenberg, R. Mazzeo, and D. Pollack, Gluing and wormholes for the Einstein constraint equations. *Comm. Math. Phys.* **231** (2002), no. 3, 529–568.

[130] E. Y. Jaffe, Asymptotic description of the formation of black holes from short-pulse data. 2020, arXiv:2003.05985.

[131] T. Johnson, The linear stability of the Schwarzschild solution to gravitational perturbations in the generalised wave gauge. *Ann. PDE* **5** (2019), no. 2, 13.

[132] D. Kastor and J. Traschen, Cosmological multi-black-hole solutions. *Phys. Rev. D* **47** (1993), no. 12, 5370.

[133] B. S. Kay and R. M. Wald, Linear stability of Schwarzschild under perturbations which are non-vanishing on the bifurcation 2-sphere. *Classical Quantum Gravity* **4** (1987), no. 4, 893.

[134] C. Kehle, Diophantine approximation as Cosmic Censor for Kerr–AdS black holes. 2020, arXiv:2007.12614.

[135] C. Kehle, Uniform boundedness and continuity at the cauchy horizon for linear waves on Reissner–Nordström–AdS black holes. *Comm. Math. Phys.* **376** (2020), no. 1, 145–200.

[136] C. Kehle, Blowup of the local energy of linear waves at the Reissner–Nordström-AdS Cauchy horizon. *Classical Quantum Gravity* **38** (2021), 21.

[137] S. Klainerman and J. Szeftel, Constructions of GCM spheres in perturbations of Kerr. 2019, arXiv:1911.00697.

[138] S. Klainerman and J. Szeftel, Effective results on uniformization and intrinsic GCM spheres in perturbations of Kerr. 2019, arXiv:1912.12195.

[139] S. Klainerman and J. Szeftel, *Global nonlinear stability of Schwarzschild space-time under polarized perturbations: (AMS-210)*. Princeton University Press, 2021.

[140] S. Klainerman and J. Szeftel, Kerr stability for small angular momentum. 2021, arXiv:2104.11857.

[141] H. Kodama and A. Ishibashi, A master equation for gravitational perturbations of maximally symmetric black holes in higher dimensions. *Progr. Theoret. Phys.* **110** (2003), no. 4, 701–722.

[142] Y. Kurylev, M. Lassas, L. Oksanen, and G. Uhlmann, Inverse problem for Einstein-scalar field equations. 2014, arXiv:1406.4776.

[143] Y. Kurylev, M. Lassas, and G. Uhlmann, Inverse problems for Lorentzian manifolds and non-linear hyperbolic equations. 2014, arXiv:1405.3386.

[144] M. Lassas, G. Uhlmann, and Y. Wang, Determination of vacuum space-times from the Einstein–Maxwell equations. 2017, arXiv:1703.10704.

[145] H. Lindblad, On the asymptotic behavior of solutions to the Einstein vacuum equations in wave coordinates. *Comm. Math. Phys.* **353** (2017), no. 1, 135–184.

[146] H. Lindblad and I. Rodnianski, The global stability of Minkowski space-time in harmonic gauge. *Ann. of Math. (2)* **171** (2010), no. 3, 1401–1477.

[147] H. Lindblad and M. Taylor, Global stability of Minkowski space for the Einstein–Vlasov system in the harmonic gauge. 2017, arXiv:1707.06079.

[148] O. Lindblad Petersen and A. Vasy, Analyticity of quasinormal modes in the Kerr and Kerr–de Sitter spacetimes. 2021, arXiv:2104.04500.

[149] J. Luk, On the local existence for the characteristic initial value problem in general relativity. *Int. Math. Res. Not.* **20** (2012), 4625–4678.

[150] J. Luk and S.-J. Oh, Proof of linear instability of the Reissner–Nordström Cauchy horizon under scalar perturbations. *Duke Math. J.* **166** (2017), no. 3, 437–493.

[151] J. Luk and S.-J. Oh, Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data I. The interior of the black hole region. *Ann. of Math.* **190** (2019), no. 1, 1–111.

[152] J. Luk and S.-J. Oh, Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data II: the exterior of the black hole region. *Ann. PDE* **5** (2019), no. 1, 6.

[153] J. Luk and I. Rodnianski, Local propagation of impulsive gravitational waves. *Comm. Pure Appl. Math.* **68** (2015), no. 4, 511–624.

[154] J. Luk and I. Rodnianski, Nonlinear interaction of impulsive gravitational waves for the vacuum Einstein equations. *Cambridge J. Math.* **5** (2017), no. 4, 435–570.

[155] J. Luk and I. Rodnianski, High-frequency limits and null dust shell solutions in general relativity. 2020, arXiv:2009.08968.

[156] J. Luk and J. Sbierski, Instability results for the wave equation in the interior of Kerr black holes. *J. Funct. Anal.* **271** (2016), no. 7, 1948–1995.

[157] S. Ma, Uniform energy bound and Morawetz estimate for extreme components of spin fields in the exterior of a slowly rotating Kerr black hole II: linearized gravity. 2017, arXiv:1708.07385.

[158] S. D. Majumdar, A class of exact solutions of Einstein's field equations. *Phys. Rev.* **72** (1947), no. 5, 390.

[159] J.-A. Marck, Parallel-tetrad on null geodesics in Kerr–Newman space-time. *Phys. Lett. A* **97** (1983), no. 4, 140–142.

[160] R. R. Mazzeo and R. B. Melrose, Meromorphic extension of the resolvent on complete spaces with asymptotically constant negative curvature. *J. Funct. Anal.* **75** (1987), no. 2, 260–310.

[161] R. B. Melrose, *The Atiyah–Patodi–Singer index theorem*. Res. Notes Math. 4, A K Peters, Ltd., Wellesley, MA, 1993.

[162] R. B. Melrose, Spectral and scattering theory for the Laplacian on asymptotically Euclidian spaces. In *Spectral and scattering theory (Sanda, 1992)*, pp. 85–130, Lect. Notes Pure Appl. Math. 161, Dekker, New York, 1994.

[163] R. B. Melrose, A. Sá Barreto, and A. Vasy, Analytic continuation and semiclassical resolvent estimates on asymptotically hyperbolic spaces. *Comm. Partial Differential Equations* **39** (2014), no. 3, 452–511.

[164] R. B. Melrose, A. Sá Barreto, and A. Vasy, Asymptotics of solutions of the wave equation on de Sitter–Schwarzschild space. *Comm. Partial Differential Equations* **39** (2014), no. 3, 512–529.

[165] J. Metcalfe, J. Sterbenz, and D. Tataru, Local energy decay for scalar fields on time dependent non-trapping backgrounds. *Amer. J. Math.* **142** (2020), no. 3.

[166] V. Moncrief, Gravitational perturbations of spherically symmetric systems. I. The exterior problem. *Ann. Physics* **88** (1974), no. 2, 323–342.

[167] K. Morgan and J. Wunsch, Generalized Price's law on fractional-order asymptotically flat stationary spacetimes. 2021, arXiv:2105.02305.

[168] G. Moschidis, The r^p-weighted energy method of Dafermos and Rodnianski in general asymptotically flat spacetimes and applications. *Ann. PDE* **2** (2016), no. 1, 1–194.

[169] G. Moschidis, A proof of the instability of AdS for the Einstein–massless Vlasov system. 2018, arXiv:1812.04268.

[170] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, Metric of a rotating, charged mass. *J. Math. Phys.* **6** (1965), no. 6, 918–919.

[171] J. R. Oppenheimer and H. Snyder, On continued gravitational contraction. *Phys. Rev.* **56** (1939), no. 5, 455.

[172] A. Papapetrou, A static solution of the equations of the gravitational field for an arbitary charge-distribution. *Proc. R. Ir. Acad., A Math. Phys. Sci.* **51** (1945), 191–204.

[173] F. Pretorius, Evolution of binary black-hole spacetimes. *Phys. Rev. Lett.* **95** (2005), 121101.

[174] F. Pretorius and W. Israel, Quasi-spherical light cones of the Kerr geometry. *Classical Quantum Gravity* **15** (1998), no. 8, 2289.

[175] R. H. Price, Nonspherical perturbations of relativistic gravitational collapse. I. Scalar and gravitational perturbations. *Phys. Rev. D* **5** (1972), no. 10, 2419.

[176] R. H. Price and L. M. Burko, Late time tails from momentarily stationary, compact initial data in Schwarzschild spacetimes. *Phys. Rev. D* **70** (2004), no. 8, 084039.

[177] T. Regge and J. A. Wheeler, Stability of a Schwarzschild singularity. *Phys. Rev.* **108** (1957), 1063–1069.

[178] A. D. Rendall, Reduction of the characteristic initial value problem to the Cauchy problem and its applications to the Einstein equations. *Proc. R. Soc. Lond. Ser. A, Math. Phys. Sci.* **427** (1990), no. 1872, 221–239.

[179] H. Ringström, Future stability of the Einstein–non-linear scalar field system. *Invent. Math.* **173** (2008), no. 1, 123–208.

[180] H. Ringström, *The Cauchy problem in general relativity*. ESI Lect. Math. Phys. 6, European Mathematical Society, 2009.

[181] H. Ringström, On the geometry of silent and anisotropic big bang singularities. 2021, arXiv:2101.04955.

[182] I. Rodnianski and Y. Shlapentokh-Rothman, The asymptotically self-similar regime for the Einstein vacuum equations. *Geom. Funct. Anal.* **28** (2018), no. 3, 755–878.

[183] I. Rodnianski and Y. Shlapentokh-Rothman, Naked singularities for the Einstein vacuum equations: the exterior solution. 2019, arXiv:1912.08478.

[184] A. Rostworowski, Higher order perturbations of anti-de Sitter space and time-periodic solutions of vacuum Einstein equations. *Phys. Rev. D* **95** (2017), no. 12, 124043.

[185] X. Saint-Raymond, A simple Nash–Moser implicit function theorem. *Enseign. Math. (2)* **35** (1989), no. 3–4, 217–226.

[186] J. Sbierski, Characterisation of the energy of Gaussian beams on Lorentzian manifolds: with applications to black hole spacetimes. *Anal. PDE* **8** (2015), no. 6, 1379–1420.

[187] J. Sbierski, On the existence of a maximal Cauchy development for the Einstein equations: a dezornification. *Ann. Henri Poincaré* **17** (2016), no. 2, 301–329.

[188] J. Sbierski, The C_0-inextendibility of the Schwarzschild spacetime and the spacelike diameter in Lorentzian geometry. *J. Differential Geom.* **108** (2018), no. 2, 319–378.

[189] J. Sbierski, On holonomy singularities in general relativity and the $C^{0,1}_{\mathrm{loc}}$ inextendibility of spacetimes. 2020, arXiv:2007.12049.

[190] V. Schlue, Global results for linear waves on expanding Kerr and Schwarzschild de Sitter cosmologies. *Comm. Math. Phys.* **334** (2015), no. 2, 977–1023.

[191] V. Schlue, Decay of the Weyl curvature in expanding black hole cosmologies. 2016, arXiv:1610.04172.

[192] Y. Shlapentokh-Rothman, Exponentially growing finite energy solutions for the Klein–Gordon equation on sub-extremal Kerr spacetimes. *Comm. Math. Phys.* **329** (2014), no. 3, 859–891.

[193] Y. Shlapentokh-Rothman, Quantitative mode stability for the wave equation on the Kerr spacetime. *Ann. Henri Poincaré* **16** (2015), no. 1, 289–345.

[194] Y. Shlapentokh-Rothman and R. Teixeira da Costa, Boundedness and decay for the Teukolsky equation on Kerr in the full subextremal range $|a| < M$: frequency space analysis. 2020, arXiv:2007.07211.

[195] M. Simpson and R. Penrose, Internal instability in a Reissner–Nordström black hole. *Internat. J. Theoret. Phys.* **7** (1973), no. 3, 183–197.

[196] D. Tataru, Unique continuation for operators with partially analytic coefficients. *J. Math. Pures Appl.* **78** (1999), no. 5, 505–521.

[197] D. Tataru, Local decay of waves on asymptotically flat stationary space-times. *Amer. J. Math.* **135** (2013), no. 2, 361–401.

[198] D. Tataru and M. Tohaneanu, A local energy estimate on Kerr black hole backgrounds. *Int. Math. Res. Not.* **2011** (2011), no. 2, 248–292.

[199] M. Taylor, The global nonlinear stability of Minkowski space for the massless Einstein–Vlasov system. *Ann. PDE* **3** (2017), no. 1, 9.

[200] S. A. Teukolsky, Perturbations of a rotating black hole. I. Fundamental equations for gravitational, electromagnetic, and neutrino-field perturbations. *Astrophys. J.* **185** (1973), 635–648.

[201] A. Vasy, Microlocal analysis of asymptotically hyperbolic and Kerr–de Sitter spaces (with an appendix by Semyon Dyatlov). *Invent. Math.* **194** (2013), no. 2, 381–513.

[202] A. Vasy, Resolvent near zero energy on Riemannian scattering (asymptotically conic) spaces. 2018, arXiv:1808.06123.

[203] A. Vasy, Limiting absorption principle on Riemannian scattering (asymptotically conic) spaces, a Lagrangian approach. *Comm. Partial Differential Equations* (to appear).

[204] A. Vasy, Resolvent near zero energy on Riemannian scattering (asymptotically conic) spaces, a Lagrangian approach. *Comm. Partial Differential Equations* (to appear).

[205] A. Vasy and M. Zworski, Semiclassical estimates in asymptotically Euclidean scattering. *Comm. Math. Phys.* **212** (2000), no. 1, 205–217.

[206] C. V. Vishveshwara, Stability of the Schwarzschild metric. *Phys. Rev. D* **1** (1970), 2870–2879.

[207] R. M. Wald, Note on the stability of the Schwarzschild metric. *J. Math. Phys.* **20** (1979), no. 6, 1056–1058.

[208] C. M. Warnick, On quasinormal modes of asymptotically anti-de Sitter black holes. *Comm. Math. Phys.* **333** (2015), no. 2, 959–1035.

[209] B. F. Whiting, Mode stability of the Kerr black hole. *J. Math. Phys.* **30** (1989), no. 6, 1301–1305.

[210] J. Wunsch and M. Zworski, Resolvent estimates for normally hyperbolic trapped sets. *Ann. Henri Poincaré* **12** (2011), no. 7, 1349–1385.

[211] F. J. Zerilli, Effective potential for even-parity Regge–Wheeler gravitational perturbation equations. *Phys. Rev. Lett.* **24** (1970), 737–738.

PETER HINTZ

ETH Zürich, Departement Mathematik, Rämistrasse 101, 8092 Zürich, Switzerland, peter.hintz@math.ethz.edu

GUSTAV HOLZEGEL

Universität Münster, Mathematisches Institut, Einsteinstrasse 62, 48149 Münster, Germany, gholzege@uni-muenster.de

11. MATHEMATICAL PHYSICS

SPIN SYSTEMS WITH HYPERBOLIC SYMMETRY: A SURVEY

ROLAND BAUERSCHMIDT AND TYLER HELMUTH

ABSTRACT

Spin systems with hyperbolic symmetry originated as simplified models for the Anderson metal–insulator transition, and were subsequently found to exactly describe probabilistic models of linearly reinforced walks and random forests. In this survey we introduce these models, discuss their origins and main features, some existing tools available for their study, recent probabilistic results, and relations to other well-studied probabilistic models. Along the way we discuss some of the many open questions that remain.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 82B20; Secondary 82B28, 60K35

KEYWORDS

Spin systems, reinforced random walks, random forests, hyperbolic symmetry, Anderson transition

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 3986–4008
DOI 10.4171/ICM2022/20

1. INTRODUCTION

Classical spin systems with spherical symmetry, such as the Ising and classical Heisenberg models, are basic models for magnetism and have been studied extensively over the last century. It is well-understood that the associated symmetry groups play an important role, particularly for the critical and low-temperature behaviour of these models. For example, the discrete $\mathbb{Z}_2$ symmetry of the Ising model is spontaneously broken at low temperatures, and in this phase truncated correlations decay exponentially. For models with continuous $O(n)$ symmetries, $n \geq 2$, low temperature truncated correlations instead decay polynomially, a reflection of the fact that the symmetry is spontaneously broken to an $O(n-1)$ symmetry.

Spin systems with hyperbolic symmetry groups are also studied in condensed matter physics, primarily because of their relevance for the Anderson delocalisation–localisation (metal–insulator) transition of random Schrödinger operators and related random matrix models [38, 69, 74]. A rigorous analysis of the Anderson transition remains an outstanding challenge; see Section 3. The essential physical phenomena of the Anderson transition are expected to be captured by the more tractable $\mathbb{H}^{2|2}$ model, a simplified spin system with hyperbolic symmetry [76]. Surprisingly, the $\mathbb{H}^{2|2}$ model and its natural generalisations are intimately connected to probabilistic lattice models. The $\mathbb{H}^2$ and $\mathbb{H}^{2|2}$ models, motivated by the Anderson transition [34,72], are exactly related to (linearly) edge-reinforced random walks and vertex-reinforced jump processes, introduced independently in the probability literature in the 1980s [29] and early 2000s [26]. A similar connection exists between the related $\mathbb{H}^{0|2}$ model and random forests [11, 22]; random forests arose earlier (for example) in connection with the Fortuin–Kasteleyn random cluster model [43]. The connections between hyperbolic spin systems and probabilistic phenomena are the main topic of this survey.

More specifically, this survey focuses on probabilistic results in line with the original physical motivation for studying hyperbolic spin systems. In particular, we focus on results for $\mathbb{Z}^d$ (and its finite approximations) for $d \geq 2$. Our perspective is that a central role is played by the continuous symmetry groups of the spin systems. There are other perspectives available, notably that of Bayesian statistics. While the latter perspective has played a role in important results, e.g. [4, 66], and has found use in statistical contexts [5, 6, 31], we will not mention it further. Similarly, there are many related works we cannot discuss; fortunately, many of these are discussed in recent surveys on closely related topics [49, 61, 70, 71].

To set the stage, the remainder of this introduction recalls the *magic formula* for edge-reinforced random walk that led to the discovery of the connections discussed in this survey. Readers familiar with the magic formula may wish to jump to Section 2, where we introduce hyperbolic spin systems, or to Section 3, which discusses the physical background. The probabilistic representations and results for reinforced random walks and random forests are discussed in Sections 4 and 5, respectively, along with questions for the future.

Magic formula for edge-reinforced random walk. Fix $\alpha > 0$, a graph $G = (\Lambda, E)$, and an initial vertex $0 \in \Lambda$. *Edge-reinforced random walk (ERRW)* with $X_0 = 0$ and initial weights α

is the stochastic process $(X_n)_{n\geq 0}$ with transitions

$$\mathbb{P}_0^{\mathrm{ERRW}(\alpha)}\big[X_{n+1}=j\,|\,(X_m)_{m\leq n}, X_n=i\big]=\frac{(\alpha+L_n^{ij})1_{ij\in E}}{\sum_{k:ik\in E}(\alpha+L_n^{ik})}, \tag{1.1}$$

where L_n^{ij} is the number of times the edge ij has been crossed up to time n (in either direction). The transition rates change rapidly if α is small, and hence this is called the *strong reinforcement* regime. *Weak reinforcement* refers to α being large. The definition can be generalised to edge-dependent weights $\alpha=(\alpha_{ij})$ in a straightforward manner.

Some intuition about ERRW can be gained by considering the case when G is a path on three vertices. Call one edge blue, one edge red, and start an ERRW at the middle vertex. If $\alpha=2$ then the law of the vector $\frac{1}{2}L_{2n}$ of (half of) the number of crossings of the edges at time $2n$ is the law of a *Pólya urn*. Pólya's urn is the process that starts with an urn containing one red and one blue ball, and then sequentially draws a ball and replaces it with two balls, both of the same colour as the drawn ball. The fundamental fact about Pólya's urn is that $\frac{1}{2n}L_{2n}$ converges to $(U,1-U)$ where U is a uniform random variable on $[0,1]$, i.e. the fraction of crossings of the blue edge is uniform. This can be proven by induction. Note that for an ordinary simple random walk this limit would be deterministic. *A priori* it is hard to predict how ERRW behaves on more complicated graphs. For example, is ERRW transient if simple random walk is transient? Does the answer depend on α?

It turns out that the connection to Pólya's urn has a far reaching generalisation. The theory of partial exchangeability guarantees that ERRW is a random walk in random environment [30]. A consequence is that $\frac{1}{n}L_n$ has a distributional limit: it is the law of the random environment. Coppersmith and Diaconis discovered that one can give an explicit formula for the limiting law on *any* finite graph. It is surprising that an explicit formula can be obtained; this explains why it has been termed the *magic formula*, see [47, 55].

To precisely formulate this result, recall that an *environment* is a set of conductances $C\colon E\to[0,\infty)$ with $\sum_{ij}C_{ij}=1$. Write C_{ij} for the conductance of the edge $\{i,j\}$ and $C_{ii}=-C_i=-\sum_j C_{ij}$. Associated to C is a reversible Markov chain (simple random walk) with transition probabilities C_{ij}/C_i whose law we denote by $\mathbb{P}_0^{\mathrm{SRW}(C)}$ when started from 0.

Theorem 1.1 (Magic formula for ERRW). *Let $G=(\Lambda,E)$ be finite. Edge-reinforced random walk with $X_0=0$ and initial weights $\alpha=(\alpha_{ij})$ is a random walk in random environment:*

$$\mathbb{P}_0^{\mathrm{ERRW}(\alpha)}[\cdot]=\int\mathbb{P}_0^{\mathrm{SRW}(C)}[\cdot]\,d\mu_\alpha(C). \tag{1.2}$$

The environment μ_α has density proportional to

$$C_0^{\frac{1}{2}}\frac{\prod_{ij\in E}C_{ij}^{\alpha_{ij}-1}}{\prod_{i\in\Lambda}C_i^{\frac{1}{2}(\alpha_i+1)}}\sqrt{\det{}^0 C} \tag{1.3}$$

with respect to Lebesgue measure on the unit simplex in $[0,\infty)^E$, where $\alpha_i=\sum_j\alpha_{ij}$ and $\det^0 C$ is the determinant of any principal cofactor of C.

Note that the matrix-tree theorem implies the determinant in (1.3) can be written as a weighted sum of spanning trees, reflecting that this term is non-local.

Sabot and Tarrès showed how to relate the density (1.3) to the $\mathbb{H}^{2|2}$ model that we will introduce in the next section. This enabled them to leverage powerful results of Disertori, Spencer, and Zirnbauer to establish the existence of a recurrence/transience phase transition for ERRW on $\mathbb{Z}^d$ for $d \geq 3$, see Section 4. In Section 5 we show that connection probabilities in the arboreal gas, a stochastic-geometric model of random forests, can be written in a form very similar to the magic formula. The derivation of this connection probability formula was inspired by [12, 13], which (at least partially) revealed the inner workings of the magic formula: horospherical coordinates (hyperbolic symmetry) and supersymmetric localisation.

2. HYPERBOLIC SPIN SYSTEMS

This section introduces the hyperbolic spin systems that we will discuss, briefly explains their characteristic symmetries, and discusses how these symmetries manifest themselves if spontaneous symmetry breaking occurs. For precise definitions of the Grassmann and Berezin integrals that are used see, e.g. [13, APPENDIX A].

The $\mathbb{H}^{2|0}$ model. The $\mathbb{H}^2 = \mathbb{H}^{2|0}$ model is defined as follows. We consider the hyperbolic plane $\mathbb{H}^2$ realised as $\mathbb{H}^2 = \{u = (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 - z^2 = -1, z > 0\}$ and equipped with the Minkowski inner product $u \cdot u' = xx' + yy' - zz'$. For a finite graph $G = (\Lambda, E)$, we consider one spin $u_i \in \mathbb{H}^2$ per vertex $i \in \Lambda$ and define the action

$$H_{\beta,h}(u) = \frac{\beta}{2} \sum_{ij \in E} (u_i - u_j) \cdot (u_i - u_j) + h \sum_{i \in \Lambda} z_i. \tag{2.1}$$

The action also has a straightforward generalisation to edge- and vertex-dependent weights $\beta = (\beta_{ij})$ and $h = (h_i)$, and we will sometimes consider this case. For $\beta > 0$ and $h = 0$, the minimisers of $H_{\beta,0}$ are constant configurations $u_i = u_j$ for all $i, j \in \Lambda$. For $h > 0$, the unique minimiser is $u_i = (0, 0, 1)$ for all i. The $\mathbb{H}^2$ model is the probability measure on spin configurations whose expectation is given, for bounded $F: (\mathbb{H}^2)^\Lambda \to \mathbb{R}$, by

$$\langle F \rangle_{\beta,h} = \frac{1}{Z_{\beta,h}} \int_{(\mathbb{H}^2)^\Lambda} \prod_{i \in \Lambda} du_i \; F(u) e^{-H_{\beta,h}(u)} \tag{2.2}$$

where du_i stands for the Haar measure on $\mathbb{H}^2$ and $Z_{\beta,h}$ is a normalisation. Parametrising $u_i \in \mathbb{H}^2$ by $(x_i, y_i) \in \mathbb{R}^2$ with $z_i = \sqrt{1 + x_i^2 + y_i^2}$, we can explicitly rewrite (2.2) as

$$\langle F \rangle_{\beta,h} = \frac{1}{Z_{\beta,h}} \int_{(\mathbb{R}^2)^\Lambda} \prod_{i \in \Lambda} \frac{dx_i \, dy_i}{z_i} F(u) e^{-H_{\beta,h}(u)}. \tag{2.3}$$

The expectation is only normalisable if $h > 0$ (or, more generally, $h_i > 0$ for some vertex i) due to the non-compactness of $\mathbb{H}^2$. It is useful to construct a version with $h = 0$ in which the field is fixed (pinned) at some distinguished vertex 0. We denote the *pinned expectation* with pinning $u_0 = (0, 0, 1)$ by $\langle \cdot \rangle_\beta^0$.

The $\mathbb{H}^{0|2}$ model. Now we consider the Grassmann algebra Ω_Λ generated by two generators ξ_i and η_i per vertex $i \in \Lambda$ and set

$$z_i = \sqrt{1 - 2\xi_i \eta_i} = 1 - \xi_i \eta_i, \tag{2.4}$$

and unite these into the formal supervector $u_i = (\xi_i, \eta_i, z_i)$. Thus u_i has two odd (anticommuting) components ξ_i and η_i and one even (commuting) component z_i. We define $u_i \cdot u_j = -\xi_i \eta_j - \xi_j \eta_i - z_i z_j$, which is again an element of Ω_Λ. These definitions are such that $u_i \cdot u_i = -1$, as in the case of $\mathbb{H}^2$ spins. Define

$$H_{\beta,h} = \frac{\beta}{2} \sum_{ij \in E} (u_i - u_j) \cdot (u_i - u_j) + h \sum_{i \in \Lambda} z_i. \tag{2.5}$$

For F a polynomial in the ξ_i and η_i set

$$\langle F \rangle_{\beta,h} = \frac{1}{Z_{\beta,h}} \int \left(\prod_{i \in \Lambda} \partial_{\eta_i} \partial_{\xi_i} \frac{1}{z_i} \right) F e^{-H_{\beta,h}}, \tag{2.6}$$

where $\int \prod_{i \in \Lambda} \partial_{\eta_i} \partial_{\xi_i}$ stands for the Grassmann integral, i.e. the top coefficient of the element of the Grassmann algebra to its right. For example,

$$\int \partial_\xi \partial_\eta e^{-\xi\eta} = \int \partial_\xi \partial_\eta (1 - \xi\eta) = \int \partial_\xi \partial_\eta \eta\xi = 1. \tag{2.7}$$

In (2.6) and (2.7) we have used the convention that smooth functions of commuting elements of the algebra are defined by Taylor expansion. By nilpotency the expansion is finite, i.e. a polynomial. The $\mathbb{H}^{0|2}$ model is the expectation (2.6); while this is not a probabilistic expectation, we will soon see that it often carries probabilistic interpretations. Generalisations to edge- and vertex-dependent weights and pinning are straightforward.

The $\mathbb{H}^{2|2}$ model. The $\mathbb{H}^{2|2}$ model is defined as the $\mathbb{H}^{0|2}$ model was, but now beginning with three commuting components x_i, y_i, z_i. Formally, this means the real coefficients of the Grassmann algebra Ω_Λ of the previous section are replaced by smooth functions of x_i and y_i. To each vertex i we associate a formal supervector $u_i = (x_i, y_i, \xi_i, \eta_i, z_i)$, where x_i and y_i are commuting, ξ_i and η_i are generators of a Grassmann algebra, and

$$z_i = \sqrt{1 + x_i^2 + y_i^2 - 2\xi_i \eta_i} = \sqrt{1 + x_i^2 + y_i^2} - \frac{\xi_i \eta_i}{\sqrt{1 + x_i^2 + y_i^2}}. \tag{2.8}$$

As for $\mathbb{H}^{0|2}$, smooth functions of commuting elements of this algebra are defined by Taylor expansion, with the expansion now performed about $(x_i, y_i) \in \mathbb{R}^{2\Lambda}$; the second equality of (2.8) is an example.

The definition (2.8) ensures that z_i has positive degree zero part, and that $u_i \cdot u_i = -1$ for the super inner product $u_i \cdot u_j = x_i x_j + y_i y_j - \xi_i \eta_j - \xi_j \eta_i - z_i z_j$. As previously, we define

$$H_{\beta,h}(u) = \frac{\beta}{2} \sum_{ij \in E} (u_i - u_j) \cdot (u_i - u_j) + h \sum_{i \in \Lambda} z_i, \tag{2.9}$$

and the associated expectation

$$\langle F\rangle_{\beta,h} = \frac{1}{(2\pi)^{|\Lambda|}} \int \Bigg(\prod_{i\in\Lambda} dx_i\, dy_i\, \partial_{\eta_i} \partial_{\xi_i} \frac{1}{z_i} \Bigg) F e^{-H_{\beta,h}}. \tag{2.10}$$

This integral combines ordinary integration and Grassmann integration and is an instance of the Berezin integral, sometimes called a superintegral [15]. One computes the Grassmann integral to obtain the top coefficient of the element of the Grassmann algebra; this is a smooth function on $\mathbb{R}^{2\Lambda}$. One then computes the ordinary Lebesgue integral of this function. Again the generalisation to edge- and vertex-dependent weights and pinning is straightforward.

Note that (2.10) does not have a normalising factor as in the definitions of the $\mathbb{H}^2$ and $\mathbb{H}^{0|2}$ models, aside from the factor $(2\pi)^{-|\Lambda|}$ that does not depend on the weights. Nonetheless, the expectation is normalised: $\langle 1\rangle_{\beta,h} = 1$ if $h > 0$. This is due to an internal supersymmetry in the model, which implies $Z_{\beta,h} = (2\pi)^{|\Lambda|}$. More generally, this supersymmetry implies a powerful localisation principle first used in this context in [34].

Theorem 2.1 (SUSY localisation for $\mathbb{H}^{2|2}$). *For $F\colon \mathbb{R}^\Lambda \times \mathbb{R}^{\Lambda\times\Lambda} \to \mathbb{R}$ smooth and with sufficient decay, and for all edge- and vertex-dependent weights $\beta = (\beta_{ij})$ and $h = (h_i)$ with some $h_i > 0$,*

$$\big\langle F\big((z_i), (u_i \cdot u_j)\big)\big\rangle_{\beta,h} = F(1, -1). \tag{2.11}$$

On the right-hand side of (2.11), 1 stands for the vector in $\mathbb{R}^\Lambda$ with all entries equal to 1, and -1 stands for the $|\Lambda| \times |\Lambda|$ matrix with all entries -1. For example, $\langle z_i\rangle_{\beta,h} = 1$ and $\langle u_i \cdot u_j\rangle_{\beta,h} = -1$.

Beyond: $\mathbb{H}^{n|2m}$. There is a natural generalisation of the above models to the broader class of $\mathbb{H}^{n|2m}$ models with $n + 1$ commuting coordinates and $2m$ anticommuting coordinates. Generalising Theorem 2.1, there is an exact correspondence between observables of the $\mathbb{H}^{n|2m}$ and $\mathbb{H}^{n+2|2m+2}$ models, see [11, SECTION 2]. For developments when $n = 0$, see [25].

Symmetries. The $\mathbb{H}^{n|2m}$ models have continuous symmetries which are analogues of the rotations of the $O(n)$ models. For example, for the hyperbolic plane $\mathbb{H}^2$, these symmetries are Lorentz boosts and rotations. The infinitesimal generator of Lorentz boosts in the xz-plane is the linear differential operator T acting as

$$Tz = x, \quad Tx = z, \quad Ty = 0. \tag{2.12}$$

If $\mathbb{H}^2$ is parametrised by $(x, y) \in \mathbb{R}^2$, then $T = z\partial_x$. For the hyperbolic sigma models, there is an infinitesimal boost $T_i = z_i \partial_{x_i}$ at each vertex i. Haar measure on $\mathbb{H}^2$ and the action $H_{\beta,0}$ with $h = 0$ are invariant under these symmetries, i.e. $\sum_i T_i H_{\beta,0} = 0$. Analogous symmetries exist for $\mathbb{H}^{0|2}$ and $\mathbb{H}^{2|2}$. If $h > 0$ then $\sum_i T_i H_{\beta,h} \neq 0$, and the symmetries are said to be explicitly broken by the external field. Important consequences of these symmetries are Ward identities. For example, when $n > 0$ (such as for the $\mathbb{H}^2$ and $\mathbb{H}^{2|2}$ models), for $h > 0$,

$$\frac{\langle z_i\rangle_{\beta,h}}{h} = \sum_j \langle x_i x_j\rangle_{\beta,h}, \tag{2.13}$$

and when $m > 0$ (such as for the $\mathbb{H}^{2|2}$ and $\mathbb{H}^{0|2}$ models),

$$\frac{\langle z_i \rangle_{\beta,h}}{h} = \sum_j \langle \xi_i \eta_j \rangle_{\beta,h}. \tag{2.14}$$

Here x_i and (ξ_i, η_i) stand for an even (bosonic) coordinate and pair of odd (fermionic) coordinates when $n, m > 0$, respectively. The proofs of these identities boil down to integration by parts; see, e.g. [34, APPENDIX B] or [11, LEMMA 2.3].

Spontaneous symmetry breaking. For the $\mathbb{H}^2$ and $\mathbb{H}^{2|2}$ models on a fixed finite graph, it is a consequence of the non-compactness of the hyperbolic symmetry that, for example, $\langle x_0^2 \rangle_{\beta,h}$ diverges as $h \downarrow 0$. Similarly, for the $\mathbb{H}^{0|2}$ model on a finite graph, symmetry implies that $\langle z_0 \rangle_{\beta,h}$ tends to 0 as $h \downarrow 0$. One of the main questions of statistical physics is whether a symmetry survives in the infinite volume limit, or if it is *spontaneously broken*. To make this precise, it is convenient to consider a finite volume criterion for this question. Consider a sequence of finite graphs Λ that approximate $\mathbb{Z}^d$ in a suitable way (denoted $\Lambda \to \mathbb{Z}^d$), and let $\langle \cdot \rangle_{\beta,h}$ be the corresponding finite volume expectations. For the $\mathbb{H}^2$ and $\mathbb{H}^{2|2}$ models, there is spontaneous symmetry breaking (SSB) for a given β if

$$\lim_{h \downarrow 0} \lim_{\Lambda \to \mathbb{Z}^d} \langle x_0^2 \rangle_{\beta,h} < \infty, \tag{2.15}$$

and similarly for the $\mathbb{H}^{0|2}$ model there is SSB if

$$\lim_{h \downarrow 0} \lim_{\Lambda \to \mathbb{Z}^d} \langle z_0 \rangle_{\beta,h} > 0. \tag{2.16}$$

These notions can be understood by noticing that when the two limits are exchanged the inequalities do not hold: in finite volume the $h = 0$ symmetries are restored in the $h \downarrow 0$ limit, while they are not in infinite volume if SSB occurs. There are other notions of SSB for hyperbolic spin models, but those in (2.15)–(2.16) capture the relevant phenomena from the perspective of the Anderson transition [34, SECTION 4.2], as well as from the perspective of the associated probabilistic models, as will be discussed in Sections 4 and 5.

We briefly summarise when SSB occurs for the $\mathbb{H}^2$, $\mathbb{H}^{0|2}$, and $\mathbb{H}^{2|2}$ models. In $d = 2$, (2.15) and (2.16) do not hold for any $\beta > 0$. These results are versions of the Mermin–Wagner theorem [12,50,56,57,64]. The situation is different in $d \geq 3$. For the $\mathbb{H}^2$ model, SSB occurs for any $\beta > 0$ as a result of convexity [49]. The $\mathbb{H}^{2|2}$ and $\mathbb{H}^{0|2}$ models, however, have phase transitions: SSB in the form of (2.15) and (2.16), respectively, occurs in $d \geq 3$ if and only if β is sufficiently large [10,34]. Once SSB is known to occur (or not), it is interesting and physically relevant to ask more precise questions, e.g. about the asymptotics of the correlation functions $\langle x_i x_j \rangle_{\beta,h}$. Sections 4 and 5 will discuss SSB and sharper questions for the $\mathbb{H}^{2|2}$ and $\mathbb{H}^{0|2}$ models.

Horospherical coordinates. An important tool for the study of the above models are horospherical coordinates for the superspaces $\mathbb{H}^{n|2m}$ with $n \geq 2$ [33,34]. For the hyperbolic plane $\mathbb{H}^2$ these are coordinates $(t, s) \in \mathbb{R}^2$ such that

$$x = \sinh(t) - \frac{1}{2} e^t |s|^2, \quad y = e^t s, \quad z = \cosh(t) + \frac{1}{2} e^t |s|^2. \tag{2.17}$$

For the space $\mathbb{H}^n$, these coordinates generalise by taking $s = (s^i) \in \mathbb{R}^{n-1}$. For the superspaces $\mathbb{H}^{n|2m}$, in addition there are m pairs Grassmann coordinates $\psi = (\psi^i)$, $\bar{\psi} = (\bar{\psi}^i)$ such that

$$\begin{aligned} x &= \sinh(t) - \frac{1}{2}e^t|s|^2 - e^t\psi\bar{\psi}, \quad y = e^t s, \quad \xi = e^t\psi, \quad \eta = e^t\bar{\psi}, \\ z &= \cosh(t) + \frac{1}{2}e^t|s|^2 + e^t\psi\bar{\psi}, \end{aligned} \tag{2.18}$$

where we are using the abbreviation $\psi\bar{\psi} = \sum_{i=1}^m \psi^i\bar{\psi}^i$ if there are m Grassmann components. In these coordinates the action becomes

$$\begin{aligned} H_{\beta,h} = \beta \sum_{ij} \Big(\big(\cosh(t_i - t_j) - 1\big) + \frac{1}{2}e^{t_i+t_j}|s_i - s_j|^2 + e^{t_i+t_j}(\psi_i - \psi_j)(\bar{\psi}_i - \bar{\psi}_j) \Big) \\ + h \sum_i \Big(\big(\cosh(t_i) - 1\big) + \frac{1}{2}e^{t_i}|s_i|^2 + e^{t_i}\psi_i\bar{\psi}_i \Big), \end{aligned} \tag{2.19}$$

and the hyperbolic reference measure is $dt\, ds\, \partial_{\bar{\psi}}\partial_{\psi} e^{(n-2m-1)\sum_i t_i}$, where $\partial_{\bar{\psi}}\partial_{\psi}$ denotes Grassmann integration if $m > 0$. A crucial feature of (2.19) is that the s and $\psi, \bar{\psi}$ variables appear quadratically in $H_{\beta,h}$ and hence can be integrated out via exact Gaussian computations. The t-marginal is thus proportional to the *positive* measure $e^{-\tilde{H}_{\beta,h}}\, dt$ where

$$\begin{aligned} \tilde{H}_{\beta,h}(t) = \beta \sum_{ij} \big(\cosh(t_i - t_j) - 1\big) + h \sum_i \big(\cosh(t_i) - 1\big) \\ + \frac{n - 2m - 1}{2}\Big(\log\det\big(-\Delta_{\beta(t)} + h(t)\big) - 2\sum_i t_i \Big). \end{aligned} \tag{2.20}$$

In (2.20), $-\Delta_{\beta(t)} + h(t)$ is the t-dependent matrix acting as

$$\big(-\Delta_{\beta(t)} f + h(t) f\big)_i = -\sum_{j\sim i} \beta e^{t_i+t_j}(f_j - f_i) + h e^{t_i} f_i. \tag{2.21}$$

The t-dependent weights $\beta_{ij}(t) = \beta e^{t_i+t_j}$ and $h_i(t) = he^{t_i}$ generalise immediately to edge- and vertex-dependent reference weights. The determinant in (2.20) arises from the Gaussian integration over the s and $\psi, \bar{\psi}$ variables. Since the t-field is distributed according to a positive measure, one can use standard tools from analysis. This is useful since, e.g. for all $\mathbb{H}^{n|2m}$ models,

$$\langle z_i \rangle_{\beta,h} = \langle z_i + x_i \rangle_{\beta,h} = \big\langle e^{t_i} \big\rangle_{\beta,h}, \tag{2.22}$$

where the first identity used that $\langle x_i \rangle_{\beta,h} = 0$, by symmetry. For the pinned expectations, analogous representations hold with $h = 0$ and $t_0 = 0$.

3. PHYSICAL BACKGROUND: ANDERSON TRANSITION

This section briefly discusses the origins of hyperbolic spin systems as simplified models for the Anderson delocalisation–localisation transition. For a more detailed survey about this, we refer in particular to [71]. Further excellent surveys include [70] and [38, 58] for a physics perspective. For general background on the Anderson transition, see [1].

Consider a random matrix $H = (H(i,j))_{i,j\in\Lambda}$ such as the Anderson Hamiltonian $H = H_\beta = -\beta\Delta + V$ where $V = (V_i)_{i\in\Lambda}$ is an i.i.d. Gaussian potential, Λ is a discrete torus approximating $\mathbb{Z}^d$, and Δ is the lattice Laplacian on Λ. The fundamental question is to determine whether or not the spectrum of H contains an absolutely continuous part in the infinite volume limit, and very closely related to this, if the eigenfunctions (often called states in this context) of H are extended or localised. Extended states correspond to a metallic phase while localised states correspond to an insulating phase. To discuss this further, define the two-point correlation function

$$\tau_{\beta,E,h}(j,k) = \mathbb{E}\big|(H_\beta - E - ih)^{-1}(j,k)\big|^2, \quad j,k \in \Lambda, \tag{3.1}$$

where $i = \sqrt{-1}$. The existence of extended states for energies near E is essentially implied by $\lim_{h\downarrow 0}\lim_{\Lambda\to\mathbb{Z}^d} \tau_{\beta,E,h}(j,j) < \infty$. For the Anderson model, it is a long-standing conjecture that this occurs in $d \geq 3$ for E inside the spectrum of $-\beta\Delta$ when β is sufficiently large. In the same setting, the more precise quantum diffusion conjecture asserts

$$\lim_{h\downarrow 0}\lim_{\Lambda\to\mathbb{Z}^d} \tau_{\beta,E,h}(j,k) \approx D(E,\beta)(-\Delta)^{-1}(j,k) \sim C(E,\beta)|j-k|^{-(d-2)}, \quad j,k \in \mathbb{Z}^d \tag{3.2}$$

for some constants C, D, and where the asymptotics hold for $|j-k| \to \infty$. This gives a hint that the conjecture might be difficult: the two-point function decays slowly, like that of the massless Gaussian free field. Such behaviour also occurs for fluctuations of spontaneously broken continuous symmetries (Goldstone modes). In [38, 74] it was argued that the origin of extended states is the existence of SSB for a (complicated) spin model with hyperbolic symmetry, and that quantum diffusion is exactly the associated Goldstone mode. The spin model is based on the supersymmetric approach to the replica trick for computing the two-point function.

We briefly indicate some parallels between the present discussion and Section 2. The elementary identity

$$\frac{1}{h}\mathbb{E}\,\mathrm{Im}(H_\beta - E - ih)^{-1}(j,j) = \sum_k \mathbb{E}\big|(H_\beta - E - ih)^{-1}(j,k)\big|^2, \tag{3.3}$$

which is also valid without expectations, is analogous to the Ward identities (2.13)–(2.14). Thus the role of $\langle z_j\rangle$ is played by $\mathbb{E}\,\mathrm{Im}(H_\beta - E - ih)^{-1}(j,j)$. In the limit $h \downarrow 0$ this is π times the density of states $\rho(E)$, i.e. the asymptotic eigenvalue distribution. The role of the two-point functions $\langle x_j x_k\rangle$ or $\langle \xi_j \eta_k\rangle$ is played by $\tau_{\beta,E,h}(j,k) = \mathbb{E}|(H_\beta - E - ih)^{-1}(j,k)|^2$. The absolute values in the latter correlation function are essential and the origin of the hyperbolic symmetry [71, SECTION 2.3]. The non-compactness of the hyperbolic symmetry manifests itself in the high temperature phase: the unboundedness of $\tau_{\beta,E,h}(j,k)$ as $h \downarrow 0$ signals an absence of delocalisation. The stronger notion of localisation corresponds to

$$\tau_{\beta,E,h}(j,k) \approx \frac{e^{-c|j-k|}}{h}. \tag{3.4}$$

The divergence as $h \downarrow 0$ is analogous to the behaviour of the $\mathbb{H}^{2|2}$ model, see Section 2, and is different from that of spin systems with compact symmetry. For further discussion, see [71].

Dictionary. The analogies between the expected behaviours of the Anderson model and the probabilistic models described in the next two sections are summarised below. In $d = 2$, $\beta_c = \infty$, while $\beta_c < \infty$ for $d \geq 3$.

	$\beta < \beta_c$	$\beta > \beta_c$
Anderson Model	localised (insulating) phase	extended (metallic) phase
VRJP	positive recurrent phase	transient phase
Arboreal gas	subcritical percolation phase	percolating phase

Logically, there is the possibility of non-extended states that are not localised, which would correspond to a null-recurrent phase for the VRJP and a phase of the arboreal gas where infinite clusters do not occur, but the cluster size distribution has infinite mean.

4. LINEARLY REINFORCED WALKS AND $\mathbb{H}^{2|2}$

Formulas arising in the study of the $\mathbb{H}^{2|2}$ model (e.g. (2.21)) have interpretations in terms of random walks, and similarities with ERRW did not go unnoticed [34, SECTION 1.5]. This was given an explanation by Sabot and Tarrès [65]; the explanation passes through another reinforced random walk, which we now introduce. Fix edge weights $\beta_{ij} > 0$ for each edge $ij \in E$, and set $\beta_{ij} = 0$ if $ij \notin E$. The *vertex-reinforced jump process (VRJP)* with $X_0 = 0$ is the continuous-time self-interacting random walk with transition probabilities

$$\mathbb{P}_0^{\mathrm{VJRP}(\beta)}\big[X_{t+dt} = j \,|\, (X_s)_{s \leq t}, X_t = i\big] = \beta_{ij} L_t^j, \quad L_t^j = 1 + \int_0^t 1_{X_s = j}\, ds. \tag{4.1}$$

The quantity L_t^j is the *local time* at j at time t, up to the shift by 1. In words, then, conditionally on the shifted local times at time t and that $X_t = i$, a VRJP jumps to site j with probability proportional to $\beta_{ij} L_t^j$. Thus previously vertices visited are preferred. The amount of local time accrued at i before jumping away has the distribution of an exponential random variable with rate $\sum_j \beta_{ij} L_t^j$. With this in mind, large edge weights β_{ij} heuristically correspond to weak reinforcement: jumps occur quickly and do not alter the local time profile too much.

Sabot and Tarrès gave an exact formula for the (properly scaled) limiting local times of the VRJP, and explained that this distribution is also the distribution of the t-field of the $\mathbb{H}^{2|2}$ model. They further showed that the magic formula for the ERRW follows from this result, see Section 4.4 below. Similarly to the ERRW, the VRJP can be expressed as a continuous-time random walk in a random environment. The next theorem is a slightly informal statement of this result. The precise formulation requires looking at the VRJP in the correct time parameterisation; see [65]. For a symmetric square matrix A with $\sum_j A_{ij} = 0$ for all rows i, we write $\det^0(A)$ for the value of any principal cofactor of A, e.g. the determinant with the first row and column of A removed.

Theorem 4.1 (Magic formula for VJRP [65]). *Let $G = (\Lambda, E)$ be a finite graph with $|\Lambda| = N$. In the exchangeable time parameterisation of the VRJP,*

$$\mathbb{P}_0^{\mathrm{VJRP}(\beta)}[\cdot] = (2\pi)^{-\frac{N-1}{2}} \int_{\mathbb{R}^{\Lambda\setminus 0}} \mathbb{P}_0^{\mathrm{SRW}(c(t))}[\cdot] e^{-\frac{\beta}{2}\sum_{i,j}\cosh(t_i-t_j)} \big(\det{}^0(-\Delta_{\beta(t)})\big)^{\frac{1}{2}} \prod_{k\in\Lambda\setminus 0} e^{-t_k}\, dt_k, \tag{4.2}$$

where $\mathbb{P}_0^{\mathrm{SRW}(c(t))}$ is the distribution of a continuous-time simple random walk with conductances $c(t)_{ij} = \beta e^{t_i+t_j}$ started at 0.

The measure on the right-hand side of (4.2) is exactly the horospherical t-marginal of the $\mathbb{H}^{2|2}$ model (with $h = 0$ and pinned at 0). The existence of a phase transition between a transient and a recurrent phase of the VRJP on $\mathbb{Z}^d$ for $d \geq 3$ now essentially follows from the following earlier results for the $\mathbb{H}^{2|2}$ model (and extensions to the pinned model):

Theorem 4.2 (SSB for $\mathbb{H}^{2|2}$ [34]). *Let $d \geq 3$ and $\beta \geq \beta_1$. There exists $C_\beta > 0$ such that*

$$\lim_{h\downarrow 0}\lim_{\Lambda\to\mathbb{Z}^d} \big\langle \cosh(t_i)^8 \big\rangle_{\beta,h} \leq C_\beta. \tag{4.3}$$

Similar statements hold for other observables and for the pinned model.

Theorem 4.3 (Localisation for $\mathbb{H}^{2|2}$ [33]). *Let $d \geq 1$ and $\beta \leq \beta_0$. There exist $C_\beta, c_\beta > 0$ such that*

$$\langle x_i x_j \rangle_{\beta,h} \leq \frac{C_\beta}{h} e^{-c_\beta|i-j|}. \tag{4.4}$$

Similar statements hold for other observables and for the pinned model.

The existence of a recurrent phase for small β has also been proved more directly from the definition of the VRJP [4]. A proof of transience that only uses the random walk point of view seems challenging, and would be of interest.

4.1. Hyperbolic symmetry and the VRJP

A more direct and general connection between hyperbolic spin systems and the VRJP was found later [12]. Towards this, observe (as was already done in [65]) that the joint process (X_t, L_t) of the VRJP and its local time is a Markov process, where $L_t = (L_t^j)_{j\in V}$. The infinitesimal generator $\mathcal{L}$ of the joint process acts on $g\colon V\times[0,\infty)^V \to \mathbb{R}$ by

$$\mathcal{L}g(i,\ell) = \sum_j \beta_{ij}\ell_j\big(g(j,\ell) - g(i,\ell)\big) + \frac{\partial g(i,\ell)}{\partial \ell_i}. \tag{4.5}$$

Write $\mathbb{E}_i^{\mathrm{VRJP}(\beta,\ell)}$ for the expectation of the joint process with initial vertex i and local times $\ell = (\ell_i)_{i\in\Lambda}$. The definition (4.1) corresponds to $\ell_i = 1$ for all i.

To connect the VRJP to hyperbolic symmetry, consider the $\mathbb{H}^2$ model, for example, and recall the infinitesimal generator T_i of Lorentz boosts in the $x_i z_i$-plane acting at vertex i from (2.12). Then under mild hypotheses on G, integration by parts and (2.12) yield

$$-\sum_j \int_{(\mathbb{H}^2)^\Lambda} \big(\mathcal{L}G(j,z)\big)x_i x_j e^{-H_{\beta,0}(u)} \prod_{k\in\Lambda} du_k = \int_{(\mathbb{H}^2)^\Lambda} (T_i x_i) G(j,z) e^{-H_{\beta,0}(u)} \prod_{k\in\Lambda} du_k. \tag{4.6}$$

Thus boosts are adjoint to the generator of the VJRP. A consequence is the next theorem.

Theorem 4.4. *Consider the $\mathbb{H}^2$ model. If $F\colon \mathbb{R}^\Lambda \to \mathbb{R}$ decays fast enough, then*

$$\langle x_i x_j F(z)\rangle_{\beta,0} = \left\langle z_i \int_0^\infty dt\, \mathbb{E}_i^{\mathrm{VRJP}(\beta,z)}\big[F(L_t)1_{X_t=j}\big]\right\rangle_\beta. \tag{4.7}$$

Sketch of proof. Normalise (4.6) and choose $G(j,\ell) = G_t(j,\ell) = \mathbb{E}_j^{\mathrm{VRJP}(\beta,\ell)} F(L_t)$. Since (X_t, L_t) is a Markov process with generator $\mathcal{L}$, we have $\mathcal{L}G_t = \partial_t G_t$. Integrating the resulting identity over t in $(0,\infty)$ gives the result. ■

Theorem 4.4 shows that $\mathbb{H}^2$ quantities can be computed in terms of the averages of VRJP quantities, the average being over the initial local time of the VRJP. This average is inconvenient for studying the VRJP itself. The computations above, however, immediately generalise to other hyperbolic spin models. For the $\mathbb{H}^{2|2}$ model, one can in addition use Theorem 2.1 to exactly compute the undesirable average. The result is the following theorem.

Theorem 4.5. *Consider the $\mathbb{H}^{2|2}$ model. Then for any $F\colon \mathbb{R}^\Lambda \to \mathbb{R}$ that decays fast enough,*

$$\langle x_i x_j F(z)\rangle_{\beta,0} = \int_0^\infty dt\, \mathbb{E}_i^{\mathrm{VRJP}(\beta)}\big[F(L_t)1_{X_t=j}\big]. \tag{4.8}$$

In particular, $\langle x_i^2\rangle_{\beta,h}$ is the expected time the VRJP started from i spends at i when killed at rate $h > 0$. This relation can be used to prove the VRJP is recurrent in two dimensions, irrespective of the reinforcement strength $\beta > 0$, by proving a Mermin–Wagner theorem for the $\mathbb{H}^{2|2}$ model [12]. Informally, Mermin–Wagner theorems assert that continuous symmetries cannot be spontaneously broken in $d = 1, 2$. As discussed earlier, for the $\mathbb{H}^{2|2}$ model SSB corresponds to a finite variance, i.e. transience.

Isomorphism theorems. Theorems 4.4 and 4.5 are examples of *isomorphism theorems*, meaning identities relating the local time field of a stochastic process to a spin system. The first example of such a result related simple random walk to the Gaussian free field and was obtained by Brydges, Fröhlich, and Spencer [19]. They were inspired by Symanzik [73]. The formulation as a distributional identity is due to Dynkin [36]; sometimes the result is called the BFS–Dynkin isomorphism. A host of other isomorphism theorems have been found in Gaussian settings, see [53]. Other isomorphism theorems for the VRJP can be obtained by the approach above, and it is possible to obtain Theorem 4.1 in this way, see [13]. Isomorphisms for the VRJP can also be obtained by expressing the VRJP as a mixture of Markov processes and using isomorphism theorems for the Markov processes; see [23].

4.2. Random Schrödinger representation and STZ field

In [33], it was observed that after conjugation by the diagonal matrix $e^{-t} = (e^{-t_i})_i$, the matrix $-\Delta_{\beta(t)} + h(t)$ in (2.21) becomes a Schrödinger operator with t-dependent potential:

$$e^{-t} \circ \big(-\Delta_{\beta(t)} + h(t)\big) \circ e^{-t} = -\Delta_\beta + V(t), \quad V_i(t) = \sum_j \beta_{ij}(e^{t_j - t_i} - 1) + h_i e^{-t_i}. \tag{4.9}$$

This point of view led to the proof of Theorem 4.3. It was later recognised that this random Schrödinger point of view can be used to obtain a powerful representation of the t-field [67]. For the pinned $\mathbb{H}^{2|2}$ model with $h = 0$ and $t_0 = 0$, the t-field measure (2.20) can be written in terms of $-\Delta_\beta + V(t)$ using that

$$e^{-\tilde{H}_\beta(t)} = e^{-\frac{1}{2}\sum_i V_i(t)}\big(\det\big(-\Delta_\beta + V(t)\big)\big)^{1/2}. \tag{4.10}$$

This suggests it might be useful to change variables from t to $V(t)$. This change of variables is not directly well-defined when $t_0 = 0$ since the set of V such that $(-\Delta_\beta + V)$ is positive definite is $|\Lambda|$-dimensional. This can be sidestepped by treating V as the fundamental variable, i.e. considering

$$e^{-\frac{1}{2}\sum_i V_i}\big(\det(-\Delta_\beta + V)\big)^{-1/2}\mathbf{1}(-\Delta_\beta + V \text{ is positive definite})\, dV. \tag{4.11}$$

The random vector $B_i = \frac{1}{2}(V_i + \sum_j \beta_{ij})$ is often called the 'β-field,' but since we use β for edge weights (inverse temperature), we will denote it by B instead and call it the *STZ field*.

Theorem 4.6. *The Laplace transform of the STZ field is given by*

$$\mathbb{E}e^{-(\lambda,B)} = \prod_i \frac{1}{(\lambda_i+1)^{1/2}} \prod_{ij} e^{-\beta_{ij}(\sqrt{\lambda_i+1}\sqrt{\lambda_j+1}-1)}. \tag{4.12}$$

Moreover, the t-field (pinned at any vertex) can be recovered in distribution from B.

In particular, the theorem implies the STZ field is 1-dependent. In [68], this remarkable property of the STZ field was used to construct an infinite volume version on $\mathbb{Z}^d$, and applied to characterise transience and recurrence of the VRJP in terms of a 0/1 law.

4.3. Phase diagram of the VRJP

The most basic qualitative question one can ask about the VRJP is whether it is recurrent or transient for a given reinforcement strength $\beta > 0$. This may in principle depend on the precise notion of recurrence used, as the VRJP is non-Markovian. As discussed above, for $d \geq 3$ the existence of a phase in which the VRJP is almost surely recurrent was established in [4,65], and an almost surely transient phase in [65]. For $d = 2$, recurrence for all $\beta > 0$ in the sense of infinite expected local time at the initial vertex was established in [12]. Proofs of almost sure recurrence followed shortly [50,64]. Similar results had previously been established for the ERRW [12,56,68].

The qualitative behaviour of the VRJP is almost completely understood on $\mathbb{Z}^d$ due to the following remarkable correlation inequality of Poudevigne.

Theorem 4.7. *For the $\mathbb{H}^{2|2}$ model and any convex function f, the expectation $\langle f(e^{t_j})\rangle_\beta^0$ is increasing in all weights $\beta = (\beta_{ij})$.*

The proof of Theorem 4.7 relies on the STZ field [62]. This inequality implies that transience is a monotone property with respect to the constant initial reinforcement parameter β. Combined with the results of the previous paragraph, this implies that the VRJP has a sharp transition from almost sure recurrence to almost sure transience on $\mathbb{Z}^d$ for constant β: recurrence for $\beta < \beta_c(d)$ and transience for $\beta > \beta_c(d)$. The behaviour at β_c is open. Poudevigne's correlation inequality also leads to a proof of recurrence in $d = 2$.

4.4. Further discussion

Back to edge-reinforced random walk. The connection of the ERRW to the $\mathbb{H}^{2|2}$ model is somewhat less direct than for the VRJP: it turns out that the ERRW is an average of VRJPs [65]. Somewhat more precisely, ERRW with initial edge weights α can be obtained from the VRJP with initial edges weights β if the β_{ij} are chosen to be independent Gamma random variables with mean α_{ij}. While this additional randomness presents some difficulties, the existence of a transient phase for the ERRW in $d \geq 3$ was obtained by similar methods to that of the VRJP [32]. In terms of the spin model, the Gamma-distributed random edge weights correspond to replacing the exponential $e^{\sum_{ij} \beta_{ij}(u_i \cdot u_j + 1)}$ by $\prod_{ij}(-u_i \cdot u_j)^{-\alpha_{ij}}$ in the (super)measure. Such a product weight is often called a Nienhuis interaction.

Interestingly, the recurrence of the ERRW in two dimensions was obtained before the recurrence of the VRJP. This was possible due to insights of Merkl and Rolles, who directly proved a Mermin–Wagner-type theorem for the ERRW by making use of the magic formula [57]. Merkl and Rolles were able to conclude recurrence of the ERRW on $\mathbb{Z}^2$ for strong reinforcement if each edge of the lattice was replaced by a long path. Sabot and Zeng's proved recurrence on $\mathbb{Z}^2$ for all reinforcement strengths by obtaining a characterisation of recurrence in terms of the STZ field [68], and showing that an estimate from [57] implies recurrence. The ergodic properties of the STZ field play a crucial role in this argument.

Beyond $\mathbb{Z}^d$. There are also results for the VRJP beyond $\mathbb{Z}^d$. The existence of a transition on trees was proven in [27], and on non-amenable graphs in [4]. A fairly complete understanding on trees has been obtained, see [7] and references therein.

Future directions. There remain many open questions. What is the critical behaviour of the VRJP and the $\mathbb{H}^{2|2}$ model on $\mathbb{Z}^d$, $d \geq 3$? Is there an upper critical dimension? For $d = 3$ aspects of this question were studied numerically in [35], and evidence was found for the existence of a multifractal structure in the $\mathbb{H}^{2|2}$ model. Multifractal structure is also expected near the Anderson transition for random Schrödinger operators. For the regular tree (Bethe lattice), further remarkable critical behaviour was observed, in part numerically, in [75]. This reference concerns a more complicated sigma model, but the main predictions also apply to the $\mathbb{H}^{2|2}$ model [35]. On $\mathbb{Z}^2$ the VRJP is believed to be *positive* recurrent, i.e. exponentially localised, but this important conjecture about the $\mathbb{H}^{2|2}$ model and the VRJP remains open. The heuristic for positive recurrence is based on the (marginal) renormalisation group flow and goes along with the prediction of asymptotic freedom at short distances [34, SECTION 4.3]. Analogous predictions based on similar heuristics exist for the $2d$ Heisenberg model, the $2d$ Anderson model, $4d$ non-abelian Yang–Mills theory, and the $2d$ arboreal gas (discussed below). Another question is to understand the VRJP in $d \geq 3$ with non-constant initial local times: Theorem 4.4 and results of [72] suggest the VRJP is always transient if started with initial local times given by the z-field of the $\mathbb{H}^2$ model. Understanding the properties of the z-field that destroy the phase transition would be interesting.

5. THE ARBOREAL GAS AND $\mathbb{H}^{0|2}$

The arboreal gas is the uniform measure on (unrooted spanning) forests of a weighted graph. More precisely, given an undirected graph $G = (\Lambda, E)$, a forest $F = (\Lambda, E(F))$ is an acyclic subgraph of G having the same vertex set as G. Given an edge weight $\beta > 0$ (inverse temperature) and a vertex weight $h \geq 0$ (external field), the probability of an edge set F under the arboreal gas measure is

$$\mathbb{P}_{\beta,h}[F] = \frac{1}{Z_{\beta,h}} \beta^{|E(F)|} \prod_{T \in F} \bigl(1 + h\bigl|V(T)\bigr|\bigr) \mathbf{1}(F \text{ is a forest}) \tag{5.1}$$

where $T \in F$ denotes that T is a tree in the forest F, i.e. a connected component of F. We write $\mathbb{P}_\beta = \mathbb{P}_{\beta,0}$. As for the VRJP, the generalisation to edge- and vertex-dependent weights $\beta = (\beta_{ij})$ and $h = (h_i)$ is straightforward, and is sometimes useful.

The arboreal gas arises naturally in the context of the q-state random cluster model (q-RCM), which we recall is the model defined by (5.1) by omitting the indicator function and instead weighting each component by a factor $q > 0$. In particular, $q = 1$ is Bernoulli bond percolation. On a finite graph, the arboreal gas edge weight β' is the limit of the q-RCM as $q, \beta \to 0$ such that $\beta/q \to \beta'$, and it is natural to think of the arboreal gas as the 0-RCM. The most fundamental question about the arboreal gas is whether or not it has a percolation phase transition. It is straightforward to establish a subcritical phase when β is small: the arboreal gas can be stochastically dominated by bond percolation [43, THEOREM 3.21], and for β small the domination is by subcritical percolation.

5.1. Phase transitions for the arboreal gas

The existence of a supercritical phase for the arboreal gas is a more subtle question than for the q-RCM with $q > 0$. One way to see this subtlety is based on symmetries. To discuss this, recall that for $q \in \{2, 3, \dots\}$ there is a connection between the q-RCM and the q-state Potts model [40]. In particular, spin–spin correlations in the q-state Potts model are equivalent to connection probabilities in the q-RCM. The results of [22, 45] extend this relationship to $q = 0$: the $\mathbb{H}^{0|2}$ model is a spin representation of the arboreal gas.

Theorem 5.1. *Let $\langle\cdot\rangle_\beta$ and $\mathbb{P}_\beta$ denote the $\mathbb{H}^{0|2}$ and arboreal gas measures on a finite graph. For vertices $i, j \in \Lambda$,*

$$\mathbb{P}_\beta[i \leftrightarrow j] = -\langle u_i \cdot u_j \rangle_\beta. \tag{5.2}$$

Moreover, the partition functions of the $\mathbb{H}^{0|2}$ model and the arboreal gas coincide.

Strictly speaking, the $\mathbb{H}^{0|2}$ formulation of Theorem 5.1 first occurred in [11] as a reformulation of [22, 45]; the hyperbolic point of view plays an important role in the proof of Theorem 5.3 below. Theorem 5.1 suggests the existence of a supercritical phase for the arboreal gas may depend on the dimension, as strong connection probabilities corresponds to a symmetry breaking phase transition for the $\mathbb{H}^{0|2}$ model. Unlike the q-Potts models with $q \in \{2, 3, \dots\}$, this model possesses a continuous symmetry, so one might expect a Mermin–Wagner theorem to prevent such a transition in $d = 2$. This is indeed true:

Theorem 5.2 ([11, THEOREM 1.3]). *Let $d = 2$. For any $\beta > 0$, there exists $c(\beta) > 0$ such that $\mathbb{P}_\beta^\Lambda[0 \leftrightarrow j] \le |j|^{-c(\beta)}$ for any $\Lambda \subset \mathbb{Z}^2$.*

It is possible to predict Theorem 5.2 without knowing about the $\mathbb{H}^{0|2}$ spin representation as follows [28]. The critical value of β for the q-RCM on $\mathbb{Z}^2$ with $q \ge 1$ is known to be $\beta_c(q) = \sqrt{q}$ [14], and this self-dual point is predicted to be the critical point for all $q > 0$. Since the arboreal gas $\mathbb{P}_{\beta'}$ is the limit of the q-RCM with $\beta = \beta' q$, if the location of the critical point is continuous as $q \downarrow 0$, it follows that $\beta_c(0) = \infty$. These heuristics support the conjecture that connection probabilities of the $2d$ arboreal gas decay exponentially for any $\beta > 0$. Independent support for this conjecture can be obtained by renormalisation group heuristics, almost exactly as for the $2d$ VRJP [22].

Turning the preceding paragraph into a rigorous proof would be very interesting. It would also be interesting to have a probabilistic proof (in terms of forests) that the arboreal gas does not have a phase transition on $\mathbb{Z}^2$. The proof of Theorem 5.2 given in [11] follows different lines. A key step is the following, which reduces the proof to an adaptation of [64, THEOREM 1].

Theorem 5.3 (Magic formula for arboreal gas). *Let $G = (\Lambda, E)$ be a finite connected graph. For vertices $0, j \in \Lambda$,*

$$\mathbb{P}_\beta[0 \leftrightarrow j] = \frac{1}{Z_\beta} \int_{\mathbb{R}^{\Lambda\setminus 0}} e^{t_j} e^{-\frac{\beta}{2}\sum_{i,k} \cosh(t_i - t_k)} (\det{}^0(-\Delta_{\beta(t)}))^{3/2} \prod_{k\in\Lambda\setminus 0} e^{-3t_k}\, dt_k, \quad (5.3)$$

where $\det^0(-\Delta_{\beta(t)})$ denotes any principal cofactor of $-\Delta_{\beta(t)}$.

In outline, the proof of Theorem 5.3 consists of three steps: Theorem 5.1, rewriting the $\mathbb{H}^{0|2}$ expectation in terms of $\mathbb{H}^{2|4}$ by SUSY localisation, and then changing to horospherical coordinates and integrating out all but the t-field. The magic formula for the VRJP from Theorem 4.1 has a strikingly similar form, but with the two occurrences of 3s in (5.3) replaced by 1s. This difference in powers is due to there being two additional Grassmann Gaussian integrals for $\mathbb{H}^{2|4}$ as compared to $\mathbb{H}^{2|2}$.

In three and more dimensions the arboreal gas does, however, undergo a percolation phase transition. To state a precise theorem, let $\Lambda_N = \mathbb{Z}^d / L^N \mathbb{Z}^d$ denote a torus of side-length L^N with L large. The next theorem immediately implies that there is a macroscopic tree occupying most of the torus with large probability.

Theorem 5.4 ([10, THEOREM 1.1]). *Let $d \ge 3$. If β is sufficiently large, then there exists $\theta_d(\beta) = 1 - O(1/\beta)$, $D(\beta) > 0$, and $\kappa > 0$ such that*

$$\mathbb{P}_\beta^{\Lambda_N}[0 \leftrightarrow j] = \theta_d(\beta)^2 + D(\beta)(-\Delta)^{-1}(0,j) + O\left(\frac{1}{\beta |j|^{d-2+\kappa}}\right) + O\left(\frac{1}{\beta L^{\kappa N}}\right). \quad (5.4)$$

Similar asymptotics hold for other correlation functions.

The polynomial correction in Theorem 5.4 is the hallmark of critical behaviour in statistical mechanics, and is a manifestation of the Goldstone mode associated with the broken continuous symmetry of the $\mathbb{H}^{0|2}$ model at low temperatures. The proof of Theorem 5.4 relies essentially on the $\mathbb{H}^{0|2}$ representation (Theorem 5.1), and is based on a combination of Ward identities and a renormalisation group analysis. The renormalisation group analysis is based in part on methods developed previously in different contexts, in particular [8, 9, 18, 20, 21].

5.2. Further discussion

In contrast to the VRJP, even the qualitative phase diagram of the arboreal gas remains incomplete: we do not know the existence of a β_c such that percolation occurs if $\beta > \beta_c$, and does not if $\beta < \beta_c$. It is also more difficult to discuss the arboreal gas directly in the infinite volume limit than the VRJP; the analogue of the STZ field does not have finite dependence and is less obviously useful, and useful correlation inequalities to this end remain conjectural, see below. Nonetheless, many open questions beckon.

Critical behaviour. There is strong evidence that the upper critical dimension of the arboreal gas is $d = 6$, just as for bond percolation, and that the critical behaviour is governed by more conventional critical behaviour as compared to the $\mathbb{H}^{2|2}$ model [28], see also [39, 48].

Comparison with percolation. Recall that the analogue of Theorem 5.4 for Bernoulli percolation has an *exponentially* decaying correction [24]. Informally, this means that supercritical percolation with the giant removed behaves like subcritical percolation. This can be given a more precise meaning in the simpler setting of the Erdős–Rényi random graph, i.e. Bernoulli percolation on the complete graph K_N, where it is known as the discrete duality principle [2, SECTION 10.5].

The polynomial correction in Theorem 5.4 shows that the arboreal gas does not satisfy a duality principle. Rather, its supercritical phase behaves like a critical model off the giant. This can again be given a more precise formulation on the complete graph K_N and on the wired regular tree, where detailed results are known [37, 51, 54, 63]. In particular, the exact cluster distribution can be determined: on K_N in the supercritical phase there is a unique giant tree, and an unbounded number of trees of size $\Theta(N^{2/3})$.

It is natural to predict that the macroscopic behaviour of the arboreal gas on the d-dimensional tori Λ_N with $d \geq 3$ is similar to that on the complete graph. In particular, one expects a unique giant tree. The next-order critical corrections can also be expected to be similar, at least when $d > 6$. In particular, the second biggest tree should then have size comparable to $|\Lambda_N|^{2/3}$. Similar results have been established for critical Bernoulli percolation in high dimensions, see [44, CHAPTER 13]. More ambitiously, we expect the order statistics of the rescaled cluster size distribution to be universal, i.e. the same as on the complete graph as determined in [51, 54]. This conjecture may be easier to explore in other settings first, e.g. on expanders, where a phase transition can be established by elementary methods [42].

Infinite-volume geometry and the UST. There is a large body of literature in probability theory concerning *uniform spanning forests* (USF), meaning weak infinite-volume limits of uniform spanning tree (UST) measures on finite graphs, see [52, CHAPTER 10]. To avoid confusion with the arboreal gas (sometimes also called the USF [46]), we will call these infinite-volume limits the UST on $\mathbb{Z}^d$. While the component structure of the UST on a finite graph is not particularly interesting, the infinite volume limit is: Pemantle proved that there is a unique connected component on $\mathbb{Z}^d$ for $d \leq 4$, and infinitely many connected components on $\mathbb{Z}^d$ for $d > 4$ [59]. This happens as ‘long connections’ can be lost in the weak limit.

On a finite graph, the UST measure is the limit $\beta \to \infty$ of the arboreal gas with edge weights β, and it is natural to wonder if the arboreal gas at low temperatures $\beta \gg 1$ has similar properties to the UST in infinite volume. For global properties, this can evidently only happen when there is a percolation transition. We are therefore lead to ask: for $d \geq 3$ at low temperatures, is it the case that for $d = 3, 4$ the infinite-volume arboreal gas has a unique infinite tree, while for $d > 4$ there are infinitely many infinite trees? Is it the case that the infinite components of the arboreal gas are topologically one-ended, as for the UST?

Negative correlation. A key tool in studying the q-RCM with $q \geq 1$ is that it is *positively associated*: for increasing functions $f, g \colon \{0, 1\}^E \to \mathbb{R}$, the covariance of f and g is non-negative. This is a special case of the FKG inequality [41]. Positive association fails for $q < 1$ and for the arboreal gas. It is believed, but not known, that these models are in fact *negatively associated*: for $f, g \colon \{0, 1\}^E \to \mathbb{R}$ depending on disjoint sets of edges, the covariance of f and g is non-positive. Negative association is more subtle than positive association, and the development of flexible, yet powerful, theoretical frameworks is an active subject [3,16,17,60]. While some of this theory applies to the arboreal gas, it remains open to prove even the special case of *negative correlation*: for distinct edges $e, f \in E$,

$$\mathbb{P}_\beta[e, f \in F] \leq \mathbb{P}_\beta[e \in F]\mathbb{P}_\beta[f \in F]. \tag{5.5}$$

Negative correlation for all weights is equivalent to all connection probabilities $\mathbb{P}_\beta[0 \leftrightarrow j] = \langle e^{t_j} \rangle_\beta^0$ being increasing in all weights $\beta = (\beta_{ij})$, where the right-hand side is in terms of the t-field of the pinned $\mathbb{H}^{0|2}$ model. The analogue for the $\mathbb{H}^{2|2}$ model is precisely Poudevigne’s inequality, Theorem 4.7. Does this inequality extend to other $\mathbb{H}^{n|2m}$ models?

6. CONCLUDING REMARKS

This survey has focused on the connections between hyperbolic spin systems and probabilistic models that share phenomenology with the Anderson transition, including a number of open questions. It is also worth repeating a question from [49]: are there other models of random walk that are related to spin systems? A partial answer was given in [13], but we expect there is more to be discovered; see, e.g. [66]. Similarly, one may search for probabilistic representations of $\mathbb{H}^{n|2m}$ models for values of n, m not discussed here.

ACKNOWLEDGMENTS

Special thanks are due to Nick Crawford and Andrew Swan for their collaboration [10–13] on the papers on which much of this survey is based. We also thank Rémy Poudevigne for explaining the origin of the STZ field as discussed in Section 4.2.

FUNDING

This work was partially supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant No. 851682 SPINRG).

REFERENCES

[1] M. Aizenman and S. Warzel, *Random operators*. Grad. Stud. Math. 168, American Mathematical Society, Providence, RI, 2015.

[2] N. Alon and J. Spencer, *The probabilistic method*. Fourth edn., Wiley-Intersci. Ser. Discrete Math. Optim., 2016.

[3] N. Anari, K. Liu, S. Gharan, and C. Vinzant, Log-concave polynomials III: Mason's ultra-log-concavity conjecture for independent sets of matroids. 2018, arXiv:1811.01600.

[4] O. Angel, N. Crawford, and G. Kozma, Localization for linearly edge reinforced random walks. *Duke Math. J.* **163** (2014), no. 5, 889–921.

[5] S. Bacallado, Bayesian analysis of variable-order, reversible Markov chains. *Ann. Statist.* **39** (2011), no. 2, 838–864.

[6] S. Bacallado, V. Pande, S. Favaro, and L. Trippa, Bayesian regularization of the length of memory in reversible sequences. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** (2016), no. 4, 933–946.

[7] A.-L. Basdevant and A. Singh, Continuous-time vertex reinforced jump processes on Galton–Watson trees. *Ann. Appl. Probab.* **22** (2012), no. 4, 1728–1743.

[8] R. Bauerschmidt, D. Brydges, and G. Slade, Critical two-point function of the 4-dimensional weakly self-avoiding walk. *Comm. Math. Phys.* **338** (2015), no. 1, 169–193.

[9] R. Bauerschmidt, D. Brydges, and G. Slade, Logarithmic correction for the susceptibility of the 4-dimensional weakly self-avoiding walk: a renormalisation group analysis. *Comm. Math. Phys.* **337** (2015), no. 2, 817–877.

[10] R. Bauerschmidt, N. Crawford, and T. Helmuth, Percolation transition for random forests in $d \geq 3$. 2021, arXiv:2107.01878.

[11] R. Bauerschmidt, N. Crawford, T. Helmuth, and A. Swan, Random spanning forests and hyperbolic symmetry. *Comm. Math. Phys.* **381** (2021), no. 3, 1223–1261.

[12] R. Bauerschmidt, T. Helmuth, and A. Swan, Dynkin isomorphism and Mermin–Wagner theorems for hyperbolic sigma models and recurrence of the two-dimensional vertex-reinforced jump process. *Ann. Probab.* **47** (2019), no. 5, 3375–3396.

[13] R. Bauerschmidt, T. Helmuth, and A. Swan, The geometry of random walk isomorphism theorems. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2021), no. 1, 408–454.

[14] V. Beffara and H. Duminil-Copin, The self-dual point of the two-dimensional random-cluster model is critical for $q \geq 1$. *Probab. Theory Related Fields* **153** (2012), no. 3–4, 511–542.

[15] F. Berezin, *Introduction to superanalysis*. Math. Phys. Appl. Math. 9, Springer, 1987.

[16] J. Borcea, P. Brändén, and T. Liggett, Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.* **22** (2009), no. 2, 521–567.

[17] P. Brändén and J. Huh, Lorentzian polynomials. *Ann. of Math. (2)* **192** (2020), no. 3, 821–891.

[18] D. Brydges, Lectures on the renormalisation group. In *Statistical mechanics*, pp. 7–93, IAS/Park City Math. Ser. 16, Amer. Math. Soc., 2009.

[19] D. Brydges, J. Fröhlich, and T. Spencer, The random walk representation of classical spin systems and correlation inequalities. *Comm. Math. Phys.* **83** (1982), no. 1, 123–150.

[20] D. Brydges and G. Slade, A renormalisation group method. I. Gaussian integration and normed algebras. *J. Stat. Phys.* **159** (2015), no. 3, 421–460.

[21] D. Brydges and G. Slade, A renormalisation group method. II. Approximation by local polynomials. *J. Stat. Phys.* **159** (2015), no. 3, 461–491.

[22] S. Caracciolo, J. Jacobsen, H. Saleur, A. Sokal, and A. Sportiello, Fermionic field theory for trees and forests. *Phys. Rev. Lett.* **aracc** (2004), no. 8, 080601, 4.

[23] Y. Chang, D.-Z. Liu, and X. Zeng, On $H^{2|2}$ isomorphism theorems and reinforced loop soup. 2019, arXiv:1911.09036.

[24] J. Chayes, L. Chayes, G. Grimmett, H. Kesten, and R. Schonmann, The correlation length for the high-density phase of Bernoulli percolation. *Ann. Probab.* **17** (1989), no. 4, 1277–1302.

[25] N. Crawford, Supersymmetric hyperbolic σ-models and decay of correlations in two dimensions. *J. Stat. Phys.* **184** (2021), Article number 32.

[26] B. Davis and S. Volkov, Continuous time vertex-reinforced jump processes. *Probab. Theory Related Fields* **123** (2002), no. 2, 281–300.

[27] B. Davis and S. Volkov, Vertex-reinforced jump processes on trees and finite graphs. *Probab. Theory Related Fields* **128** (2004), no. 1, 42–62.

[28] Y. Deng, T. Garoni, and A. Sokal, Ferromagnetic phase transition for the spanning-forest model ($q \to 0$ limit of the Potts model) in three or more dimensions. *Phys. Rev. Lett.* **98** (2007), 030602.

[29] P. Diaconis, Recent progress on de Finetti's notions of exchangeability. In *Bayesian statistics, 3 (Valencia, 1987)*, pp. 111–125, Oxford Sci. Publ., 1988.

[30] P. Diaconis and D. Freedman, Finite exchangeable sequences. *Ann. Probab.* **8** (1980), no. 4, 745–764.

[31] P. Diaconis and S. Rolles, Bayesian analysis for reversible Markov chains. *Ann. Statist.* **34** (2006), no. 3, 1270–1292.

[32] M. Disertori, C. Sabot, and P. Tarrès, Transience of edge-reinforced random walk. *Comm. Math. Phys.* **339** (2015), no. 1, 121–148.

[33] M. Disertori and T. Spencer, Anderson localization for a supersymmetric sigma model. *Comm. Math. Phys.* **300** (2010), no. 3, 659–671.

[34] M. Disertori, T. Spencer, and M. Zirnbauer, Quasi-diffusion in a 3D supersymmetric hyperbolic sigma model. *Comm. Math. Phys.* **300** (2010), no. 2, 435–486.

[35] T. Dupré, Localization transition in three dimensions: Monte Carlo simulation of a nonlinear σ model. *Phys. Rev. B* **54** (1996), 12763–12774.

[36] E. Dynkin, Markov processes as a tool in field theory. *J. Funct. Anal.* **50** (1983), no. 2, 167–187.

[37] P. Easo, The wired arboreal gas on regular trees. 2021, arXiv:2108.04335.

[38] K. Efetov, Supersymmetry and theory of disordered metals. *Adv. Phys.* **32** (1983), no. 1, 53–127.

[39] L. Fei, S. Giombi, I. Klebanov, and G. Tarnopolsky, Critical Sp(N) models in 6−ε dimensions and higher spin dS/CFT. *J. High Energy Phys.* (2015), no. 9, 076, front matter+13.

[40] C. Fortuin and P. Kasteleyn, On the random-cluster model. I. Introduction and relation to other models. *Physica* **57** (1972), 536–564.

[41] C. Fortuin, P. Kasteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.* **22** (1971), 89–103.

[42] A. Goel, S. Khanna, S. Raghvendra, and H. Zhang, Connectivity in random forests and credit networks. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2037–2048, 2015.

[43] G. Grimmett, *The random-cluster model*. Grundlehren Math. Wiss. 333, Springer, Berlin, 2006.

[44] M. Heydenreich and R. van der Hofstad, *Progress in high-dimensional percolation and random graphs*. CRM Short Courses, Springer, 2017.

[45] J. L. Jacobsen and H. Saleur, The arboreal gas and the supersphere sigma model. *Nuclear Phys. B* **716** (2005), no. 3, 439–461.

[46] J. Kahn and M. Neiman, Negative correlation and log-concavity. *Random Structures Algorithms* **37** (2010), no. 3, 367–388.

[47] M. Keane and S. Rolles, Edge-reinforced random walk on finite graphs. In *Infinite dimensional stochastic analysis (Amsterdam, 1999)*, pp. 217–234, Verh. Afd. Natuurkd. 1. Reeks. K. Ned. Akad. Wet. 52, R. Neth. Acad. Arts Sci., Amsterdam, 2000.

[48] I. Klebanov, Critical field theories with osp(1|2m) symmetry. 2021, arXiv:2111.12648.

[49] G. Kozma, Reinforced random walk. In *European Congress of Mathematics*, pp. 429–443, Eur. Math. Soc., Zürich, 2013.

[50] G. Kozma and R. Peled, Power-law decay of weights and recurrence of the two-dimensional VRJP. *Electron. J. Probab.* **26** (2021), no. 82, 1–19.

[51] T. Łuczak and B. Pittel, Components of random forests. *Combin. Probab. Comput.* **1** (1992), no. 1, 35–52.

[52] R. Lyons and Y. Peres, *Probability on trees and networks*. Camb. Ser. Stat. Probab. Math. 42, 2016.

[53] M. Marcus and J. Rosen, *Markov processes, Gaussian processes, and local times*. Cambridge Stud. Adv. Math. 100, 2006.

[54] J. Martin and D. Yeo, Critical random forests. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** (2018), no. 2, 913–960.

[55] F. Merkl, A. Öry, and S. Rolles, The 'magic formula' for linearly edge-reinforced random walks. *Stat. Neerl.* **62** (2008), no. 3, 345–363.

[56] F. Merkl and S. Rolles, Bounding a random environment for two-dimensional edge-reinforced random walk. *Electron. J. Probab.* **13** (2008), no. 19, 530–-565.

[57] F. Merkl and S. Rolles, Recurrence of edge-reinforced random walk on a two-dimensional graph. *Ann. Probab.* **37** (2009), no. 5, 1679–1714.

[58] A. Mirlin, Statistics of energy levels and eigenfunctions in disordered and chaotic systems: supersymmetry approach. In *New directions in quantum chaos (Villa Monastero, 1999)*, pp. 223–298, Proc. Internat. School Phys. Enrico Fermi 143, IOS, Amsterdam, 2000.

[59] R. Pemantle, Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.* **19** (1991), no. 4, 1559–1574.

[60] R. Pemantle, Towards a theory of negative dependence. *J. Math. Phys.* **41** (2000), 1371–1390.

[61] R. Pemantle, A survey of random processes with reinforcement. *Probab. Surv.* **4** (2007), 1–79.

[62] R. Poudevigne, Monotonicity and phase transition for the VRJP and the ERRW. *J. Eur. Math. Soc.* to appear (2022).

[63] G. Ray and B. Xiao, Forests on wired regular trees. 2021, arXiv:2108.04287.

[64] C. Sabot, Polynomial localization of the 2D-vertex reinforced jump process. *Electron. Commun. Probab.* **26** (2021), Paper No. 1, 9.

[65] C. Sabot and P. Tarrès, Edge-reinforced random walk, vertex-reinforced jump process and the supersymmetric hyperbolic sigma model. *J. Eur. Math. Soc. (JEMS)* **17** (2015), no. 9, 2353–2378.

[66] C. Sabot and P. Tarrès, The $*$-Vertex-Reinforced Jump Process. 2021, arXiv:2102.08988.

[67] C. Sabot, P. Tarrès, and X. Zeng, The vertex reinforced jump process and a random Schrödinger operator on finite graphs. *Ann. Probab.* **45** (2017), no. 6A, 3967–3986.

[68] C. Sabot and X. Zeng, A random Schrödinger operator associated with the Vertex Reinforced Jump Process on infinite graphs. *J. Amer. Math. Soc.* **32** (2019), no. 2, 311–349.

[69] L. Schäfer and F. Wegner, Disordered system with n orbitals per site: Lagrange formulation, hyperbolic symmetry, and Goldstone modes. *Z. Phys. B* **38** (1980), no. 2, 113–126.

[70] T. Spencer, SUSY statistical mechanics and random band matrices. In *Quantum many body systems*, pp. 125–177, Lecture Notes in Math. 2051, Springer, Heidelberg, 2012.

[71] T. Spencer, Duality, statistical mechanics, and random matrices. In *Current developments in mathematics 2012*, pp. 229–260, 2013.

[72] T. Spencer and M. Zirnbauer, Spontaneous symmetry breaking of a hyperbolic sigma model in three dimensions. *Comm. Math. Phys.* **252** (2004), no. 1–3, 167–187.

[73] K. Symanzik, Euclidean quantum field theory. In *Local quantum field theory*, edited by R. Jost, Academic Press, New York, 1969.

[74] F. Wegner, The mobility edge problem: continuous symmetry and a conjecture. *Z. Phys. B* **35** (1979), 207–210.

[75] M. Zirnbauer, Localization transition on the Bethe lattice. *Phys. Rev. B (3)* **34** (1986), no. 9, 6394–6408.

[76] M. Zirnbauer, Fourier analysis on a hyperbolic supermanifold with constant curvature. *Comm. Math. Phys.* **141** (1991), no. 3, 503–522.

ROLAND BAUERSCHMIDT

DPMMS, University of Cambridge, Cambridge CB3 0WB, UK, rb812@cam.ac.uk

TYLER HELMUTH

Mathematical Sciences, Durham University, Durham DH1 3LE, UK,
tyler.helmuth@durham.ac.uk

THE KAC MODEL: VARIATIONS ON A THEME

FEDERICO BONETTO, ERIC CARLEN, AND MICHAEL LOSS

ABSTRACT

The Kac master equation provides a simple framework to understand systems of particles that interact through pairwise collisions. This article is a short review of results, chief among them is the approach to equilibrium for a gas of particles that undergo energy and momentum preserving collisions, as well as results on the entropy and information decay for a one-dimensional Kac system coupled to a reservoir. The principles underlying the Kac master equation can be extended to a Quantum Master Equation (QME) where the time evolution acts on density matrices and is a completely positive trace-preserving map. There is a rich set of equilibrium states and there is also a notion of propagation of chaos that leads to the Quantum Kac–Boltzmann equation. Likewise, the gap of the generator of the QME can be computed in certain special cases.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 82C22; Secondary 60J25, 60J10, 05C50

KEYWORDS

Kac model, entropy decay, heat reservoir, graph Laplacian, spectral Gap

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4010–4025
DOI 10.4171/ICM2022/194

1. INTRODUCTION

In 1956 Mark Kac published his influential paper “Foundations of kinetic theory” [14] in which he laid out a program to explain various issues about the interaction of a large number of particles. It is based on a model that, in its simplest form, describes a spatially homogenous gas of N particles undergoing pair collisions. Using this simple probabilistic model, he elucidated Boltzmann’s chaos hypotheses and its propagation and gave a derivation of what is now known as the Kac–Boltzmann equation. He also formulated a quantitative version of approach to equilibrium known as Kac’s conjecture.

The model can be described as follows: The states of the system are described by the velocities of N particles $\vec{v} = (v_1, \dots, v_N)$. If $T_{i,j}$ is the waiting time for the collision of the pair (i, j) then the first collision occurs at time $T = \min_{i,j}\{T_{i,j}\}$. If we assume that these variables are independent and that $\Pr\{T_{i,j} > t\} = e^{-2t/(N-1)}$ then T is exponentially distributed as well, with parameter N, i.e., $\Pr\{T > t\} = e^{-Nt}$. This simply reflects the fact that, on average, the collision time of a particular particle with any other particle is shortened by a factor of $1/(N-1)$. With this choice, the average number of collisions per unit time a given particle undergoes does not depend on N, as in the classical Grad–Boltzmann limit for the realistic Boltzmann equation. At time T, the pair furnishing the minimum collides and the velocity vector jumps

$$(v_1, \dots, v_i, \dots, v_j, \dots, v_N) \to (v_1, \dots, v_i^*, \dots, v_j^*, \dots, v_N).$$

The velocities v_i^* and v_j^* are then chosen according to the rule

$$(v_i, v_j) \to \big(v_i^*(\theta), v_j^*(\theta)\big) = (v_i \cos\theta - v_j \sin\theta, v_i \sin\theta + v_j \cos\theta),$$

where θ is picked randomly and uniformly in $[0, 2\pi)$. Obviously, the kinetic energy $\sum_{j=1}^N v_j^2$ is preserved during this process (we assume that all particles have the same mass 2 and the total energy is N) and hence we may describe this process as an evolution on the space of probability distributions $F(\vec{v}) \in L^1(\mathbb{S}^{N-1}(\sqrt{N}), d\sigma_N)$ (where σ_N is the uniform probability measure on the sphere), given by the Kac master equation

$$\partial_t F(\vec{v}, t) = -\mathcal{L}_N F(\vec{v}, t), \quad F(\vec{v}, 0) = F_0(\vec{v}), \quad \mathcal{L}_N = N(I - Q_N), \tag{1.1}$$

where

$$Q_N \phi(\vec{v}) = \frac{1}{\binom{N}{2}} \sum_{i<j} \frac{1}{2\pi} \int_0^{2\pi} \phi\big(\dots, v_i^*(\theta), \dots, v_j^*(\theta), \dots\big) d\theta.$$

The solution can be written as

$$F(\vec{v}, t) = e^{-Nt} \sum_{k=0}^{\infty} \frac{(Nt)^k}{k!} Q_N^k F_0(\vec{v}).$$

We shall henceforth assume that the initial condition and hence the solution are symmetric, i.e., invariant under permutations of the particle labels.

On the sphere, the v_j variables are not independent, but any finite collection is asymptotically independent as the dimension $N \to \infty$. Such an “asymptotic” independence is known as “chaos” and this gets propagated by the Kac evolution. More precisely, a sequence of distributions $\{F_N(v_1,\dots,v_N)\}_{N=1}^{\infty}$ is chaotic with marginal $f:\mathbb{R}\to\mathbb{R}_+$ if for any integer k and any bounded continuous function $\phi:\mathbb{R}^k\to\mathbb{R}$,

$$\begin{aligned}\lim_{N\to\infty}&\int_{\mathbb{S}^{N-1}(\sqrt{N})} F_N(v_1,\dots,v_N)\phi(v_1,\dots,v_k)d\sigma_N\\ &=\int_{\mathbb{R}^k}\prod_{\ell=1}^{k} f(v_\ell)\phi(v_1,\dots,v_k)dv_1\cdots dv_k.\end{aligned}$$

Kac's theorem states that if $F_{0N}(\cdot)$ is a chaotic sequence with marginal $f_0(\cdot)$ then the solution of the Kac master equation $F_N(\cdot,t)$ is chaotic with marginal $f(\cdot,t)$, which is a solution of the Kac–Boltzmann equation

$$\begin{aligned}\partial_t f(v,t)=\frac{1}{\pi}\int_{-\infty}^{\infty}dw\int_0^{2\pi}d\theta\\ \times\left[f(v\cos\theta+w\sin\theta,t)f(-v\sin\theta+w\cos\theta,t)-f(v,t)f(w,t)\right]\end{aligned}$$

with initial condition $f_0(v)$.

This model has several limitations. For one, it conserves only the energy since in one dimension particles that undergo energy and momentum preserving collisions either keep their velocities or exchange them. Moreover, in physical models, the likelihood of collision outcomes depends on momentum transfer and the scattering angles, while in this simplest version of the model, all energy conserving outcomes are equally likely.

The description of the Kac model for 3-dimensional momentum-preserving collisions is more involved. The velocities are now vectors in $\mathbb{R}^3$ and the evolution will take place on the space $L^1(\mathcal{S}_{N,E,p},d\sigma_N)$ where $\mathcal{S}_{N,E,p}$ consists of all vectors in $\mathbb{R}^{3N}$ with total energy NE and total momentum Np, i.e.,

$$\frac{1}{N}\sum_{j=1}^{N}|v_j|^2=E,\quad \frac{1}{N}\sum_{j=1}^{N}v_j=p.$$

The measure σ_N is the normalized Euclidean measure induced from $\mathbb{R}^{3N}$ on $S_{N,E,p}$. The inverse collision time for the pair (i,j) depends now on the velocities v_i,v_j of the pair, and we set it to be

$$\lambda_{i,j}=\frac{N}{\binom{N}{2}}|v_i-v_j|^{\alpha}.$$

The momentum transfer for hard spheres is $|v_i-v_j|$, i.e., $\alpha=1$. To specify the collision process, one must parametrize the collisions, and a convenient way is to set

$$\begin{aligned}v_i^*(\sigma)&=\frac{v_i+v_j}{2}+\frac{|v_i-v_j|}{2}\sigma,\\ v_j^*(\sigma)&=\frac{v_i+v_j}{2}-\frac{|v_i-v_j|}{2}\sigma,\end{aligned}\tag{1.2}$$

where $\sigma\in\mathbb{S}^2$. A particular kinematically possible collision is selected according to the following rule: in the specification of the process, there is a given nonnegative and even function

b on $[-1, 1]$ such that for any fixed $\sigma' \in S^2$, with $d\sigma$ denoting the uniform probability measure on $\mathbb{S}^2$,

$$\int_{\mathbb{S}^2} b(\sigma \cdot \sigma')d\sigma = 1 \quad \text{or, equivalently,} \quad \frac{1}{2}\int_{-1}^{1} b(t)dt = 1. \tag{1.3}$$

The most important case is

$$b(x) = 1. \tag{1.4}$$

When $\alpha = 1$ and b is given by (1.4), the Kac process models "hard sphere" or "billiard ball" collisions [15]. There are two standard parameterizations of the set of energy and momentum conserving collisions, the "σ parameterization" given by (1.2), and the "$\vec{n}$ parameterization". While the latter is often used in physics texts and in [15], the former, used here, has advantages: first, in this parameterization, b is constant, and second, it is not due to a nonconstant Jacobian relating the two parameterizations. See Appendix A.1 of [4] for more information; equation (A.18) of [4] is the formula relating the b functions for the two representations.

The generator of the Markov process is given by

$$L_{N,\alpha}F(\vec{v}) = -N\binom{N}{2}^{-1}\sum_{i<j}|v_i - v_j|^{\alpha}\big[F(\vec{v}) - [F]^{(i,j)}(\vec{v})\big]$$

where

$$[F]^{(i,j)}(\vec{v}) = \int_{S^2} b\left(\sigma \cdot \frac{v_i - v_j}{|v_i - v_j|}\right)F(R_{i,j,\sigma}\vec{v})d\sigma \tag{1.5}$$

and $(R_{i,j,\sigma}\vec{v})_k = \begin{cases} v_i^*(\sigma), & k = i, \\ v_j^*(\sigma), & k = j, \\ v_k, & k \neq i, j. \end{cases}$ The corresponding master equation then takes the form $\partial_t F = -L_{N,\alpha}F$ with the initial condition $F(\cdot, 0) = F_0(\cdot)$.

For this model, propagation of chaos was proved by Mischler and Mouhot in [16]. This is much more complicated than for the previous model since, due to the dependence on the velocities, the operator $L_{N,\alpha}$ is not uniformly bounded. For a general view on propagation of chaos, see [17].

2. KAC'S CONJECTURE

It is easy to see that the only equilibrium for the evolution given in (1.1) is the constant function. The operator $\mathcal{L}_N$ given in (1.1), as an operator in $L^2(\mathbb{S}^{N-1}(\sqrt{N}), d\sigma_N)$, is self-adjoint. The eigenvalue 0 is nondegenerate and the gap, i.e, the first nonzero eigenvalue Δ_N, is a measure for the approach to equilibrium since

$$\big\|F(t) - 1\big\|_2 \leq e^{-\Delta_N t}\|F_0 - 1\|_2.$$

Kac conjectured that there exists a constant $C > 0$ independent of N such that $\Delta_N \geq C$ for all N. This conjecture was proved in [13] and shortly thereafter in [6] the gap was explicitly computed to be

$$\Delta_N = \frac{1}{2}\frac{N+2}{N-1}.$$

The argument is an induction procedure and can be adapted to many other situations. It is maybe useful to explain it in this simple case. It is easy to see that $\mathcal{L}_N$ is self-adjoint. Then the gap is defined to be

$$\inf_{F\perp 1, \|F\|_2=1} \langle F, \mathcal{L}F\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the inner product in $L^2(\mathbb{S}^{N-1}(\sqrt{N}), d\sigma_N)$. An elementary argument shows that

$$\mathcal{L}_N = \frac{N}{N-1}\sum_{k=1}^{N} \mathcal{L}_{N-1}^{(k)},$$

where $\mathcal{L}_{N-1}^{(k)}$ is the generator for the $N-1$ particle Kac operator with particle k removed. Denote by P_k the orthogonal projection onto the space of functions on the sphere that depend only on the variable v_k. One can think of $P_k f$ as taking the average of the function f over all rotations that fix the k-axis. Then

$$\langle F, \mathcal{L}_N F\rangle = \frac{N}{N-1}\frac{1}{N}\sum_{k=1}^{N}\big\langle F, \mathcal{L}_{N-1}^{(k)}F\big\rangle = \frac{N}{N-1}\frac{1}{N}\sum_{k=1}^{N}\big\langle (F-P_kF), \mathcal{L}_{N-1}^{(k)}(F-P_kF)\big\rangle$$

since $\mathcal{L}_{N-1}^{(k)}$ does not act on functions that depend only on the variable v_k. By the induction assumption,

$$\big\langle (F-P_kF), \mathcal{L}_{N-1}^{(k)}(F-P_kF)\big\rangle \geq \Delta_{N-1}\|F-P_kF\|_2^2$$

because $F-P_kF \perp 1$ on the sphere with the variable v_k fixed. Hence we have the lower bound

$$\langle F, \mathcal{L}_N F\rangle \geq \frac{N}{N-1}\Delta_{N-1}\frac{1}{N}\sum_{k=1}^{N}\|F-P_kF\|_2^2 = \frac{N}{N-1}\Delta_{N-1}\big[\big\langle F, (I-P)F\big\rangle\big],$$

where

$$P = \frac{1}{N}\sum_{k=1}^{N} P_k.$$

In other words, the gap Δ_N is the product of the gap Δ_{N-1} multiplied by the gap of $I-P$. Now one observes that the eigenfunction of P that belong to a nonzero eigenvalue must be sums of functions of one variable. Using this, it is not very difficult to compute Λ_N, the gap of $I-P$ being

$$\Lambda_N = \frac{N-1}{N}\left(1-\frac{3}{N^2-1}\right),$$

(see, e.g., [6] and [7] for details), and hence

$$\Delta_N \geq \Delta_{N-1}\left(1-\frac{3}{N^2-1}\right)$$

and, iterating this bound using the fact that $\Delta_2 = 2$, one obtains

$$\Delta_N \geq \prod_{j=3}^{N}\left(1-\frac{3}{j^2-1}\right)\Delta_2 = \frac{1}{2}\frac{N+2}{N-1}.$$

This strategy can be used to estimate the gap for the three-dimensional momentum-preserving collisions [9]. Associated with the generator $L_{N,\alpha}$ is the quadratic form $\mathcal{E}(f, f) = -\langle f, L_{N,\alpha} f\rangle_{L^2(\mathcal{S}_{N,E,p})}$, i.e.,

$$\mathcal{E}(f, f) = \\ \frac{N}{2}\binom{N}{2}^{-1} \sum_{i<j} \int_{\mathcal{S}_{N,E,p}} \int_{\mathbb{S}^2} |v_i - v_j|^\alpha b\left(\sigma \cdot \frac{v_i - v_j}{|v_i - v_j|}\right)\left[f(\vec{v}) - f(R_{i,j,\sigma}\vec{v})\right]^2 d\sigma d\sigma_N. \tag{2.1}$$

It is easy to see that for $L_{N,\alpha}$ the constant function 1 is the only equilibrium. It is straightforward to see that $L_{N,\alpha}$ is self-adjoint on $L^2(\mathcal{S}_{N,E,p})$. The gap is the distance between the lowest and the next lowest eigenvalue of $L_{N,\alpha}$, i.e.,

$$\Delta_{N,\alpha}(E, p) = \inf\left\{\mathcal{E}(f, f) : \langle f, 1\rangle_{L^2(\mathcal{S}_{N,E,p})} = 0 \text{ and } \|f\|^2_{L^2(\mathcal{S}_{N,E,p})} = 1\right\}. \tag{2.2}$$

Using a unitary transformation mapping $L^2(S_{N,1,0})$ to $L^2(S_{N,E,p})$ (see [9]), one writes

$$\Delta_{N,\alpha}(E, p) = \left(E - |p|^2\right)^{\alpha/2} \Delta_{N,\alpha}(1, 0), \tag{2.3}$$

and we call $\Delta_{N,\alpha}(1, 0)$ the "spectral gap for the Kac model."

Theorem 2.1 (Spectral gap for the Kac model with $0 \le \alpha \le 2$). *For each continuous non-negative even function b on $[-1, 1]$ satisfying (1.3) and for each $\alpha \in [0, 2]$, there is a strictly positive constant K depending only on b and α, and explicitly computable, such that*

$$\Delta_{N,\alpha} \ge K > 0$$

for all N. In particular, this is true with b given by (1.4) and $\alpha = 1$, the 3-dimensional hard sphere Kac model.

This theorem was conjectured by Kac [15]. The proof is considerably more involved than that for the one-dimensional gas, and we refer the reader to [9] for the details. The fundamental idea is to find a replacement for the operator $(I - P)$. On the space $L^2(S_{N,1,0})$ consider the operator

$$\hat{L}_{N,\alpha} f = -\frac{1}{N} \sum_{k=1}^{N} \left[\frac{N^2 - (1 + |v_k|^2)N}{(N-1)^2}\right]^{\alpha/2} [f - P_k f], \tag{2.4}$$

where P_k is the orthogonal projection defined by the map $\phi(\vec{v}) \to f(v_k)$ given defined by the relation

$$\int_{S_{N,1,0}} \phi(\vec{v}) g(v_k) d\sigma_N = \int_{S_{N,1,0}} f(v_k) g(v_k) d\sigma_N.$$

This operator is again self-adjoint on $L^2(S_{N,1,0})$, has 0 as its lowest eigenvalue, and we denote its gap by $\widehat{\Delta}_{N,\alpha}$. The reason for this operator is that it provides again an inductive approach to the whole problem. We have

Theorem 2.2. *For all $N \ge 3$,*

$$\Delta_{N,\alpha} \ge \frac{N}{N-1} \Delta_{N-1,\alpha} \widehat{\Delta}_{N,\alpha}. \tag{2.5}$$

In a further step, one proves, and this is the main work,

Theorem 2.3. *For all $N \geq 3$ and all $\alpha \in [0,2]$, $\widehat{\Delta}_{N,\alpha} > 0$. Moreover, there is a constant C independent of N such that*

$$\widehat{\Delta}_{N,\alpha} \geq 1 - \frac{1}{N} - \frac{C}{N^{3/2}}. \tag{2.6}$$

As a corollary, one obtains for any N large,

$$\Delta_{N,\alpha} \geq \prod_{j=N_0+1}^{N} \left(1 - \frac{C}{j^{1/2}(j-1)}\right) \Delta_{N_0,\alpha},$$

where N_0 is chosen such that $1 - \frac{C}{N_0^{1/2}(N_0-1)} > 0$. One now observes that

$$\lim_{N\to\infty} \prod_{j=N_0+1}^{N} \left(1 - \frac{C}{j^{1/2}(j-1)}\right) =: D > 0.$$

3. APPROACH TO EQUILIBRIUM IN ENTROPY

For simplicity we restrict ourselves to the one-dimensional case. To measure the approach to equilibrium in terms of the gap is unsatisfactory because the square norm of a probability distribution in general grows exponentially with the dimension. The right quantity is the entropy, relative to the uniform measure on the sphere,

$$S_N(F) = \int_{\mathbb{S}^{N-1}(\sqrt{n})} F(\vec{v}) \log F(\vec{v}) d\sigma_N,$$

which is proportional to the number of particles for the case of approximate independence. For the connection between the entropy and a strengthened notion of chaos know as entropic chaos, we refer to [5].

The dissipation is defined as

$$D_N(F) := -\frac{d}{dt} S_N\big(F(t)\big)\Big|_{t=0} = \int_{\mathbb{S}^{N-1}(\sqrt{N})} \mathcal{L}_N F \log F d\sigma_N.$$

It is not hard to see that the dissipation is positive. The entropy production is then defined by

$$\Gamma_N = \inf_F \frac{D_N(F)}{S_N(F)}.$$

It was shown by Villani [18] that

$$\Gamma_N \geq \frac{2}{N-1}, \tag{3.1}$$

which leads to exponential decay of the entropy

$$S_N\big(F(t)\big) \leq e^{-\frac{2}{N-1}t} S(F_0).$$

Obviously, the rate vanishes as the particle number tends to infinity. It was shown by Einav [11] that the entropy production estimate (3.1) is essentially correct by producing a trial function that yields an upper bound very close to Villani's estimate. While such considerations do not preclude the possibility that the entropy stays essentially constant for a time of order 1 and then decays exponentially with a rate independent of N, this seems somewhat unlikely. If one imagines a gas where few particles contain most of the energy, it will presumably take a long time until these particles have imparted their energy to the others and the system is near some equilibrium. This intuition was used in [11] to produce a state with very small entropy production.

A reasonable approach is to consider a system of M particles that interact with a reservoir of N particles that are initially at equilibrium. One envisions that N is large when compared to M. The Kac evolution for this situation can be written as

$$\partial_t F = -(\lambda_S \mathcal{L}_M F + \lambda_R \mathcal{L}_N + \mu \mathcal{I}_{M,N})F,$$

where $F = F(\vec{v}, \vec{w})$ and $\vec{v} = (v_1, \dots, v_M)$ are the velocities of the particles in the system and $\vec{w} = (w_1, \dots, w_N)$ are the velocities of the particles in the reservoir; λ_S, λ_R are constants and the term $\mu \mathcal{I}_{M,N}$, describing the interaction between the system and the reservoir, is given by

$$\mu \mathcal{I}_{M,N} F = \frac{\mu}{N} \sum_{i=1}^{M} \sum_{j=1}^{N} R_{i,j} F,$$

where

$$R_{i,j} F(\vec{v}, \vec{w}) = \frac{1}{2\pi} \int_0^{2\pi} F\big(\dots, v_i^*(\theta), \dots, w_j^*(\theta), \dots\big) d\theta.$$

The factors μ/N are chosen in such a way that the average collision time between a fixed particle in the system with any particle in the reservoir is of order μ. To keep the problem simple, we consider distributions on $\mathbb{R}^{M+N}$ and assume that the initial condition is of the form

$$F_0(\vec{v}, \vec{w}) = f_0(\vec{v}) e^{-\pi|\vec{w}|^2},$$

where f_0 is normalized. The quantity of interest is the relative entropy of the system at time t which is defined by

$$S\big(f(t)|\gamma\big) = \int_{\mathbb{R}^M} f(\vec{v}, t) \log \frac{f(\vec{v}, t)}{\gamma} d\vec{v},$$

where

$$f(\vec{v}, t) = \int_{\mathbb{R}^N} F(\vec{v}, \vec{w}, t) d\vec{w}$$

and $\gamma = e^{-\pi|\vec{v}|^2}$. The following theorem is proved in [1]

Theorem 3.1. *For any positive integers N, M, we have that*

$$S\big(f(t)|\gamma\big) \le \left(\frac{M}{N+M} + e^{-\frac{\mu(N+M)}{2N} t} \frac{N}{N+M} \right) S(f_0|\gamma).$$

This theorem states that for $N \gg M$ the entropy decays with a rate approximately $\mu/2$ to a very small fraction of the original entropy. The original proof in [1] is rather cumbersome using Brascamp–Lieb inequalities. Another, simpler, proof based on the concept of information can be found in [2]. One should emphasize that the reservoir does not stay in equilibrium in this process, and it is interesting to compare the process with a system interacting with a thermostat, i.e., a system interacting with an "infinite reservoir." Such a model can be described by the following master equation:

$$\partial_t f = -\lambda \mathcal{L}_M f - \mu \sum_{j=1}^{M} (I - R_j) f,$$

where

$$R_j f(\vec{v}) = \int_{\mathbb{R}} dw \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{-\pi(-v_j \sin(\theta) + w \cos(\theta))^2} \times f\big(v_1, \dots, v_j \cos(\theta) + w \sin(\theta), \dots, v_M\big).$$

This describes the collision between the particle with label j and a particle randomly picked from the Gaussian ensemble with temperature $\frac{1}{2\pi}$. It is easy to see that $e^{-\pi|\vec{v}|^2}$ is the equilibrium for this process.

The following theorem is proved in [3]

Theorem 3.2. *The relative entropy with respect to γ satisfies the estimate*

$$S\big(f(t)|\gamma\big) \le e^{-\frac{\mu}{2}t} S(f_0|\gamma).$$

It is satisfying to see that the result in Theorem 3.1 takes the form of Theorem 3.2 as $N \to \infty$.

4. A QUANTUM KAC MODEL

In [8], the list of Kac models was extended by a Quantum Markov Semigroup that describes pair collisions of quantum particles. The energy of a single particle is given by a Hamiltonian h on a Hilbert space $\mathcal{H}$ which we take to be finite dimensional for simplicity. A state of the system of N particles is described by a density matrix ρ on $\otimes^N \mathcal{H}$, i.e., a self-adjoint positive trace class operator with unit trace. One specifies the binary collisions by a family of unitary operators $U(\sigma)$ on the two particle Hilbert space $\mathcal{H}_2 = \mathcal{H} \otimes \mathcal{H}$ that commute with $H_2 = h \otimes I + I \otimes h$. Here σ lives in a measure space $(\mathcal{C}, \nu)$. The precise conditions will be given below. The collision operators $\mathcal{Q} : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ is given by

$$\mathcal{Q}(A) = \int_{\mathcal{C}} d\nu(\sigma) U(\sigma) A U^*(\sigma),$$

where $\mathcal{B}(\mathcal{H}_2)$ denotes the space of bounded operators on $\mathcal{H}_2$.

The measure ν is a probability measure and it is easily seen that $\mathcal{Q}$ is a trace-preserving map that is positivity preserving, in fact completely positive. Since $U(\sigma)$ commutes with H_2, this collision process preserves energy, i.e., if all the eigenstates of ρ have the same energy so does $\mathcal{Q}(\rho)$. Naturally, one wants that the collision of particle 1 with particle 2 and the collision of particle 2 with particle 1 leads to the same result. If V denotes the swap operation

$$V(\phi \otimes \psi) = \psi \otimes \phi$$

then one imposes the condition that

$$\{U(\sigma) : \sigma \in \mathcal{C}\} = \{VU(\sigma)V^* : \sigma \in \mathcal{C}\}$$

and the map $\sigma \to \sigma'$ is such that $VU(\sigma)V^* = U(\sigma')$ is a measurable transformation that leaves ν invariant. It is also desirable that the collision satisfies local reversibility, i.e., that

$$\{U(\sigma) : \sigma \in \mathcal{C}\} = \{U^*(\sigma) : \sigma \in \mathcal{C}\}$$

and the map $\sigma \to \sigma'$ is measurable and leaves ν invariant. One easily sees that for any two operators A, B on $\mathcal{H}_2$ one has

$$\mathrm{Tr}\big(A^*\mathcal{Q}(B)\big) = \mathrm{Tr}\big(\mathcal{Q}(A)^*B\big),$$

i.e., the operation is self-adjoint on the Hilbert space $\mathcal{B}(\mathcal{H}_2)$ with inner product $(A, B) = \mathrm{Tr}(A^*B)$. We call $(\mathcal{C}, U, \nu)$ satisfying these conditions a collision specification.

Denote by $\mathcal{A}_2$ the commutative subalgebra of $\mathcal{B}(\mathcal{H}_2)$ consisting of all operators that are of the form $f(H_2)$ where $f : \sigma(H_2) \to \mathbb{C}$ is a continuous bounded function. Obviously, $\mathcal{A}_2$ is a subset of $\{U(\sigma) : \sigma \in \mathcal{C}\}'$, the commutant of $\{U(\sigma) : \sigma \in \mathcal{C}\}$. We shall require that the two particle collisions are ergodic, that is,

$$\mathcal{A}_2 = \{U(\sigma) : \sigma \in \mathcal{C}\}'.$$

4.1. Example

The following example taken from [8] is useful for understanding these concepts. For the simplest possible example, take $\mathcal{H} = \mathbb{C}^2$, so that $\mathcal{H}_2 = (\mathbb{C}^2)^{\otimes 2}$. Define the single particle Hamiltonian h by $h = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right]$. Identify $\mathbb{C}^2 \otimes \mathbb{C}^2$ with $\mathbb{C}^4$ using the basis

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The standard physics notation for this basis is simply

$$|00\rangle, \quad |10\rangle, \quad |01\rangle, \quad |11\rangle, \tag{4.1}$$

which will be useful. With this identification of $\mathbb{C}^2 \otimes \mathbb{C}^2$ with $\mathbb{C}^4$,

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \otimes \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{bmatrix} =: A \otimes B \text{ is represented by } \begin{bmatrix} b_{1,1}A & b_{1,2}A \\ b_{2,1}A & b_{2,2}A \end{bmatrix}.$$

(Switching the order of the second and third basis elements swaps the roles of A and B in the block matrix representation of the tensor product $A \otimes B$.)

In this basis,

$$H_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes I + I \otimes \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Therefore, the spectrum of $H_2 = \{0, 1, 2\}$ and

$$P_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad P_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now define $\mathcal{C} = \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$ identifying each copy of $\mathbb{S}^1$ with the unit circle in $\mathbb{C}$ so that the general point in $\sigma \in \mathcal{C}$ has the form $\sigma = (e^{i\varphi}, e^{i\theta}, e^{i\psi}, e^{i\eta})$. Then define

$$U(\sigma) := \begin{bmatrix} e^{i\theta} & 0 & 0 & 0 \\ 0 & e^{i\psi}\cos\theta & -e^{i\varphi}\sin\theta & 0 \\ 0 & e^{-i\varphi}\sin\theta & e^{-i\psi}\cos\theta & 0 \\ 0 & 0 & 0 & e^{i\eta} \end{bmatrix}.$$

Choosing ν to be the uniform probability measure on $\mathcal{C}$ gives us a collision specification $(\mathcal{C}, U, \nu)$.

A simple computation shows that for every operator A on $\mathcal{H}_2 = \mathbb{C}^2 \otimes \mathbb{C}^2$ identified as the 4×4 matrix with entries $a_{i,j}$ using the basis (4.1),

$$\begin{aligned} \mathcal{Q}(A) = \int_{\mathcal{C}} d\nu(\sigma) U(\sigma) A U^*(\sigma) &= \begin{bmatrix} a_{1,1} & 0 & 0 & 0 \\ 0 & \frac{1}{2}(a_{2,2} + a_{3,3}) & 0 & 0 \\ 0 & 0 & \frac{1}{2}(a_{2,2} + a_{3,3}) & 0 \\ 0 & 0 & 0 & a_{4,4} \end{bmatrix} \\ &= a_{1,1} P_0 + \frac{a_{2,2} + a_{3,3}}{2} P_1 + a_{4,4} P_2 \in \mathcal{A}_2. \end{aligned}$$

Therefore,

$$\{U(\sigma) : \sigma \in \mathcal{C}\}' \subset \operatorname{ran}(\mathcal{Q}) \subset \mathcal{A}_2 \subset \{U(\sigma) : \sigma \in \mathcal{C}\}',$$

showing that $(\mathcal{C}, U, \nu)$ is ergodic.

Using these preliminaries, it is now straightforward to write the corresponding Quantum Master Equation (QME) as

$$\partial_t \rho = -\mathfrak{L}_N(\rho),$$

with

$$\mathfrak{L}_N(A) = N \binom{N}{2}^{-1} \sum_{i<j} \left[A - \mathcal{Q}_{i,j}(A)\right] \tag{4.2}$$

and where the unitaries in the definition of $\mathcal{Q}_{i,j}$ act nontrivially only on the ith and jth factors in the tensor product $\otimes^N \mathcal{H}$. This is a trace preserving completely positive map, i.e., a Quantum Evolution.

4.2. Propagation of chaos

A density matrix is symmetric if it is invariant under the swap operation between any two factors in the tensor product $\otimes^N \mathcal{H}$. A sequence of symmetric density matrices $\{\rho_N\}_{N=1}^{\infty}$ is chaotic with marginal ϱ, or in short ϱ-chaotic, if

$$\lim_{N\to\infty} \mathrm{Tr}_{2,\dots,N}\rho_N = \varrho \quad \text{and} \quad \lim_{N\to\infty} \mathrm{Tr}_{k+1,\dots,N}\rho_N = \otimes^k \varrho,$$

where $\mathrm{Tr}_{k+1,\dots,N}$ is the trace taken in the factors $k+1,\dots,N$. A trivial example of a chaotic sequence is $\otimes^N \varrho$, but one can also construct chaotic sequences that have a sharply defined energy for large N.

We have (see [8])

Theorem 4.1. *Let $\{U(\sigma) : \sigma \in \mathcal{C}\}$ be a set of collision operators and let ν be a given Borel probability measure on $\mathcal{C}$. Let $\mathfrak{L}_N$ be defined in terms of these as in (4.2). Then the semigroup $\mathcal{P}_{N,t} = e^{t\mathfrak{L}_N}$ propagates chaos for all t, meaning that if $\{\rho_N\}_{N\in\mathbb{N}}$ is a ϱ-chaotic sequence then, for each t, $\{\mathcal{P}_{N,t}\varrho_N\}_{N\in\mathbb{N}}$ is a $\varrho(t)$-chaotic sequence for some $\varrho(t) = \lim_{N\to\infty}(\mathcal{P}_{N,t}\varrho_N)^{(1)}$, where in particular this limit of the one-particle marginal exists and is a density matrix.*

As expected, the marginal density matrix $\varrho(t)$ satisfies a Quantum Kac–Boltzmann equation

$$\frac{d}{dt}\varrho(t) = 2\big(\varrho(t) \star \varrho(t) - \varrho(t)\big),$$

where, quite generally, for operators in $\mathcal{B}(\mathcal{H})$,

$$A \star B = \mathrm{Tr}_2\big[d\nu(\sigma)U(\sigma)[A \otimes B]U^*(\sigma)\big] = \mathrm{Tr}_2\big[\mathcal{Q}(A \otimes B)\big]$$

is the Quantum Wild Convolution.

4.3. Equilibrium states

An equilibrium density matrix for the evolution (4.2) is given by all those density matrices ρ_N that satisfy

$$\mathfrak{L}_N(\rho_N) = 0.$$

Recall that the N-particle Hamiltonian is $H_N = \sum_{j=1}^N h_j$ where h_j is the single particle Hamiltonian acting on the jth factor. List the eigenvalues of h as $e_1,\dots,e_K$ counting their multiplicities and denote the corresponding eigenvectors by $\phi_1,\dots,\phi_K$. Using the multi-index notation $\alpha = (\alpha_1,\dots,\alpha_N)$ where $\alpha_j \in \{1,\dots,K\}$, $j = 1,\dots,N$, the eigenvalues of H_N are given by $E_\alpha = \sum_{j=1}^N e_{\alpha_j}$ and $\Psi_\alpha = \phi_{\alpha_1} \otimes \cdots \otimes \phi_{\alpha_N}$ are the eigenvectors.

It is not very difficult to show that the set $\mathfrak{C}_N$ of equilibrium states form a commutative von Neumann algebra, and hence it is generated by the minimal projections. The algebra $\mathcal{A}_N$ consisting of all operators of the form $f(H_N)$ where f is a bounded continuous function is a subalgebra of $\mathfrak{C}_N$ and it is generated by the spectral projections of the Hamiltonian H_N. Define two multiindices α, α' to be adjacent if for some pair (i,j), $e_{\alpha_i} + e_{\alpha_j} = e_{\alpha'_i} + e_{\alpha'_j}$ and $\alpha_k = \alpha'_k, k \neq i, j$. With this notion of adjacency the multiindices α form a graph, the adjacency graph $\mathcal{G}_N$. We denote by $\gamma_1, \ldots, \gamma_n$ the connected components of $\mathcal{G}_N$. In [8] the following theorem is proved.

Theorem 4.2. *The minimal projections of $\mathfrak{C}_N$ are in one-to-one correspondence with the connected components of the adjacency graph $\mathcal{G}_N$ and are given by*

$$\mathcal{P}_k = \sum_{\alpha \in \gamma_k} |\Psi_\alpha\rangle\langle\Psi_\alpha|.$$

Ergodicity in our context is the notion that the only equilibrium states of the quantum Kac model are given by the algebra $\mathcal{A}_N$. By the above theorem, this is the case if the connected components of the adjacency graph are determined by the energies of the Hamiltonian H_N. The occupation number representation is useful in this context. We write $E_\alpha = \sum_{j=1}^{K} k_j(\alpha) e_j$ where $k_j(\alpha)$ denotes the number of times the index j occurs in α. Thus, if the energies of h, $\{e_1, \ldots, e_K\}$, are rationally independent then any eigenvalue of H_N is uniquely determined by the occupation numbers $k_1(\alpha), \ldots, k_K(\alpha)$ (see below). Hence, in this case we have that the minimal projections of $\mathfrak{C}_N$ are eigenprojections of H_N and hence $\mathfrak{C}_N = \mathcal{A}_N$.

Here is an example where $\mathfrak{C}_N \neq \mathcal{A}_N$. Assume the single particle Hamiltonian has the eigenvalues $1, 2, 4$ with the corresponding eigenvectors ψ_1, ψ_2, ψ_3. Then pick n_1 to be even integers and set

$$n_2 = N - \frac{3}{2} n_1, \quad n_3 = \frac{1}{2} n_1.$$

Then

$$n_1 + 2n_2 + 4n_3 = 2N, \quad n_1 + n_2 + n_3 = N.$$

The number $e = 2N$ is an eigenvalue of the Hamiltonian H_N and it is degenerate. The eigenvectors are of the form $\psi_{\alpha_1} \otimes \cdots \otimes \psi_{\alpha_N}$ where $\alpha_j \in \{1, 2, 3\}$. We set $\alpha = (\alpha_1, \ldots, \alpha_N)$ and let $n_1(\alpha)$ be the number of ψ_1 factors, $n_2(\alpha)$ the number of ψ_2 factors, and $n_3(\alpha)$ the number of ψ_3 factors. If α and β are adjacent then the condition $e_{\alpha_k} + e_{\alpha_\ell} = e_{\beta_k} + e_{\beta_\ell}$ implies that either $e_{\alpha_k} = e_{\beta_k}$ and $e_{\alpha_\ell} = e_{\beta_\ell}$ or $e_{\alpha_k} = e_{\beta_\ell}$ and $e_{\alpha_\ell} = e_{\beta_k}$; anything else is not possible. Hence for any of the indices α and β to be adjacent, we must have that $n_1(\alpha) = n_1(\beta), n_2(\alpha) = n_2(\beta), n_3(\alpha) = n_3(\beta)$. Thus, if these triples are different, but with the same N and e, the two states are not adjacent and hence $\mathcal{G}_{e,N}$, the adjacency graph for a fixed energy e, is not connected. The number of elements in a connected component of $\mathcal{G}_{e,N}$ is given by

$$\frac{N!}{n_1! n_2! n_3!},$$

where $N = n_1 + n_2 + n_3$.

The Quantum Kac Master Equation, being a completely positive map, can be written in terms of Kraus operators (see [10]). The collision specifications yield that the Kraus operators are self-adjoint, and the hence the QKME can be brought into a Lindblad form $\partial_t \rho = \sum_k [V_k, [V_k, \rho]]$. An example, closely related to Example 4.1, is the following Lindblad equation $\partial_t \rho = L_N(\rho)$, where

$$L_N(\rho) = \frac{1}{N-1} \sum_{[\alpha,\beta] \in \mathcal{E}_N} \big[L_{\alpha,\beta}, [L_{\alpha,\beta}, \rho]\big].$$

Here, $\mathcal{E}_N$ is the edge set of the graph $\mathcal{G}_N$. With $F_{\alpha,\beta} = |\Psi_\alpha\rangle\langle\Psi_\beta|$, the "angular" momentum operators $L_{\alpha,\beta}$ are given by

$$L_{\alpha,\beta} = F_{\alpha,\beta} - F_{\beta,\alpha}.$$

Note that in Example 4.1 the operator given by the collision specifications is L_N up to a factor that commutes with the angular momenta $L_{\alpha,\beta}$. The interesting point is that the gap of the generator L_N is given by the gap of the combinatorial or graph Laplacian on $\mathcal{G}_N$. To describe this we shall assume that the eigenvalues of h are rationally independent. The energies of the Hamiltonian H_N are given by

$$E(\alpha) = \sum_{j=1}^{K} k_j(\alpha) e_j,$$

where the $k_j(\alpha)$ are integers and $\sum_{j=1}^{K} k_j(\alpha) = N$. Since the e_js are rationally independent, the eigenvalues of H_N are in a one-to-one correspondence with the "occupation numbers" $\mathbf{k}(\alpha) = (k_1(\alpha), \dots, k_K(\alpha))$. Next, note that $\mathbf{k}(\alpha) = \mathbf{k}(\beta)$ if and only if α and β are related by a finite sequence of pair transpositions. Thus, $H\Psi_\alpha = E\Psi_\alpha$ and $H_N\Psi_\beta = E\Psi_\beta$ if and only if α and β are adjacent in $\mathcal{G}_N$. In other words, there is a one-to-one correspondence between the eigenspaces of H_N and the connected components of $\mathcal{G}_N$. This is precisely the case in Example 4.1. Indeed, the energies of $\mathcal{H}_N$ are given by

$$E(\alpha) = k_1(\alpha) \times 0 + k_2(\alpha) \times 1,$$

and with $k_1(\alpha) + k_2(\alpha) = N$ the occupation numbers determine $E(\alpha)$ uniquely. The vertices of the graph $\mathcal{G}_N$ are given by multiindices of length N consisting of 1s and 0s and if two indices are connected then one can be transformed into the other by a series of transpositions. Thus, in this case multiindices are adjacent if and only if they have the same number of 0s and hence 1s and, clearly, the occupation numbers determine the connected components of $\mathcal{G}_N$ uniquely. The subgraphs given by the connected components are well known under the name Johnson Graphs or Johnson Association Schemes. In particular, the eigenvalues of the graph Laplacian of these graphs are all known as are the eigenvectors [12]. The following theorem is a special case of a result that will appear in [10].

Theorem 4.3. *Assume that the eigenalues of h are rationally independent and $N > 2$. Then the gap of $\mathcal{L}_N$ is*

$$\frac{2N}{N-1}.$$

FUNDING

This work was partially supported by U.S. NSF grants DMS-1907643 (F.B.), DMS-2055282 (E.C.) and DMS-1856645 (M.L.).

REFERENCES

[1] F. Bonetto, A. Geisinger, M. Loss, and T. Ried, Entropy decay for the Kac evolution. *Comm. Math. Phys.* **363** (2018), no. 3, 847–875.

[2] F. Bonetto, R. Han, and M. Loss, Decay of information for the Kac evolution. *Ann. Henri Poincaré* **22** (2021), no. 9, 2975–2993.

[3] F. Bonetto, M. Loss, and R. Vaidyanathan, The Kac model coupled to a thermostat. *J. Stat. Phys.* **156** (2014), no. 4, 647–667.

[4] E. A. Carlen, J. A. Carrillo, and M. C. Carvalho, Strong convergence towards homogeneous cooling states for dissipative Maxwell models. *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **26** (2009), no. 5, 1675–1700.

[5] E. A. Carlen, M. C. Carvalho, J. Le Roux, M. Loss, and C. Villani, Entropy and chaos in the Kac model. *Kinet. Relat. Models* **3** (2010), no. 1, 85–122.

[6] E. Carlen, M. C. Carvalho, and M. Loss, Many-body aspects of approach to equilibrium. In *Séminaire: Équations aux Dérivées Partielles, 2000–2001, Sémin. Équ. Dériv. Partielles, pages Exp* No. XIX, 12, École Polytech, Palaiseau, 2001.

[7] E. A. Carlen, M. C. Carvalho, and M. Loss, Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta Math.* **191** (2003), no. 1, 1–54.

[8] E. A. Carlen, M. C. Carvalho, and M. P. Loss, Chaos, ergodicity, and equilibria in a quantum Kac model. *Adv. Math.* **358** (2019), 106827, 50 pp.

[9] E. Carlen, M. Carvalho, and M. Loss, Spectral gaps for reversible Markov processes with chaotic invariant measures: The Kac process with hard sphere collisions in three dimensions. *Ann. Probab.* **48** (2020), no. 6, 2807–2844.

[10] E. Carlen and M. Loss, in preparation.

[11] A. Einav, On Villani's conjecture concerning entropy production for the Kac master equation. *Kinet. Relat. Models* **4** (2011), no. 2, 479–497.

[12] Y. Filmus, An orthogonal basis for functions over a slice of the Boolean hypercube. *Electron. J. Combin.* **23** (2016), no. 1, Paper 1.23, 27.

[13] E. Janvresse, Spectral gap for Kac's model of Boltzmann equation. *Ann. Probab.* **29** (2001), no. 1, 288–304.

[14] M. Kac, Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*, pp. 171–197, University of California Press, Berkeley–Los Angeles, CA, 1956.

[15] M. Kac, *Probability and related topics in physical sciences (Proceedings of the Summer Seminar, Boulder, Colorado)*. Lect. Appl. Math. 1957, Interscience Publishers, London–New York, 1959. With special lectures by G. E. Uhlenbeck, A. R. Hibbs, and B. van der Pol.

[16] S. Mischler and C. Mouhot, About Kac's program in kinetic theory. *C. R. Math. Acad. Sci. Paris* **349** (2011), no. 23–24, 1245–1250.

[17] A.-S. Sznitman, Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, pp. 165–251, Lecture Notes in Math. 1464, Springer, Berlin, 1991.

[18] C. Villani, Cercignani's conjecture is sometimes true and always almost true. *Comm. Math. Phys.* **234** (2003), no. 3, 455–490.

FEDERICO BONETTO

School of Mathematics, Georgia Institute of Technology, 686 Cherry St., Atlanta, GA 30332-0160, USA, bonetto@math.gatech.edu

ERIC CARLEN

Department of Mathematics, Hill Center, Rutgers University, 110 Frelinghuysen Road Piscataway NJ 08854-8019, carlen@math.rutgers.edu

MICHAEL LOSS

School of Mathematics, Georgia Institute of Technology, 686 Cherry St., Atlanta, GA 30332-0160, USA, loss@math.gatech.edu

ON THE ENERGY OF DILUTE BOSE GASES

SØREN FOURNAIS AND JAN PHILIP SOLOVEJ

ABSTRACT

A fundamental problem in quantum mechanics is to understand the structure and the energy of ground states of interacting systems of many particles. The quantum correlations in ground states or low lying energy states are supposed to explain phenomena such as superfluidity or superconductivity.

A long-standing conjecture in mathematical physics has been to establish a universal two-term asymptotic formula for the ground state energy of a system of bosons in the dilute limit of low density predicted by the theory of superfluidity. We discuss a recent proof of this formula.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 81V73; Secondary 81V70, 82B10

KEYWORDS

Quantum many-body theory, Bose gases, superfluidity

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4026–4038
DOI 10.4171/ICM2022/139

1. INTRODUCTION

Physical systems of many interacting particles are highly complex and extremely difficult to analyze due to the correlations between the particles.

Many-particle *quantum* systems are particularly difficult because of the added complexity caused by entanglement leading to quantum correlations. Exotic phenomena such as superfluidity and superconductivity are due to such quantum correlations. We are still very far from being able to give a full mathematical explanation of these phenomena, but recent years have seen some progress on these very fundamental issues.

We will give a short account of progress on a particularly fundamental aspect of the analysis of quantum many-particle systems. The question is to understand the ground state, i.e., the state of lowest energy, of an interacting quantum system of identical particles in three dimensions. Consider a large, i.e., thermodynamic, system of density $\rho > 0$ of identical nonrelativistic particles. The only assumption we make about the interaction between these particles is that it is a repulsive two-body interaction. The question is what is the ground state energy density of such a system. In a seminal paper from 1957 [12], Lee, Huang, and Yang predicted that there is a universal asymptotic formula for the energy density $e(\rho)$ in the dilute limit given by

$$e(\rho) = (\hbar^2/2m)4\pi\rho^2 a\left(1 + \frac{128}{15\sqrt{\pi}}\sqrt{\rho a^3} + o(\sqrt{\rho a^3})\right). \tag{1.1}$$

The formula is referred to as universal because there is a two-term asymptotic expansion depending on the interaction potential through only one parameter, *the scattering length* a. We will define it below. Above $\hbar$ is Planck's constant and m is the mass of the particles. The diluteness of the system is measured in terms of the dimensionless parameter ρa^3, i.e., the expected number of particles in a cube of size a. In [12] the prediction was based on a heuristic analysis of the case of a hard core potential of radius a, i.e., particles move freely except that they cannot get closer than a distance a from each other. Formula (1.1) can also be understood heuristically from Bogolubov's theory of superfluidity from 1947 [4]. We will describe a recent proof [8,9] that establishes the formula for a very large class of repulsive interaction potentials. The Lee–Huang–Yang formula has been tested experimentally on a gas of ^{7}Li atoms in [18]. Here the coefficient which in the formula is $\frac{128}{15\sqrt{\pi}} = 4.81$ was measured to be 4.5 ± 0.7 in excellent agreement with the theoretical value.

This paper is organized as follows. In Section 2 we explain the mathematical formulation of many-particle quantum systems with two-body interactions. We, in particular, introduce the thermodynamic limit of the ground state energy density for translation-invariant systems. In Section 3 we consider the simple case of just two particles and use it to introduce the scattering length and give the precise statements of the main theorems. In Section 4 we briefly introduce the second quantized formalism and give the heuristics behind Bogolubov's approximation that leads to his theory of superfluidity for weakly interacting Bose gases. We will also explain how the Lee–Huang–Yang (LHY) formula can be heuristically derived from the Bogolubov approximation. In Section 5 we sketch the ingredients of the rigorous proof of the LHY formula. The details of what is being discussed here can be found in [8,9].

2. QUANTUM MANY-PARTICLE HAMILTONIANS WITH 2-BODY INTERACTIONS

We consider N identical particles moving in a box $\Omega = [0, L^3]$ described by the basic two-body Hamiltonian

$$H_N = \sum_{i=1}^{N} -\Delta_i + \sum_{1\leq i<j\leq N} V(x_i - x_j) \tag{2.1}$$

acting as a self-adjoint operator on an appropriate domain on $\mathcal{H}_N = L^2(\Omega^N) = \bigotimes^N L^2(\Omega)$. We have chosen units such that $\hbar = 2m = 1$, where m is the mass of the particles. The first sum in the Hamiltonian describes the kinetic energy of the nonrelativistic particles. For simplicity, we may assume that we have periodic boundary conditions such that Ω represents a torus, but, as we shall see, this is not really important. In the periodic case we see that the Hamiltonian above is translation invariant. The second sum in the Hamiltonian is the interaction. The only assumption we make about the interaction potential V is that it is repulsive, i.e., it could be any measurable function $V : \mathbb{R}^3 \to [0, \infty]$, spherically symmetric, and has suffient decay. For simplicity, we will here assume that it has compact support, but this can be relaxed considerably (see [9]).

A particularly interesting example is the hard core potential

$$V(x) = \begin{cases} 0, & |x| > a, \\ \infty, & |x| \leq a. \end{cases} \tag{2.2}$$

Since the Hamiltonian is symmetric under interchange of particles, it could also be restricted to the fully symmetric subspace $\mathcal{H}_N^{\mathrm{B}} = \bigvee^N L^2(\Omega)$ or the fully antisymmetric subspace $\mathcal{H}_N^{\mathrm{F}} = \bigwedge^N L^2(\Omega)$. In the first case we describe bosons, while in the second case we describe fermions.

The spectrum of the operator H_N will be discrete. The lowest eigenvalue is referred to as the *ground state energy*

$$E(N, \Omega, V) = \inf \mathrm{Spec}_{\mathcal{H}_N} H_N = \inf \mathrm{Spec}_{\mathcal{H}_N^{\mathrm{B}}} H_N. \tag{2.3}$$

Note that for the ground state energy it does not play any role whether we consider the full space $\mathcal{H}_N$ or the bosonic subspace $\mathcal{H}_N^{\mathrm{B}}$: By a classical theorem, the ground state eigenvector will be symmetric. As physical particles are either bosons or fermions, we refer to our analysis as the ground state of a Bose gas, but from a mathematical point of view this restriction is not important. Nevertheless, we shall in Section 4.1 use the second quantized techniques developed particularly for Bose gases.

The important quantity that we will analyze is the thermodynamic limit of the ground state energy density

$$e(\rho, V) = \lim_{L\to\infty, N/L^3=\rho} L^{-3} E(N, \Omega, V), \tag{2.4}$$

where we have fixed the density of particles to be $\rho \geq 0$. It is not difficult to see that the limit in (2.4) exists and it is indeed independent on the type of boundary condition that was chosen for the Hamiltonian. An alternative formulation would be not to fix the density but to introduce a chemical potential μ and define

$$e_{gc}(\mu, V) = \lim_{L\to\infty,} L^{-3} \inf_{N\geq 0}\big(E(N,\Omega,V) - \mu N\big). \tag{2.5}$$

This is referred to as the grand canonical (gc) formalism. The two "energy densities" are related by a Legendre transform

$$e(\rho, V) = \sup_{\mu}\big(e_{gc}(\mu, V) + \mu\rho\big). \tag{2.6}$$

3. THE 2-PARTICLE CASE AND THE SCATTERING LENGTH

In the case that we have only two particles $N = 2$ there exists a length a, called the *scattering length* such that

$$E(2,\Omega,V) = 8\pi a L^{-3}\big(1 + O(a/L)\big). \tag{3.1}$$

The problem can be studied by analyzing the simple Schrödinger operator $-\Delta + \frac{1}{2}V$ on $L^2(\mathbb{R}^3)$. Indeed, we introduce the *scattering solution*, i.e., the unique function $\varphi : \mathbb{R}^3 \to [0,\infty)$ satisfying the zero energy equation

$$\left(-\Delta + \frac{1}{2}V\right)\varphi = 0$$

with the limiting condition $\lim_{x\to\infty}\varphi = 1$. Then, in terms of the scattering length, the scattering solution satisfies $\varphi(x) = 1 - a/|x|$ for x outside a ball containing the support of V. Moreover,

$$\int V\varphi = 8\pi a.$$

Since we also have $0 \leq \varphi \leq 1$, we see that $8\pi a \leq \int V$. In the case of the hard core (2.2), the scattering length is indeed the radius a of the core. In this case $\int V = \infty$, whereas the scattering length is finite.

We are now in a position to state the main result on the Lee–Huang–Yang asymptotics. The asymptotic formula is proved by giving upper and lower bounds for the energy density $e(\rho, V)$. The upper bound is proved by constructing approximate trial ground state eigenfunctions that reproduce the asymptotics. Establishing a matching lower bound is usually considered more difficult as it requires ideas of how to control unimportant parts of the Hamiltonian. The upper bound has, however, proved to be very difficult too, and today the lower bound requires fewer assumptions on the potential than the upper bound.

The main results on the upper and lower bounds establishing the LHY formula are given in the next two theorems.

Theorem 3.1 (The lower bound in the LHY formula). *If $V : \mathbb{R}^3 \to [0,\infty]$ is measurable, spherically symmetric with compact support then there exist a constant $C > 0$, depending only on the support of V, and an explicit number $\eta > 0$ such that*

$$e(\rho) \geq 4\pi\rho^2 a\left(1 + \frac{128}{15\sqrt{\pi}}\sqrt{\rho a^3} - C(\rho a^3)^{\frac{1}{2}+\eta}\right). \tag{3.2}$$

This lower bound was established in [8,9].

Theorem 3.2 (The upper bound in the LHY formula). *If $0 \le V \in L^3(\mathbb{R}^3)$, spherically symmetric with compact support then there exists a constant $C > 0$ depending on the potential V such that*

$$e(\rho) \le 4\pi\rho^2 a\left(1 + \frac{128}{15\sqrt{\pi}}\sqrt{\rho a^3} + C(\rho a^3)^{\frac{1}{2}+\frac{1}{10}}\right). \tag{3.3}$$

The upper bound as stated here was proved in [1]. The first proof of an upper bound giving the first two terms was established in [19].

It is not difficult to heuristically understand the leading term as we shall now explain. In the dilute limit, it is natural to expect to find the energy to be the energy of two particles times the number of pairs. This would then, indeed, lead to an approximation for $e(\rho, V)$ given by

$$\lim_{L\to\infty, N/L^3=\rho} L^{-3}\frac{N(N-1)}{2}E(2,\Omega,V) = 4\pi\rho^2 a.$$

It is, however, already difficult to get an upper bound that reproduces this correctly. To illustrate this difficulty, notice that the simple constant trial state

$$\Psi_L = L^{-3N/2} \tag{3.4}$$

that minimizes the kinetic energy gives

$$\lim_{L\to\infty, N/L=\rho} L^{-3}\langle\Psi_L, H_N\Psi_L\rangle = \frac{1}{2}\rho^2\int V$$

which, as we saw above, can be much bigger than $4\pi\rho^2 a$. The main difficulty is to understand how to improve the large value $\int V$ with the smaller scattering length expression $8\pi a$.

We end this section by giving a short review of the history of the formula which has been a major open problem in mathematical physics for over 60 years. Additional details can be found in [14]. The leading term in the LHY expansion (1.1) was predicted by Lenz in [13]. The first to analyze it rigorously was Freeman Dyson in [6], i.e., in the same volume of Physical Review in which the paper of Lee, Huang, and Yang appeared. Dyson, indeed, proved an upper bound which gave the correct leading term for the hard sphere gas. In the case of the hard sphere gas, Dyson's upper bound still today gives the best known error term of order $(\rho a^3)^{1/3}$, which is unfortunately not of the LHY order. Dyson also gave a lower bound of the right leading order, but with a wrong constant. It took another 40 years before Lieb and Yngvason in [17] established the lower bound with the correct constant. Ten years later Erdős, Schlein, and Yau noticed in [7] that the Gaussian or quasi-free states in Bogolubov's theory of superconductivity can be used to give an upper bound that is correct to leading order and has an error term of the same order as the second term in the LHY formula but with a wrong constant. Later Yau and Yin [19] improved on the quasi-free states to get the correct LHY formula as an upper bound. Both [7] and [19] require some regularity of the potential and do not work for the hard core. In [5,10] the correct second term in the LHY formula was derived for sufficiently soft potentials, i.e., potentials that were allowed to depend in particular ways on the diluteness parameter. Finally, the correct LHY lower bound was proved first for general L^1 potentials in [8] and then in the most general case stated above in [9]. For gases confined to boxes of size $(\rho a)^{-1/2}$, the LHY formula was derived in [2, 3]. The length scale $(\rho a)^{-1/2}$ is called the *healing length* and its relevance will become clear when we discuss Bogolubov's theory in the next section.

4. BOGOLUBOV'S THEORY OF SUPERFLUIDITY

The LHY formula for the ground state energy can heuristically be understood from Bogoliubov's theory of superfluidity [4]. Thus in some sense establishing the LHY formula validates Bogolubov's theory. We will briefly describe this here, but it will require us to take a little detour into the second quantized formalism.

4.1. Second quantized formalism

For any function $f \in L^2(\Omega)$, we introduce the bosonic *annihilation operator*

$$a(f) : \mathcal{H}_N^{\mathrm{B}} \to \mathcal{H}_{N-1}^{\mathrm{B}}$$

defined by

$$\big(a(f)\Psi\big)(x_1,\ldots,x_{N-1}) = \sqrt{N}\int_\Omega \overline{f(x_N)}\Psi(x_1,\ldots,x_{N-1},x_N)\,dx_N.$$

The bosonic *creation operator* $a^\dagger(f) : \mathcal{H}_{N-1}^{\mathrm{B}} \to \mathcal{H}_N^{\mathrm{B}}$ is the adjoint $a^\dagger(f) = a(f)^*$ of $a(f)$. We here use the standard notation in physics to indicate the adjoint with a †. We deliberately did not put a subscript N on the creation or annihilation operators because we want to use the same notation independently of N. Indeed, this will allow us to write the famous commutation relations

$$\big[a(f), a(g)\big] = 0, \quad \big[a(f), a^\dagger(g)\big] = (g, f)_{L^2(\Omega)}.$$

Using the second quantization formalism, we can rewrite the Hamiltonian H_N (at least for L large enough) as

$$\begin{aligned} H &= \sum_{p\in\frac{2\pi}{L}\mathbb{Z}^3} p^2 a_p^\dagger a_p + \frac{1}{2L^3}\sum_{p,q,k\in\frac{2\pi}{L}\mathbb{Z}^3} \hat{V}(k) a_{p+k}^\dagger a_{q-k}^\dagger a_q a_p \\ &= \sum_{p\in\frac{2\pi}{L}\mathbb{Z}^3} p^2 a_p^\dagger a_p + \frac{N-1}{2}\rho\hat{V}(0) + \frac{1}{2L^3}\sum_{0\neq k\in\frac{2\pi}{L}\mathbb{Z}^3}\sum_{p,q\in\frac{2\pi}{L}\mathbb{Z}^3} \hat{V}(k) a_{p+k}^\dagger a_{q-k}^\dagger a_q a_p, \end{aligned} \tag{4.1}$$

where we used the short hand notation $a_p = a(L^{-3/2}\exp(ipx))$. These operators satisfy the commutation relations

$$\big[a_p, a_q\big] = 0, \quad \big[a_p, a_q^\dagger\big] = \delta_{p,q}. \tag{4.2}$$

We have also introduced the Fourier transform

$$\hat{V}(k) = \int_{\mathbb{R}^3} \exp(-ipx)V(x)dx.$$

4.2. The Bogolubov's approximation

In his 1947 paper [4], Bogolubov introduces an approximation to the Hamiltonian H_N, or in fact to the operator H in (4.1), that forms the basis of his theory of superfluidity. Bogolubov's approximation may be divided into three steps.

Step 1. Condensation and c-number substitution. The assumption is that the ground state or low lying energy states represent a condensate, i.e., have many particles with momentum $p = 0$. If all particles had momentum $p = 0$, we would get the state (3.4) which we know does not have the correct ground state energy. It is, however, still possible that the expectation of the operator $a_0^\dagger a_0$ in the ground state is close to the total number of particles N. The second ingredient in this first step of the approximation is to replace the operators a_0 and $a_0^\dagger$ by the number $\sqrt{N}$ in the Hamiltonian H in (4.1). This is referred to as c-number substitution. It will lead to an operator that no longer maps $\mathcal{H}_N^{\mathrm{B}}$ to itself. We consider it instead as an operator on the *bosonic Fock space* $\bigoplus_{M=0}^{\infty} \mathcal{H}_M^{\mathrm{B}}$.

Step 2. The Bogolubov's Hamiltonian. The first step results in a Hamiltonian that will have terms containing zero, two, three, or four factors $a_p^\dagger$ or a_p with $p \neq 0$. There are no terms with only one $a_p^\dagger$ or a_p with $p \neq 0$ because of momentum conservation. The second step in the approximation is to assume that we may consider $a_p^\dagger$ or a_p with $p \neq 0$ to be small and therefore ignore terms with three or more such factors. This will lead to the Bogolubov's Hamiltonian

$$H_{\mathrm{Bog}} = \sum_{0 \neq p \in \frac{2\pi}{L}\mathbb{Z}^3} \left((p^2 + \rho \hat{V}(p)) a_p^\dagger a_p + \frac{1}{2}\rho \hat{V}(p)\big(a_p^\dagger a_{-p}^\dagger + a_{-p} a_p\big) \right) + \frac{N-1}{2}\rho \hat{V}(0). \tag{4.3}$$

Step 3. Diagonalizing the Bogolubov's Hamiltonian. It is not difficult to diagonalize the Bogolubov's Hamiltonian if we apply the following simple lemma whose proof is elementary.

Lemma 4.1 (Simple case of Bogoliubov's diagonalization). *For $\mathcal{A} > 0$, $\mathcal{B} \in \mathbb{R}$ satisfying $|\mathcal{B}| \leq \mathcal{A}$, we have the operator identity*

$$\mathcal{A}(a_p^\dagger a_p + a_{-p}^\dagger a_{-p}) + \mathcal{B}(a_p^\dagger a_{-p}^\dagger + a_{-p} a_p) = \mathcal{D}(b_p^\dagger b_p + b_{-p}^\dagger b_{-p}) - (\mathcal{A} - \sqrt{\mathcal{A}^2 - \mathcal{B}^2}), \tag{4.4}$$

where

$$\mathcal{D} := \sqrt{\mathcal{A}^2 - \mathcal{B}^2}, \tag{4.5}$$

and

$$b_p := (1-\alpha^2)^{-1/2}(a_p + \alpha a_{-p}^\dagger), \quad b_{-p} := (1-\alpha^2)^{-1/2}(a_{-p} + \alpha a_p^\dagger), \tag{4.6}$$

with

$$\alpha := \mathcal{B}^{-1}(\mathcal{A} - \sqrt{\mathcal{A}^2 - \mathcal{B}^2}). \tag{4.7}$$

Note that the operators b_p and $b_p^\dagger$ satisfy the same commutation relations (4.2) as the operators a_p and $a_p^\dagger$. We see that the Bogolubov's Hamiltonian may be rewritten as

$$H_{\mathrm{Bog}} = \bigg(\sum_{p \in \frac{2\pi}{L}\mathbb{Z}^3} \varepsilon(p) b_p^\dagger b_p \bigg) + E_L \tag{4.8}$$

with

$$\varepsilon(p) = \sqrt{\big(p^2 + \rho\hat{V}(p)\big)^2 - \big(\rho\hat{V}(p)\big)^2}, \tag{4.9}$$

and where the ground state energy of H_{Bog} is

$$E_L = \frac{N-1}{2}\rho\hat{V}(0) - \frac{1}{2}\sum_{0 \neq p \in \frac{2\pi}{L}\mathbb{Z}^3} \Big(\big(p^2 + \rho\hat{V}(p)\big) - \sqrt{\big(p^2 + \rho\hat{V}(p)\big)^2 - \big(\rho\hat{V}(p)\big)^2}\Big). \tag{4.10}$$

The ground state of the Bogolubov's Hamiltonian is the vacuum state for the operators b_p. Such vacuum states of general bosonic annihilation operators are referred to as (pure) quasi-free or Gaussian states.

In the thermodynamic limit, we have $\lim_{L\to\infty, N/L^3=\rho} \frac{E_L}{L^3} = e_{\mathrm{Bog}}(\rho, V)$, where

$$\begin{aligned} e_{\mathrm{Bog}}(\rho, V) = {} & \frac{1}{2}\rho^2 \int V \\ & - \frac{1}{2}(2\pi)^{-3} \int_{\mathbb{R}^3} \Big(\big(p^2 + \rho\hat{V}(p)\big) - \sqrt{\big(p^2 + \rho\hat{V}(p)\big)^2 - \big(\rho\hat{V}(p)\big)^2}\Big)\, dp. \end{aligned} \tag{4.11}$$

We may rewrite this as

$$\begin{aligned} e_{\mathrm{Bog}}(\rho, V) = {} & 4\pi\rho^2(a_0 + a_1) \\ & - \frac{1}{16\pi^3} \int p^2 + \rho\hat{V}(p) - \sqrt{p^4 + 2\rho\hat{V}(p)p^2} - \rho^2 \frac{\hat{V}(p)^2}{2p^2}\, dp, \end{aligned} \tag{4.12}$$

where we have introduced the notation

$$a_0 = \frac{1}{8\pi}\int V, \quad a_1 = \frac{-1}{(8\pi)^2} \iint \frac{V(x)V(y)}{|x-y|}\, dx dy = \frac{-1}{64\pi^4} \int \frac{\hat{V}(p)^2}{2p^2}\, dp. \tag{4.13}$$

In fact, a_0 and a_1 are the first two terms in what is called the *Born series* for the scattering length a. In the last integral in (4.12), we can change variable $p = \sqrt{8\pi\rho a_0} q$ and arrive at

$$\begin{aligned} & \int p^2 + \rho\hat{V}(p) - \sqrt{p^4 + 2\rho\hat{V}(p)p^2} - \rho^2 \frac{\hat{V}(p)^2}{2p^2}\, dp \\ & \quad = (8\pi\rho a_0)^{5/2} \int q^2 + W_\rho(q) - \sqrt{q^4 + 2W_\rho(q)q^2} - \frac{W_\rho(q)^2}{2q^2}\, dq, \end{aligned} \tag{4.14}$$

where we wrote $W_\rho(q) = (8\pi a_0)^{-1}\hat{V}(\sqrt{8\pi\rho a_0} q)$. In the dilute limit, we may assume that $(\rho a_0)^{-1/2}$ is much longer than the range of the potential and hence we can, to leading order in the integral, replace $W_\rho(q)$ by $W_\rho(0) = 1$. Since

$$\int_{\mathbb{R}^3} q^2 + 1 - \sqrt{q^4 + 2q^2} - \frac{1}{2q^2}\, dq = -\frac{32\sqrt{2}\pi}{15},$$

we arrive at

$$e_{\mathrm{Bog}}(\rho, V) \approx 4\pi\rho^2(a_0 + a_1) + 4\pi\rho^2 a_0 \frac{128}{15\sqrt{\pi}} \sqrt{\rho a_0^3}. \tag{4.15}$$

If we replace the first two Born terms $a_0 + a_1$ by the scattering length a in the first term and a_0 by a in the second term above, we arrive at the Lee–Huang–Yang formula. We note that the change of variable $p = \sqrt{8\pi\rho a_0} q$ in the integral above shows that the relevant momenta that contribute to the LHY formula are of order of the inverse healing length $\sqrt{\rho a}$.

Understanding the validity of the Bogolubov's approximation and the validity of these last replacements above were the major challenges in establishing the LHY formula rigorously. We address this in the next section. We will end this section with a few further remarks about the Bogolubov's approximation and Bogolubov's theory of superfluidity.

In his treatment [11] of superfluidity in helium, Landau realized the importance of a linear dispersion law, i.e., that the energies of excitations grow linearly with momentum. The slope in the linear dispersion represents the critical velocity for superfluidity, i.e., the velocity below which objects can move through the fluid without creating excitations. We see that the dispersion $\varepsilon(p)$ in (4.9), indeed, has a nonvanishing linear slope $\lim_{p\to 0} |\nabla\varepsilon(p)| = \sqrt{2\rho\hat{V}(0)}$. This is the central point in Bogolubov's theory of superfluidity in weakly interacting Bose gases.

5. RIGOROUS PROOF OF THE LEE–HUANG–YANG FORMULA

In this section we very briefly sketch the rigorous arguments, leading to the lower bound in Theorem 3.1. The details can be found in [8,9].

An important ingredient in the Bogolubov's approximation was the assumption of condensation. It is still a great mathematical challenge to establish Bose condensation in nontrapped translation invariant Bose gases. To circumvent this, the first step in the rigorous derivation of the Lee–Huang–Yang formula in [8,9] is a localization to boxes that are essentially of the order of the healing length. On this scale, it turns out that the gas will look sufficiently condensed. In other words, it is not possible to show that most particles in a thermodynamic box are in a state of momentum zero. We can, however, show that most particles have momenta small compared to the inverse healing length.

For the rigorous lower bound, the localization is achieved by an operator estimate on the Hamiltonian

$$H_N - \mu N \geq \int h_u \, du, \tag{5.1}$$

where we introduced the chemical potential μ that we will write $\mu = 8\pi\rho_\mu a$. The reason for this choice is that if we insert the leading term in the LHY formula $4\pi\rho^2 a$ for the energy density then the choice of ρ that minimizes $4\pi\rho^2 a - \mu\rho = 4\pi\rho^2 a - 8\pi\rho_\mu a\rho$ is indeed $\rho = \rho_\mu$. The operators h_u above represent translations by $u \in \mathbb{R}^3$ of a Hamiltonian h_0 localized to a box $[0, \ell]^3$ with length $\ell = K_\ell(\rho_\mu a)^{-1/2}$ for a sufficiently large constant K_ℓ, i.e., we are localizing on scales that are large compared to the healing length.

To describe the localized Hamiltonian h_0, we introduce the orthogonal projection P that projects onto the one-dimensional space of constant functions in $L^2([0,\ell]^3)$ and the projection $Q = I - P$ onto the orthogonal complement. We also introduce a sufficiently regular function $\chi : \mathbb{R}^3 \to [0,\infty)$ supported on $[0,1]^3$ and let $\chi_\ell(x) = \chi(x/\ell)$.

The localized Hamiltonian then has the form

$$h_0 = \sum_{i=1}^{N} T_i - \rho_\mu \sum_{i=1}^{N} \int w_1(x_i, y)dy + \sum_{1\leq i<j\leq N} w(x_i, x_j), \tag{5.2}$$

where the localized kinetic energy is

$$T = Q\chi_\ell K(\Delta)\chi_\ell Q + QG_0Q \tag{5.3}$$

with $K(t)$ being a function that is essentially the identity for $t \gg \ell^{-2}$ and G_0 an operator that ensures a sufficient gap above zero in the kinetic energy, i.e., an important property is $G_0 \geq (\text{const})\ell^{-2}$. The exact forms of K and G_0 are complicated and can be found in [8,9]. The potential function is

$$w(x,y) = \chi_\ell(x)\frac{V(x-y)}{\chi * \chi(x/\ell)}\chi_\ell(y), \quad w_1(x,y) = w(x,y)\varphi(x-y), \tag{5.4}$$

where we recall that φ is the scattering solution. For the potential part of the Hamiltonian, it is not difficult to see that (5.1) is actually an identity. It is for the kinetic energy that it becomes a lower bound.

In order to establish condensation, it is necessary to obtain an a priori lower bound on the ground state energy of h_0 of the correct LHY order. This is achieved in [8,9] by doing a further localization that we shall not discuss here. Such an a priori lower bound establishes a bound on the expectation of the number of noncondensed particles $n_+ = \sum_{i=1}^N Q_i$. Indeed, the bound on the gap operator G_0 and the a priori bound imply that any state that does not already satisfy an LHY lower bound would have

$$(\text{const})\ell^{-2}\langle n_+\rangle \leq \langle G_0\rangle \leq \rho_\mu^2 a\sqrt{\rho_\mu a^3}\ell^3,$$

i.e., $\langle n_+\rangle \leq C\rho_\mu^2 a\ell^5 = K_\ell^2\sqrt{\rho_\mu a^3}\rho_\mu\ell^3$. In other words, the expected number of noncondensed particles is smaller by the (small) factor $K_\ell^2\sqrt{\rho_\mu a^3}$ compared to the expected number of particles $\rho_\mu\ell^3$ in the box. Unfortunately, it is not sufficient to control the *expected* number of noncondensed particles. There are terms that require controlling powers of the number of noncondensed particles. To achieve control of powers, we establish in [8] a stronger version of condensation, namely that it is enough to restrict attention to the part of the Hilbert space where we have the operator bound $n_+ \leq \mathcal{M}$ for some appropriately chosen parameter $\mathcal{M}$. Unfortunately, in order to treat the hard core potential, in [9] we had to work with a much more complicated restriction, namely that $n_+^I \leq \mathcal{M}_I$ where n_+^I represents the number of particles with kinetic energy in an interval $I = (0, K_I\ell^{-2})$, for an appropriately large constant K_I. This means $n_+^I = \sum_{i=1}^N \mathbb{1}_I(T)_i$. Note that n_+^I would be equal to n_+ if $K_I = \infty$. The point is that only restricting this operator allows us to choose a much smaller $\mathcal{M}_I$ than we would if we had to restrict n_+. The argument required to restrict the Hilbert space uses a method developed in [16] referred to as localization of large matrices.

Having established control on the number of noncondensed particles we use c-number substitution to treat the condensed particles. This can be done rigorously using the method in [15].

A very central point in our analysis is a decomposition of the localized interaction potential using in a very particular way the scattering solution φ. As $0 \le \varphi \le 1$, it is convenient to introduce the function $\omega = 1 - \varphi$ satisfying $0 \le \omega \le 1$ and tending to zero at infinity. The following decomposition is an elementary, but crucial, identity from [8]:

$$-\rho_\mu \sum_{i=1}^{N} \int w_1(x_i, y)\, dy + \sum_{1 \le i < j \le N} w(x_i, x_j) = \mathcal{Q}_0^{\mathrm{ren}} + \mathcal{Q}_1^{\mathrm{ren}} + \mathcal{Q}_2^{\mathrm{ren}} + \mathcal{Q}_3^{\mathrm{ren}} + \mathcal{Q}_4^{\mathrm{ren}},$$

where

$$\begin{aligned}
\mathcal{Q}_4^{\mathrm{ren}} &:= \frac{1}{2} \sum_{i \neq j} \bigl[Q_i Q_j + (P_i P_j + P_i Q_j + Q_i P_j)\omega(x_i - x_j)\bigr] w(x_i, x_j) \\
&\qquad \times \bigl[Q_j Q_i + \omega(x_i - x_j)(P_j P_i + P_j Q_i + Q_j P_i)\bigr], \\
\mathcal{Q}_3^{\mathrm{ren}} &:= \sum_{i \neq j} P_i Q_j w_1(x_i, x_j) Q_j Q_i + h.c., \\
\mathcal{Q}_2^{\mathrm{ren}} &:= \sum_{i \neq j} P_i Q_j w_2(x_i, x_j) P_j Q_i + \sum_{i \neq j} P_i Q_j w_2(x_i, x_j) Q_j P_i \\
&\qquad - \rho_\mu \sum_{i=1}^{N} Q_i \int w_1(x_i, y)\, dy Q_i + \frac{1}{2} \sum_{i \neq j} \bigl(P_i P_j w_1(x_i, x_j) Q_j Q_i + h.c.\bigr), \\
\mathcal{Q}_1^{\mathrm{ren}} &:= \sum_{i,j} P_j Q_i w_2(x_i, x_j) P_i P_j - \rho_\mu \sum_{i} Q_i \int w_1(x_i, y)\, dy P_i + h.c., \\
\mathcal{Q}_0^{\mathrm{ren}} &:= \frac{1}{2} \sum_{i \neq j} P_i P_j w_2(x_i, x_j) P_j P_i - \rho_\mu \sum_{i} P_i \int w_1(x_i, y)\, dy P_i.
\end{aligned}$$

Here $w_2(x, y) = w_1(x, y)(1 + \omega(x, y))$. The main observation is that the term $\mathcal{Q}_4^{\mathrm{ren}}$ is nonnegative and can be ignored for a lower bound. We think of the terms with zero to four Q's as being similar to the corresponding terms in the Bogolubov's analysis with zero to four factors of a_p or $a_p^\dagger$ with $p \neq 0$. Note that in ignoring the term $\mathcal{Q}_4^{\mathrm{ren}}$ we are not simply ignoring the term with four Q's as Bogolubov did.

The term $\mathcal{Q}_2^{\mathrm{ren}}$, together with the kinetic energy, can be rewritten in a form similar to a Bogolubov's Hamiltonian and can be diagonalized using a Bogolubov-type diagonalization argument. Note that $\mathcal{Q}_2^{\mathrm{ren}}$ contains the potentials w_1 and w_2. The potential w_1 is a localization of $V\varphi$ that satisfies $\widehat{V\varphi}(0) = 8\pi a$. This is the reason that our analysis will immediately lead to the scattering length appearing and not only the Born approximations a_0 and $a_0 + a_1$.

The appearance of w_2 in $\mathcal{Q}_2^{\mathrm{ren}}$ means that the analysis of the $\mathcal{Q}_2^{\mathrm{ren}}$ does not directly give the LHY formula. The additional contributions from the difference between w_1 and w_2 will, however, be exactly canceled by a careful analysis of the term $\mathcal{Q}_3^{\mathrm{ren}}$. This term can again be approximately diagonalized, this time together with the excitation Hamiltonian from the Bogolubov's diagonalization, i.e., the analog of the first term in (4.8). This, however, first requires estimating the operator PQw_1QQ appearing in $\mathcal{Q}_3^{\mathrm{ren}}$ in terms of an operator

$PQ_Lw_1Q_HQ_H$ where Q_L essentially projects onto appropriately low (but still nonzero) momenta and Q_H projects onto high momenta disjoint from the low momenta. Note that the operator $PQ_Lw_1Q_HQ_H$ is quadratic in Q_H which is why it allows for a Bogolubov's treatment similar to the treatment of $\mathcal{Q}_2^{\text{ren}}$.

In Bogolubov's case there was no term corresponding to $\mathcal{Q}_1^{\text{ren}}$ because of momentum conservation. Here our spatial localization breaks momentum conservation, and we therefore have a term $\mathcal{Q}_1^{\text{ren}}$. This term can fairly easily be treated together with the $\mathcal{Q}_2^{\text{ren}}$ term in the first Bogolubov's diagonalization.

Putting these ingredients together is rather technical but eventually leads to the rigorous lower bound in Theorem 3.1.

FUNDING

SF was partially supported by a Sapere Aude grant from the Independent Research Fund Denmark, Grant number DFF–4181-00221, by the Charles Simonyi Endowment, and by an Elite Research Prize from the Danish Ministry of Higher Education and Science. JPS was partially supported by the Villum Centre of Excellence for the Mathematics of Quantum Theory (QMATH).

REFERENCES

[1] G. Basti, S. Cenatiempo, and B. Schlein, A new second order upper bound for the ground state energy of dilute Bose gases. 2021, arXiv:2101.06222.

[2] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, Complete Bose–Einstein condensation in the Gross–Pitaevskii regime. *Comm. Math. Phys.* **359** (2018), 975–1026.

[3] C. Boccato, C. Brennecke, S. Cenatiempo, and B. Schlein, Bogoliubov theory in the Gross–Pitaevskii limit. *Acta Math.* **222** (2019), 219–335.

[4] N. N. Bogolyubov, On the theory of superfluidity. *Proc. Inst. Math. Kiev* **9** (1947), 89–103. Rus. Trans. *Izv. Akad. Nauk Ser. Fiz.* **11** (1947), 77, Eng. Trans. *J. Phys. (USSR)* **11** (1947), 23.

[5] B. Brietzke and J. P. Solovej, The second order correction to the ground state energy of the dilute Bose gas. *Ann. Henri Poincaré* **21** (2020), 571–626.

[6] F. J. Dyson, Ground-state energy of a hard-sphere gas. *Phys. Rev.* **106** (1957), 20–26.

[7] L. Erdős, B. Schlein, and H.-T. Yau, Ground-state energy of a low-density Bose gas: A second-order upper bound. *Phys. Rev. A* **78** (2008).

[8] S. Fournais and J. P. Solovej, The energy of dilute bose gases. *Ann. of Math.* **192** (2020), 893–976.

[9] S. Fournais and J. P. Solovej, The energy of dilute Bose gases II. 2021, arXiv:2108.12022.

[10] A. Giuliani and R. Seiringer, The ground state energy of the weakly interacting bose gas at high density. *J. Stat. Phys.* **135** (2009), 915–934.

[11] L. D. Landau, The theory of superfluidity of helium II. *J. Phys. USSR* **5** (1941), 71–79.

[12] T. D. Lee, K. Huang, and C. N. Yang, Eigenvalues and eigenfunctions of a Bose system of hard spheres and its low-temperature properties. *Phys. Rev.* **106** (1957), 1135–1145.

[13] W. Lenz, Die Wellenfunktion und Geschwindigkeitsverteilung des entarteten Gases. *Z. Phys.* **56** (1929), 778–789.

[14] E. H. Lieb, R. Seiringer, J. P. Solovej, and J. Yngvason, *The mathematics of the Bose gas and its condensation*. Birkhäuser, 2005.

[15] E. H. Lieb, R. Seiringer, and J. Yngvason, Justification of c-number substitutions in bosonic Hamiltonians. *Phys. Rev. Lett.* **94** (2005), 080401.

[16] E. H. Lieb and J. P. Solovej, Ground state energy of the one-component charged Bose gas. *Comm. Math. Phys.* **217** (2001), 127–163.

[17] E. H. Lieb and J. Yngvason, Ground state energy of the low density Bose gas. *Phys. Rev. Lett.* **80** (1998), 2504–2507.

[18] N. Navon, S. Piatecki, K. Günter, B. Rem, T.-C. Nguyen, F. Chevy, W. Krauth, and C. Salomon, Dynamics and thermodynamics of the low-temperature strongly interacting Bose gas. *Phys. Rev. Lett.* **107** (2011), 135301.

[19] H. T. Yau and J. Yin, The second order upper bound for the ground state energy of a Bose gas. *J. Stat. Phys.* **136** (2009), 453–503.

SØREN FOURNAIS

Department of Mathematics, Aarhus University, Ny Munkegade 118, DK-8000 Aarhus C, Denmark, fournais@math.au.dk

JAN PHILIP SOLOVEJ

Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, solovej@math.ku.dk

SCALING LIMITS AND UNIVERSALITY OF ISING AND DIMER MODELS

ALESSANDRO GIULIANI

ABSTRACT

After having introduced the notion of universality in statistical mechanics and its importance for our comprehension of the macroscopic behavior of interacting systems, I review recent progress in the understanding of the scaling limit of lattice critical models, including a quantitative characterization of the limiting distribution and the robustness of the limit under perturbations of the microscopic Hamiltonian. Specifically, I focus on two classes of non-exactly-solvable two-dimensional systems: nonplanar Ising models and interacting dimers. In both settings, I describe the conjectures on the expected structure of the scaling limit, review the progress towards their proof, and state some of the recent results on the universality of the limit, which I contributed to. Finally, I outline the ideas and methods involved in the proofs, describe some of the perspectives opened by these results, and propose several open problems.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 82B20; Secondary 82B27, 82B28

KEYWORDS

Non-planar Ising models, interacting dimer models, scaling limit, Renormalization Group

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4040–4072
DOI 10.4171/ICM2022/131

1. UNIVERSALITY IN STATISTICAL MECHANICS: A MATHEMATICAL CHALLENGE

Statistical Mechanics (SM) aims at explaining the macroscopic behavior of matter in its different states, starting from a microscopic description of the system, which involves an extremely large number of elementary components, such as atoms, molecules, or spins. Due to the complexity of the microscopic structure of realistic materials and to the necessity of handling models that are accessible to theoretical and numerical treatments, the mathematical modeling of any system one may wish to study inevitably requires approximations and simplifications, often quite drastic: essentially all the models studied in equilibrium and nonequilibrium statistical mechanics are “toy models,” even the most challenging ones. As illustrative examples, think to the description of magnets in terms of Ising, XY, or Heisenberg models; of disordered materials in terms of the Anderson model and the interacting extensions thereof (in the context of electrical conduction in the presence of lattice defects) or of the Edwards–Anderson model (in the context of spin glasses); of anisotropic liquids in terms of monomer–dimer systems; and so on. The oversimplifications underlying the definitions of these models cast a dark light on the physical reliability of their predictions. A priori, there is no reason why the thermodynamic and correlation functions of real magnets, liquids, or conducting materials should behave quantitatively (or even qualitatively) in the same way as those of the Ising model, the Anderson model, the dimer–monomer model, etc.

Predictions based on these popular but oversimplified models can be reliable only if they can be shown to be robust under the choice of the microscopic Hamiltonian, that is, if they depend only upon general features such as symmetry, dimensionality, etc. Vaguely speaking, robustness of the macroscopic behavior of SM systems is the content of the *universality* principle, which will be stated more precisely in two concrete mathematical settings below. For the moment, let us just observe that, in view of the previous considerations, this principle can be seen as *the* justification for the use of toy models in the description of complex materials and, in a sense, it is what makes SM predictive and useful as a whole.

Away from the critical point, where, typically, the correlations among fluctuations of local observables decay exponentially to zero at large distances, the universality of the behavior of the system at the macroscopic and mesoscopic level is closely related to the Law of Large Numbers and to the Central Limit Theorem (CLT) for weakly correlated random variables: averages of local observables converge almost surely to their expectation (the macroscopic value of the corresponding thermodynamic function), and their fluctuations around the mean converge, after appropriate rescaling, to normal random variables.

Things are much more subtle and interesting in the vicinity of a phase transition, where correlations among faraway fluctuations of local observables become so important that the CLT has no a priori reason to hold, and will in general not hold. The understanding of phase transitions is one of the central goals of SM since at least a hundred years. The existence of several different kinds of phase transition and the characterization of the corresponding low- and high-temperature phases are among the great successes of the SM of the 20th century. On the other hand, a complete understanding of the behavior of the system at, or close to, the critical point, is still missing, and poses several exciting challenges for mathematical physics and probability. Let us focus here on the case of *continuous* phase transitions. In such a case, do fluctuations of local observables admit an interesting, non-Gaussian, mesoscopic limit? Is the limit *robust* under a large class of perturbations of the interaction among the microscopic constituents of the system?

These are some of the most fundamental problems of equilibrium SM since the 1960s. The theory of Wilsonian Renormalization Group (RG) [94–96], which is among the greatest success of theoretical physics in the last century, was developed for quantitatively answering to these questions. It predicts that the "scaling limit" describing the large scale behavior of correlations at a continuous phase transition is in great generality a Euclidean Field Theory, which can be determined as the fixed point of an explicit semigroup (the Wilsonian RG transformation), acting on an often vaguely-defined "space of Hamiltonians." The Gaussian, often dubbed "trivial," fixed points of the Wilsonian RG transformations correspond to off-critical systems or to the simplest critical ones. What about nontrivial, i.e., non-Gaussian, fixed points? By construction, any such fixed point turns out to be scale invariant. It has been argued that, under some reasonable additional hypotheses on the structure of the correlation functions and the locality of theory, such fixed points are conformally invariant [84,85,98], i.e., described by a Euclidean Conformal Field Theory (CFT). If we trust this picture, we can take an axiomatic point of view, i.e., we can try to *classify* the admissible nontrivial fixed point by classifying all the possible CFTs and, whenever possible, characterize their structure (by, e.g., computing their correlation functions); this task has been essentially completed in two dimensions (2D), thanks to the rich structure of the 2D conformal group (see, e.g., [13,66] for the case of the "discrete series" of models with central charge $0 < c < 1$, and [71] for the case of Liouville theory). In three dimensions (3D), this axiomatic point of view recently led to some spectacular developments, which allowed to compute the critical exponents associated with the non-Gaussian behavior of local observables for several nontrivial fixed point theories, including one that is believed to describe the scaling limit of the 3D Ising model at its critical point [83].

Once that the candidate scaling limits have been constructed in this way, one is left with identifying the right one for any given class of microscopic Hamiltonians. While heuristically one can appeal, e.g., to symmetry considerations or to numerical constraints on the decay exponents of correlation functions to guess the right scaling limit for a given microscopic Hamiltonian, the tasks of mathematically proving that the scaling limit of the critical theory exists and it is conformal invariant, that such limit coincides with one of the candidate Euclidean CFTs, and that it is robust under a large class of perturbations of the microscopic interaction, are among the great challenges of modern mathematical SM.

Even in very specific, simple, settings, many of the natural questions arising from the above premises remain open to date. However, in the last decades there has been remarkable progress from different viewpoints, which allowed to exhibit the first examples of conformally invariant, universal, scaling limits, rigorously constructed starting from lattice microscopic models. These mathematical results are mostly restricted to 2D, which is the case I will focus on from now on. Two complementary approaches that have been, and are being, successfully used to rigorously understand universality and conformal invariance of 2D lattice SM systems are: a probabilistic one, based on random geometry, percolation and discrete holomorphicity; and a field theoretic one, based on constructive RG ideas.

The first, probabilistic, method led to the complete proof of conformal invariance of the scaling limits of the 2D planar Ising [31–33,39,55,56,91] and dimer [3,67,68,70] models. It has the advantage of being flexible in treating geometric deformations of the domain and of the underlying lattice, thus leading to the first proofs of universality with respect to these kinds of deformations. The limitation of this approach is that it is mostly restricted to exactly solved models at the "free Fermi point" (i.e., exactly solvable models, whose solution can be expressed in determinant form, as for Ising and dimers) and it is not flexible in dealing with perturbations of the microscopic Hamiltonian (there are a few important exceptions, notably [5,90], which suggest possible directions for future extensions and developments). The second, field theoretic, method led to the construction of the bulk scaling limit of several interacting, non-solvable, models, such as Ashkin–Teller and 8-vertex (8V) models [14,45,74], interacting dimers and 6-vertex (6V) models [48–50], the sine-Gordon model on the Kosterlitz–Thouless critical line [40], and many others [1,9,10,17,18,24,34]: remarkably, many of these models have non-determinantal scaling limits, and the results are robust under a large class of microscopic perturbations of the lattice Hamiltonian. Moreover, this approach led to the proof of several CFT predictions, such as scaling relations among critical exponents and amplitudes [14,20,50], bosonization identities [10,15], and expression for the universal subleading contributions to the critical free energy [46]. A limitation of this approach is that it is restricted to "weakly interacting" cases, that is, to models that are close to a Gaussian model or to a free Fermi model. Moreover, it is not yet flexible enough for dealing with non-translationally-invariant situations, including geometric perturbations of the domain or of the underlying lattice. However, recent progress in simple domains with boundaries [6,7] opens new perspectives for applications to general geometries and for an effective combinations of constructive RG ideas with probabilistic ones.

In the following, I will review some of these advances in the specific contexts of non-planar 2D Ising models and of non-integrable perturbations of 2D dimer models, focusing on a selection of results obtained via the constructive RG, whose development and application to the theory of universality in 2D SM systems I contributed to. In Section 2, I discuss a class of non-planar Ising models: I will first define the setting, then state the conjectures on the universality of the scaling limit at the critical point, and then, after having reviewed the known results in the integrable, planar, model, I will state our main results on the existence and universality of the scaling limit for the multipoint energy correlations in the plane and in the cylinder, see Theorems 2.1 and 2.2 below. In Section 3, I discuss a class of non-integrable dimer models; also in this case, after having defined the setting, stated the expected structure of the scaling limit and reviewed some of the known results in the integrable case, I will state our main results on the fine asymptotics of the dimer–dimer correlations and on the universality of the scaling limit of the height fluctuations, see Theorems 3.1 and 3.2. In Section 4, I will informally describe the methods of proof, and, in Section 5, I will comment on perspectives and open problems.

2. THE SCALING LIMIT OF NONPLANAR ISING MODELS

Consider a finite, simply connected, region of the plane, $\Omega \subset \mathbb{R}^2$, and let $\Omega_a = \Omega \cap a\mathbb{Z}^2$ be its discretization on the square grid of lattice spacing $a > 0$. At each site x of Ω_a, we assign an Ising spin $\sigma_x \in \{+,-\}$ and, given the spin configuration $\sigma \in \{+,-\}^{\Omega_a} \equiv \Sigma_a$, we assume that its energy, or Hamiltonian, has the following form:

$$H^{\lambda;\emptyset}_{a,\Omega}(\sigma) = -J \sum_{\langle x,y\rangle} \sigma_x\sigma_y - \lambda \sum_{X\subset\Omega_a} V(X)\sigma_X, \tag{2.1}$$

where $J > 0$; the first sum runs over (unordered) nearest neighbor pairs of sites in Ω_a; in the second sum, given a subset X of Ω_a, we denoted $\sigma_X := \prod_{x\in X}\sigma_x$, and V is a translationally invariant interaction, supported on *even* sets X, of finite range proportional to a[1]. In general, we will require V be neither a pair interaction (i.e., to be supported on sets of cardinality 2) nor ferromagnetic (i.e., non-negative). We will refer to the first term on the right-hand side of (2.1) as to the nearest-neighbor interaction, of strength J (to be fixed once and for all), and to the second term as to the multi-spin interaction, of strength λ (to be thought of as being small compared to J). The model describes the magnetic properties of thin ferromagnetic films with an out-of-plane easy-axis of magnetization and short range "exchange" interactions among the magnetic moments of the ions. For $\lambda = 0$, the Hamiltonian reduces to that of the planar, nearest-neighbor, Ising model, originally introduced by Lenz in 1920 **[73]**, which in 2D is exactly solvable in a very strong sense, as originally proved by Onsager **[81]**, see Section 2.1 below. For $\lambda \neq 0$, the multispin interaction breaks planarity (i.e., the interaction cannot be represented in terms of couplings associated with the edges of a planar graph with edge set Ω_a), as well as the integrability of the model.

The apex $\emptyset$ on $H^{\lambda;\emptyset}_{a,\Omega}$ refers to the boundary conditions (b.c.), which we implicitly assumed to be "open," or "free," i.e., we assumed that there is no spin in the complement of Ω_a interacting with those in Ω_a. In a similar way, we can define the Hamiltonians $H^{\lambda;+}_{a,\Omega}$ (resp. $H^{\lambda;-}_{a,\Omega}$) with $+$ (resp. $-$) b.c., by including in its definition the interactions between the spins σ in Ω_a and a configuration of spins identically equal to $+1$ (resp. -1) in its complement, Ω^c_a. Analogously, if Ω is a 2D torus (resp. a 2D cylinder), we denote by Ω_a its discretization of lattice spacing $a > 0$ and let $H^{\lambda;\mathrm{per}}_{a,\Omega}$ (resp. $H^{\lambda;\mathrm{cyl}}_{a,\Omega}$) be the spin Hamiltonian defined as in (2.1), with the first sum including the nearest neighbor pairs winding up over the torus (resp. cylinder) and the interaction V being translationally invariant with respect to the natural translations on the torus (resp. cylinder).

The finite volume Gibbs measure with inverse temperature $\beta > 0$ and # b.c., with $\# \in \{\emptyset, +, -, \mathrm{per}, \mathrm{cyl}\}$, is characterized by the probability weight

$$\mathbb{P}^{\lambda;\#}_{\beta;a,\Omega}(\sigma) = \frac{1}{Z^{\lambda;\#}_{a,\Omega}} e^{-\beta H^{\lambda;\#}_{a,\Omega}(\sigma)}, \quad \forall \sigma \in \Sigma_a, \tag{2.2}$$

where $Z^{\lambda;\#}_{a,\Omega} = \sum_{\sigma\in\Sigma_a} e^{-\beta H^{\lambda;\#}_{a,\Omega}(\sigma)}$ is the partition function. Given an observable $A : \Sigma_a \to \mathbb{R}$, we denote its average with respect to the probability weight (2.2) by $\mathbb{E}^{\lambda;\#}_{\beta;a,\Omega}(A)$.

[1] More precisely, we assume $V(X) = V_0(X/a)$ for a fixed, finite range, potential $V_0 : \mathbb{Z}^2 \to \mathbb{R}$.

"Truncated," or "connected," expectations are denoted by semicolons, e.g., $\mathbb{E}^{\lambda;\#}_{a,\Omega}(A_1;A_2) = \mathbb{E}^{\lambda;\#}_{a,\Omega}(A_1 A_2) - \mathbb{E}^{\lambda;\#}_{a,\Omega}(A_1)\mathbb{E}^{\lambda;\#}_{a,\Omega}(A_2)$.

It is well known that the system displays a phase transition, in the following sense. Fix $a > 0$ and take λ sufficiently small compared to J. Denote by $\Omega \nearrow \mathbb{R}^2$ the "thermodynamic limit" obtained by (say) centering Ω at the origin, rescaling its linear dimensions by L, and letting $L \to \infty$. Then:

- If β is small enough, for any finite $X \subset a\mathbb{Z}^2$, the limit $\mathbb{E}^{\lambda}_{\beta;a,\mathbb{R}^2}(\sigma_X) := \lim_{\Omega\nearrow\mathbb{R}^2} \mathbb{E}^{\lambda;\#}_{\beta;a,\Omega}(\sigma_X)$ is independent of the b.c. #, it is translationally invariant and characterized by the fact that $\mathbb{E}^{\lambda}_{\beta;a,\mathbb{R}^2}(\sigma_x) = 0$, $\forall x \in a\mathbb{Z}^2$, and that the truncated correlations $\mathbb{E}^{\lambda}_{\beta;a,\mathbb{R}^2}(\sigma_X;\sigma_Y)$ decay exponentially to zero as the distance between the finite sets $X, Y \subset a\mathbb{Z}^2$ diverges.
- If β is large enough, for any finite $X \subset a\mathbb{Z}^2$, the limits $\mathbb{E}^{\lambda;\pm}_{\beta;a,\mathbb{R}^2}(\sigma_X) := \lim_{\Omega\nearrow\mathbb{R}^2} \mathbb{E}^{\lambda;\pm}_{\beta;a,\Omega}(\sigma_X)$ with b.c. $+$ or $-$ exist, they are translationally invariant, but are different if $|X|$ is odd: in particular, $\mathbb{E}^{\lambda;+}_{\beta;a,\mathbb{R}^2}(\sigma_x) = -\mathbb{E}^{\lambda;-}_{\beta;a,\mathbb{R}^2}(\sigma_x)$ is positive and independent of x, and $\mathbb{E}^{\lambda;\pm}_{\beta;a,\mathbb{R}^2}(\sigma_X;\sigma_Y)$ decay exponentially to zero as the distance between the finite sets $X, Y \subset a\mathbb{Z}^2$ diverges.

The two scenarios described above are usually referred to as "high-temperature" and "low-temperature" phases, respectively. If λV is a ferromagnetic pair interaction, it is known [4] that they extend to two contiguous intervals $(0, \beta_c)$ and (β_c, ∞) separated by a critical inverse temperature $\beta_c = \beta_c(\lambda)$, at which, for any finite $X \subset a\mathbb{Z}^2$, the limit $\mathbb{E}^{\lambda}_{\beta_c;a,\mathbb{R}^2}(\sigma_X) := \lim_{\Omega\nearrow\mathbb{R}^2} \mathbb{E}^{\lambda;\#}_{\beta_c;a,\Omega}(\sigma_X)$ is independent of #, and $\mathbb{E}^{\lambda}_{\beta_c;a,\mathbb{R}^2}(\sigma_x) = 0$, $\forall x \in a\mathbb{Z}^2$. The same is expected to hold for a general (translationally invariant, even, of finite range $\propto a$) interaction V, provided that λ is small enough. Moreover, it is expected that $\mathbb{E}^{\lambda}_{\beta_c;a,\mathbb{R}^2}(\sigma_X;\sigma_Y)$ decays *algebraically* to zero as the distance among the finite sets $X, Y \subset a\mathbb{Z}^2$ diverges. Even more, the scaling limit of the correlations is expected to exist and to be universal, in the following sense. Fix $\beta = \beta_c$, and let Ω be a prescribed subset of the plane, or a 2D torus, or a 2D cylinder. Define the rescaled spin and energy variables as

$$\sigma(x) = a^{-1/8}\sigma_{[x]}, \quad \varepsilon_j(x) = a^{-1}\big(\sigma_{[x]}\sigma_{[x]+a\hat{e}_j} - \mathbb{E}^{\lambda;\#}_{\beta_c;a,\Omega}(\sigma_{[x]}\sigma_{[x]+a\hat{e}_j})\big), \tag{2.3}$$

where, for $x \in \Omega$, $[x] = a\lfloor a^{-1}x\rfloor$, $\hat{e}_j$ is the unit coordinate vector in direction $j \in \{1,2\}$, and $\# \in \{\emptyset, +, -, \mathrm{per}, \mathrm{cyl}\}$. Then it is expected that, for any tuple of distinct points $x_1, \dots, x_n, y_1, \dots, y_m$ of Ω and any choice of $j_1, \dots, j_m \in \{1,2\}$, the limit

$$\lim_{a\to 0} \mathbb{E}^{\lambda;\#}_{\beta_c;a,\Omega}\big(\sigma(x_1)\cdots\sigma(x_n)\varepsilon_{j_1}(y_1)\cdots\varepsilon_{j_m}(y_m)\big) \tag{2.4}$$

exists, it is conformally covariant under Riemann mappings of the domain Ω into an arbitrary new domain Ω' and, moreover, it depends on λ in an extremely simple, multiplicative, way, i.e., one expects that there exist two constants $Z_1 = Z_1(\lambda)$ and $Z_2 = Z_2(\lambda)$ such that the limit in (2.4) equals $Z_1^n Z_2^m$ times the limit obtained in the nearest neighbor case $\lambda = 0$. This is what universality predicts in this context and whose proof represents a key challenge in mathematical SM for the incoming years. As anticipated in the introduction, lately there has been remarkable progress towards its proof, as reviewed in the next subsections.

2.1. The nearest neighbor case

As mentioned above, for $\lambda = 0$ the model is exactly solvable in a remarkably strong sense [58, 60, 63, 65, 78–81, 88, 97]; in particular, the partition function can be written as the Pfaffian of a suitable complex adjacency matrix K of a graph, known as the Fisher graph, obtained by suitably decorating the one associated with Ω_a [41, 64]; moreover, correlation functions of local observables can be expressed in terms of Pfaffians of submatrices of K^{-1}. The exact solution provides, among other things, closed formulas for the free energy, specific heat, magnetization, and the large distance asymptotics of the spin correlations. The critical temperature is known to be $\beta_c = \beta_c(0) = (2J)^{-1}\log(\sqrt{2}+1)$, at which, letting first $\Omega \nearrow \mathbb{R}^2$ and then $a \to 0$, one finds, for any pair of distinct points $x_1, x_2 \in \mathbb{R}^2$ and any choice of $j_1, j_2 \in \{1,2\}$,

$$\begin{aligned}\lim_{a\to 0}\lim_{\Omega\nearrow\mathbb{R}^2}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}\big(\sigma(x_1)\sigma(x_2)\big) &= \frac{A}{|x-y|^{1/4}},\\ \lim_{a\to 0}\lim_{\Omega\nearrow\mathbb{R}^2}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}\big(\varepsilon_{j_1}(x_1);\varepsilon_{j_2}(x_2)\big) &= \frac{1}{\pi^2}\frac{1}{|x-y|^2},\end{aligned} \tag{2.5}$$

irrespective of the boundary conditions (in the first line, $A = 0.70338016\ldots$). The exact solution, in the form reviewed, e.g., in [79], also allows one to compute the multipoint energy correlations in the infinite plane limit: for any n-tuple of distinct points $x_1, \ldots, x_n$ and any $j_1, \ldots, j_n \in \{1,2\}$, we have

$$\lim_{a\to 0}\lim_{\Omega\nearrow\mathbb{R}^2}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}\big(\varepsilon_{j_1}(x_1)\cdots\varepsilon_{j_n}(x_n)\big) = \pi^{-n}\big|\mathrm{Pf}\,M(z_1,\ldots,z_n)\big|^2, \tag{2.6}$$

irrespective of the boundary conditions, where $z_j = (x_j)_1 + i(x_j)_2$ is the complex representative of the point x_j, and $M(z_1,\ldots,z_n)$ is the $n\times n$ antisymmetric matrix of elements $M_{ij}(z_1,\ldots,z_n) = \frac{\mathbb{1}_{i\neq j}}{z_i - z_j}$. While these results are classical, conformal covariance of the limits in finite domain remained elusive for decades. In the case of the multipoint energy correlations, a rigorous proof is due to Hongler [54], who proved that, for any open, simply connected region Ω of the plane, letting $\varphi : \Omega \to \mathbb{H}$ be the conformal mapping from Ω to the upper half-plane (both thought of as subsets of $\mathbb{C}$), and defining $\mathbb{E}^{0;\#}_{\beta_c;\Omega}(\varepsilon(z_1)\cdots\varepsilon(z_n)) := \lim_{a\to 0}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}(\varepsilon_{j_1}(x_1)\cdots\varepsilon_{j_n}(x_n))$ for any n-tuple $x_1,\ldots,x_n \in \Omega$ and $j_1,\ldots,j_n \in \{1,2\}$ (here, as above, z_j is the complex representative of x_j), then, for $\# \in \{\emptyset, +, -\}$,

$$\mathbb{E}^{0;\#}_{\beta_c;\Omega}\big(\varepsilon(z_1)\cdots\varepsilon(z_n)\big) = \left(\prod_{i=1}^n\big|\varphi'(z_i)\big|\right)\mathbb{E}^{0;\#}_{\beta_c;\mathbb{H}}\big(\varepsilon\big(\varphi(z_1)\big)\cdots\varepsilon\big(\varphi(z_n)\big)\big). \tag{2.7}$$

Moreover, the right side is explicit: in fact,

$$\mathbb{E}^{0;\emptyset}_{\beta_c;\mathbb{H}}\big(\varepsilon(z_1)\cdots\varepsilon(z_n)\big) = (-1)^n\mathbb{E}^{0;+}_{\beta_c;\mathbb{H}}\big(\varepsilon(z_1)\cdots\varepsilon(z_n)\big),$$

and

$$\mathbb{E}^{0;\pm}_{\beta_c;\mathbb{H}}\big(\varepsilon(z_1)\cdots\varepsilon(z_n)\big) = (i\pi)^{-n}\,\mathrm{Pf}\,M(z_1,\ldots,z_n,\overline{z_n},\ldots,\overline{z_1}).$$

The scaling limit of the multipoint spin correlations is more subtle. Kadanoff [61] first guessed its expression in the special case of n colinear points. In the general case, the result was conjectured on the basis of CFT methods: in fact, after the work of Belavin, Polyakov, and Zamolodchikov [13], it became clear that the scaling limit of any mixed multipoint spin-energy correlation should coincide with the corresponding correlations of the CFT minimal model with central charge $c = 1/2$, which can be explicitly computed via Coulomb Gas methods. A rigorous proof of the validity of the expected formula is very recent, compared with the history of the Ising model, and is due to Dubedat [36, 37] and to Chelkak, Hongler, and Izyurov [31, 32], who proved that for even n, letting again z_j be the complex representative of x_j,

$$\lim_{a\to 0}\lim_{\Omega\nearrow\mathbb{R}^2}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}\big(\sigma(x_1)\cdots\sigma(x_n)\big) = \Bigg[\Big(\frac{A}{\sqrt{2}}\Big)^n \sum_{\substack{\mu_1,\dots,\mu_n=\pm:\\ \mu_1+\dots+\mu_n=0}}\ \prod_{1\le i<j\le n}|z_i - z_j|^{\mu_i\mu_j/2}\Bigg]^{1/2}. \tag{2.8}$$

Even more, [31] proved that in a finite domain Ω with $\# \in \{\emptyset, +, -\}$ boundary conditions, the limiting spin correlations, $\mathbb{E}^{0;\#}_{\beta_c;\Omega}(\sigma(z_1)\cdots\sigma(z_n)) := \lim_{a\to 0}\mathbb{E}^{0;\#}_{\beta_c;a,\Omega}(\sigma(x_1)\cdots\sigma(x_n))$ are conformally covariant in a sense analogous to (2.7), i.e.,

$$\mathbb{E}^{0;\#}_{\beta_c;\Omega}\big(\sigma(z_1)\cdots\sigma(z_n)\big) = \Bigg(\prod_{i=1}^{n}\big|\varphi'(z_i)\big|^{1/8}\Bigg)\mathbb{E}^{0;\#}_{\beta_c;\mathbb{H}}\big(\sigma\big(\varphi(z_1)\big)\cdots\sigma\big(\varphi(z_n)\big)\big), \tag{2.9}$$

and, again, the right-hand side is explicit, see [31, EQ. (1.2)].

2.2. The nonplanar case

The remarkable results reviewed in the previous subsection are crucially based on the underlying exact solvability and discrete holomorphicity of the model. For $\lambda \neq 0$, neither of these properties holds, and completely different methods must be employed for constructing the scaling limit. As explained in the introduction, the natural framework for treating the effect of interactions are multiscale methods, rigorously implementing Wilson's RG ideas in the present context. A constructive approach based on these ideas was proposed in [82, 92], and successfully used in [14, 45, 74] to compute the large distance asymptotics of correlation functions and prove several instances of universality in spin and vertex models such as Ashkin–Teller, the 8V model and non-integrable variants thereof. See [76] for a review of these developments until 2010. In the context of non-planar Ising models, a decade ago we successfully employed these methods to compute and prove universality of the bulk energy correlations, as summarized in the following theorem.

Theorem 2.1 ([44]). *Fix a potential V that, besides being even and translationally invariant, has finite range proportional to a and is invariant under discrete rotations and reflections. Then there exist $\lambda_0 > 0$ and two functions $\beta_c = \beta_c(\lambda)$ and $Z_2(\lambda)$, real-analytic in λ for $|\lambda| \le \lambda_0$, such that, if $\Omega = \Omega_L$ is a 2D torus of side L, for any $n > 1$, any n-tuple of distinct* *points $x_1, \dots, x_n \in \mathbb{R}^2$, and any choice of $j_1, \dots, j_n \in \{1, 2\}$,*

$$\lim_{a\to 0} \lim_{L\to\infty} \mathbb{E}^{\lambda;\mathrm{per}}_{\beta_c;a,\Omega}\big(\varepsilon_{j_1}(x_1)\cdots\varepsilon_{j_n}(x_n)\big) = Z_2^n \pi^{-n} \big|\mathrm{Pf}\, M(z_1, \dots, z_n)\big|^2, \tag{2.10}$$

where $M(z_1, \dots, z_n)$ is the same defined after (2.6).

Note that the right-hand side of (2.10) is equal to Z_2^n times the bulk scaling limit of the nearest-neighbor model (2.6): therefore, this theorem proves the universality conjecture for the full-plane multipoint energy correlations of a large class of perturbations of the standard 2D Ising model. The proof of the theorem, which is based on multiscale cluster expansion methods (see Section 4 below) gives much more information than that summarized in its statement: for instance, it provides an explicit, and essentially optimal, speed of convergence to the limit, as well as a constructive algorithm for computing $\beta_c(\lambda)$ and $Z_2(\lambda)$; moreover, it can be adapted to more general cases, e.g., it requires neither that V is invariant under discrete rotations and reflections (the assumption has just the effect of simplifying the explicit expression on the right-hand side of (2.10)) nor that β is fixed exactly at β_c: choosing $\beta = \beta_c + am_0$, we prove in **[44]** that the scaling limit of the truncated energy correlations is non-trivial (it realizes the so-called "massive scaling limit" in the temperature direction) and it decays exponentially to zero at large distances, with rate proportional to m_0.

Theorem 2.1 and its proof are restricted to a translational-invariant setting, which guarantees, in particular, that the effective potentials used to describe the system at length scales much larger than the lattice spacing, see Section 4 for details, can be parametrized by a finite number of "relevant" and "marginal" scale-dependent couplings (using a terminology borrowed from the Wilsonian RG jargon), which are in fact *constants*, rather than functions of the position x in the domain where the system is defined on. If we are interested in constructing the scaling limit of the correlations in a finite domain Ω, then we need to keep track of such x-dependence, and to control the boundedness of the relevant and marginal couplings, as the length scale increases, uniformly in x. From a technical point of view, the x-dependence of the scale-dependent couplings potentially induces additional logarithmic divergences in the theory, arising from the integration of the degrees of freedom supported in the vicinity of the boundary. To date, there are no systematic, well-developed methods for dealing with these divergences and related technical issues: the multiscale cluster expansion, which the proof of Theorem 2.1 is based on, is not well developed yet in the case of critical theories in finite domains, where boundaries are present and affect the form of correlation functions in the scaling limit. This is a severe limitation for the rigorous construction of scaling limits in finite domains and for the study of their conformal covariance with respect to deformations of the domain. Recently, we managed to overcome several of these technical issues and to provide the first construction of non-planar Ising models in a domain with boundary, in cylindrical geometry:

Theorem 2.2 (**[6,7]**). *Fix V as in Theorem 2.1 and let $\lambda_0, \beta_c = \beta_c(\lambda)$ and $Z_2 = Z_2(\lambda)$ be the same introduced there. Let Ω be a 2D cylinder with arbitrary sides $\ell_1, \ell_2 > 0$, periodic in the horizontal direction. Then, for any $n > 1$, any n-tuple of distinct points $x_1, \dots, x_n \in \mathbb{R}^2$,* *and any choice of* $j_1, \dots, j_n \in \{1,2\}$,

$$\lim_{a\to 0} \mathbb{E}^{\lambda;\mathrm{cyl}}_{\beta_c;a,\Omega}\big(\varepsilon_{j_1}(x_1)\cdots\varepsilon_{j_n}(x_n)\big) = Z_2^n \lim_{a\to 0} \mathbb{E}^{0;\mathrm{cyl}}_{\beta_c;a,\Omega}\big(\varepsilon_{j_1}(x_1)\cdots\varepsilon_{j_n}(x_n)\big) \\ = Z_2^n(-\pi)^{-n}\,\mathrm{Pf}\,A(x_1,\dots,x_n). \tag{2.11}$$

Here $A(x_1,\dots,x_n)$ *is the* $2n\times 2n$ *antisymmetric matrix with elements* $A_{(i,a),(j,b)}(x_1,\dots,x_n) = \mathbb{1}_{i\neq j}\, g^{\mathrm{cyl}}_{a,b}(x_i,x_j)$, *where* $i,j\in\{1,\dots,n\}$, $a,b\in\{1,2\}$,

$$g^{\mathrm{cyl}}_{a,b}(x,y) = \sum_{n\in\mathbb{Z}^2}(-1)^n\big[g_{a,b}(x-y+\ell_n)+(-1)^a g_{a,b}(x-\tilde{y}+\ell_n)\big], \tag{2.12}$$

$g_{a,b}$ *are the matrix elements of*

$$g(x) = |x|^{-2}\begin{pmatrix} x_1 & x_2 \\ x_2 & -x_1\end{pmatrix},$$

$\ell_n = (n_1\ell_1, n_2\ell_2)$, *and* $\tilde{y} = (y_1,-y_2)$.

Also in this case, as for Theorem 2.1, the proof of the theorem provides bounds on the speed of convergence to the limit, and it does not rely on the assumptions that V is invariant under discrete rotations and reflections, and that the inverse temperature is fixed exactly at β_c (these were made just to simplify the statement). The key new ingredients in the proof, compared with that of Theorem 2.1, are the following (I use again the Wilsonian RG jargon, for additional details see Section 4 below): (1) proof that the scaling dimension of boundary operators is better by one dimension than their bulk counterparts, (2) a cancelation mechanism based on an approximate image rule for the fermionic two-point function allows us to control the RG flow of the marginal boundary terms. I expect that these novel ingredients will play an important role in future developments in the mathematical construction of the scaling limit of critical 2D SM models in domains with boundaries.

Let me emphasize that the result summarized in Theorem 2.2 is uniform in ℓ_1, ℓ_2.[2] Letting $\ell_1, \ell_2 \to \infty$, we obtain the correlations in the half-plane (or, if desired, those in the full-plane, depending on the way in which we perform the limit). The proof can be generalized to the computation of the scaling limit of the boundary spin correlations, which can be shown to be the Pfaffian of an explicit antisymmetric matrix. A limitation of our result, intrinsic in the multiscale cluster expansion method employed, is the restriction to small values of λ. A related result [5] is the recent proof that, if V is a ferromagnetic pair interaction, then the scaling limit of the boundary spin correlations has a Pfaffian structure. The proof is based on a random current representation and a multiscale application of Russo–Seymour–Welsh-type bounds [86,89] on the crossing probabilities of the currents, and applies to ferromagnetic pair interaction of *any strength*, i.e., remarkably, the result is nonperturbative. A limitation is that the scaling limit of the boundary spin correlations constructed in [5] and the associated

[2] Strictly speaking, the proof in [6] requires ℓ_1/ℓ_2 to be bounded from above and below. This limitation can be easily overcome: if $\ell_1 \ll \ell_2$ or $\ell_1 \gg \ell_2$, one needs to separately study the contributions from the intermediate length scales between ℓ_1 and ℓ_2, which is easy to do by the multiscale methods of [6]: in fact, at these scales, the systems effectively behaves as a 1D Ising system (whose thermodynamic behavior is "trivial") with dressed parameters.

critical exponents are not explicit: in this sense, [5] offers a complementary perspective to ours, both with respect to the results and of the techniques employed.

3. THE SCALING LIMIT OF INTERACTING DIMER MODELS

Let us now consider a different setting, the one of 2D dimer models, where the notion of universality and the nature of the scaling limit is more subtle than for Ising models. Dimer models (possibly in the presence of vacancies, called “monomers”) were introduced in the equilibrium setting by Fowler and Rushbrooke in 1935 [42] as simplified models for liquids of anisotropic molecules. Here we consider such models in the limit of no vacancies. Let us define the setting precisely: similarly to the previous section, let Ω be a simply connected region of the plane, or a 2D torus, or a 2D cylinder. Let $\mathbb{L}$ be the infinite square grid of lattice spacing $1/\sqrt{2}$ and axes tilted by 45° with respect to the standard horizontal and vertical axes, and let Ω_a be a discretization of Ω on $a\mathbb{L}$, $a>0$. Note that the graph G_{Ω_a} associated with Ω_a, i.e., the one with vertex set Ω_a and edge set consisting of the links connecting nearest neighbor sites of Ω_a, is bipartite, and we color its vertices black and white so that neighboring vertices have different colors, with the convention that the origin is black. An edge e of G_{Ω_a} is said to be of type $r\in\{1,2,3,4\}$ if its white endpoint is to the NE, NW, SW, SE of its black endpoint, respectively. For any e, we let $r(e)$ be its type and $x(e)$ the coordinate of its black site (note that $x(e)\in a\mathbb{Z}^2$). A dimer covering, or “allowed dimer configuration,” of the graph G_{Ω_a} is a subset of its edges that covers every vertex exactly once. We denote by $\mathfrak{D}=\mathfrak{D}(\Omega_a)$ the set of allowed dimer configurations in Ω_a, and we assume that Ω_a has been chosen in such a way that $\mathfrak{D}\neq\emptyset$. The class of dimer models we are interested in are defined by the following probability measure on $\mathfrak{D}$:

$$\mathbb{P}^{\lambda}_{a,\Omega}(D)=\frac{1}{Z^{\lambda}_{a,\Omega}}\Big(\prod_{e\in D}t_{r(e)}\Big)e^{\lambda V(D)},\quad \forall D\in\mathfrak{D},\tag{3.1}$$

where t_r, with $r\in\{1,2,3,4\}$, are the weights of the 4 different dimer types; V is a translationally invariant interaction, of finite range proportional to a; λ is the interaction strength, to be thought of as “small”; $Z^{\lambda}_{a,\Omega}=\sum_{D\in\mathfrak{D}}(\prod_{e\in D}t_{r(e)})e^{\lambda V(D)}$ is the partition function. With no loss of generality, we can fix $t_4=1$, and we shall do so in the following. For $\lambda=0$, the model is exactly solvable in a very strong sense, as originally proved by Kasteleyn [64] and by Temperley and Fisher [93], see Section 3.1 below. In general, for $\lambda\neq 0$, the model is not exactly solvable anymore, even though there is a special choice of the interaction V, of nearest neighbor type, for which it reduces to the 6V model, which is solvable by Bethe ansatz, see [50, SECTION 2.3] and [11].

In analogy with the notation of the previous section, we denote by $\mathbb{E}^{\lambda}_{a,\Omega}(A)$ the average of an observable $A:\mathfrak{D}\to\mathbb{R}$ with respect to the probability weight in (3.1); again, truncated expectations are denoted by semicolons. Given an edge e, we denote by $\mathbb{1}_e$ the corresponding “dimer observable,” i.e., the characteristic function of the event “e belongs to the dimer configuration”; (truncated) expectations of products of dimer observables will be referred to as (truncated) dimer correlations. Another important observable is the height function h, which is defined on the faces of G_{Ω_a} as follows: fix arbitrarily a face η_0 of G_{Ω_a}, and set $h(\eta_0) = 0$; the value of the height on the other faces is fixed by letting the gradients be

$$h(\eta') - h(\eta) = \sum_{e \in C_{\eta \to \eta'}} \sigma_e(\mathbb{1}_e - 1/4), \tag{3.2}$$

where $C_{\eta\to\eta'}$ is a nearest neighbor path on the dual of G_{Ω_a} from the face η to the face η', the sum is over the edges crossed by this path, and σ_e is a sign, equal to $+$ or $-$ depending on whether the oriented path $C_{\eta\to\eta'}$ crosses e with the white site on the right or left, respectively. The definition (3.2) is well posed because the right side does not depend[3] on the choice of the path $C_{\eta\to\eta'}$.

The probability measure $\mathbb{P}^{\lambda}_{a,\Omega}$ depends on the parameters t_1, t_2, t_3, λ and on the interaction V. Let us fix the latter once and for all. We are interested in identifying choices of t_1, t_2, t_3, λ producing a nontrivial scaling limit as $a \to 0$ and/or $\Omega \nearrow \mathbb{R}^2$. However, contrary to the Ising case, the properties of the limiting distribution are extremely sensitive to the shape of Ω, to the choice of its discretization and on the boundary conditions. Let us first consider the case that $\Omega = \Omega_L$ is a torus, centered at the origin, whose horizontal and vertical sides are both of length L. In the limit $L \to \infty$, the expectation of the height function converges to a linear profile with slope $\rho = \rho(t_1, t_2, t_3, \lambda) \in \mathbb{R}^2$:

$$\lim_{L\to\infty} \mathbb{E}^{\lambda}_{a,\Omega}\big(ah(\eta_x)\big) = \rho \cdot x, \quad \forall x \in \mathbb{R}^2 \tag{3.3}$$

where, for $x \in \mathbb{R}^2$, η_x is the face whose bottom vertex is black, of coordinate $[x] := a\lfloor a^{-1}x\rfloor$. An alternative way of computing the function $\rho(t_1, t_2, t_3, \lambda)$ is via the Legendre transform of the free energy of the system with respect to a suitable "magnetic field" $B \in \mathbb{R}^2$: let $t_1(B) = t_1 e^{-B_1}$, $t_2(B) = t_2 e^{-B_1 - B_2}$, $t_3(B) = t_3 e^{-B_2}$, and

$$F(B) := \lim_{L\to\infty} L^{-2} \log Z^{\lambda}_{a,\Omega}(B), \tag{3.4}$$

with $Z^{\lambda}_{a,\Omega}(B) = \sum_{D\in\mathfrak{D}}(\prod_{e\in D} t_{r(e)}(B))e^{\lambda V(D)}$ the partition function with B-dependent weights, and define the *surface tension* $\sigma : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ as

$$\sigma(s) = \sup_B\big\{s \cdot B - (B_1 + B_2)/2 - F(B)\big\}. \tag{3.5}$$

Then the average slope ρ in (3.3) is the unique minimizer of σ with respect to s. If ρ belongs to the region $\mathcal{C}$ where σ is *strictly convex* and twice differentiable, we also expect that the height fluctuations on top of the linear profile with slope ρ are universally described by a Gaussian Free Field (GFF), in the sense that, for any C^∞ compactly supported test function $\psi : \mathbb{R}^2 \to \mathbb{R}$ such that $\int_{\mathbb{R}^2} \psi(x)dx = 0$ and any $\alpha \in \mathbb{R}$, letting $h^a(\psi) = a^2 \sum_{x\in\Omega_a} \psi(x)(h(\eta_x) - a^{-1}\rho \cdot x)$,

$$\lim_{a\to 0}\lim_{L\to\infty} \mathbb{E}^{\lambda}_{a,\Omega}(e^{i\alpha h^a(\psi)}) = e^{-\frac{\alpha^2}{2}\int_{\mathbb{R}^2} dx \int_{\mathbb{R}^2} dy\, \psi(x)\psi(y) G_\rho(x,y)} \tag{3.6}$$

[3] More precisely, the values of the right-hand side of (3.2) computed along two paths $C_{\eta\to\eta'}$ and $C'_{\eta\to\eta'}$ are the same if the loop obtained by concatenating $C_{\eta\to\eta'}$ with the path obtained by reversing the orientation of $C'_{\eta\to\eta'}$ is contractible. If Ω is a torus, then the two values may differ by a quantity depending on the windings of such a loop.

where G_ρ is the Green's function, i.e., the inverse of $-\Delta_\rho := -\sum_{i,j=1}^2 \partial_i(\sigma_{ij}(\rho)\partial_j)$, with $\sigma_{ij}(\rho)$ the elements of the Hessian of σ at ρ.

We are now in the position of formulating a conjecture on the scaling limit of the dimer model in more general domains, involving the surface tension σ introduced above and the region $\mathcal{C}$ of slopes where σ is strictly convex and twice differentiable. Suppose, e.g., that Ω is an open, finite, simply connected region of the plane. Let $\bar{h} : \Omega \to \mathbb{R}$ be a continuous function that extends continuously to $\partial\Omega$, which is a "dimer limit shape," in the sense that it is the unique minimizer of $\int_\Omega \sigma(\nabla h)dx$ with boundary condition $h|_{\partial\Omega} = \bar{h}|_{\partial\Omega}$, and suppose that it has "no frozen regions," in the sense that $\nabla\bar{h}$ belongs to $\mathcal{C}$ for almost-every $x \in \Omega$. Then we expect that there exists a sequence of discretizations Ω_a of Ω such that the average limiting height profile is exactly $\bar{h}$, i.e., $\lim_{a\to 0} \mathbb{E}^\lambda_{a,\Omega}(ah(\eta_x)) = \bar{h}(x)$, $\forall x \in \Omega$, and the scaling limit of the height fluctuations around $\bar{h}$ is a GFF, in the sense that, for any C^∞ compactly supported test function $\psi : \Omega \to \mathbb{R}$,

$$\lim_{a\to 0} \mathbb{E}^\lambda_{a,\Omega}\big(e^{i\alpha h^a(\psi)}\big) = e^{-\frac{\alpha^2}{2}\int_{\mathbb{R}^2} dx \int_{\mathbb{R}^2} dy\, \psi(x)\psi(y) G_{\bar{h},\Omega}(x,y)} \tag{3.7}$$

where $G_{\bar{h},\Omega}$ is the inverse of the operator $-\Delta_{\bar{h}}$ on Ω defined by

$$(-\Delta_{\bar{h}} f)(x) := -\sum_{i,j=1}^2 \partial_i\big(\sigma_{ij}\big(\bar{h}(x)\big)\partial_j f(x)\big), \tag{3.8}$$

with zero Dirichlet boundary conditions at $\partial\Omega$. The expected structure of the scaling limit is even richer than what emerges from the previous discussion, e.g., it turns out that the GFF nature of the height fluctuations is strictly (and subtly) related to the mesoscopic and macroscopic behavior of the dimer correlations, as it will become clearer from the discussion in the next two subsections. Moreover, the conjectured GFF nature of the height field comes together with complementary (and even harder-to-prove) predictions on the monomer and "vertex," or "electric," correlation functions, whose precise description, however, goes beyond the purpose of this review.

3.1. The non-interacting case

As anticipated above, and in analogy with what we saw for the Ising model, at $\lambda = 0$ the dimer model is exactly solvable [64,93]. Let us review here a few aspects of the solution, and let us focus for simplicity on the case that $\Omega = \Omega_L$ is a square torus of side L, as described before (3.3). For any finite a and L, the partition function can be expressed as the linear combination of the determinants of four variants of the so-called Kasteleyn matrix $K = K(t_1, t_2, t_3)$ (a complex adjacency matrix of G_{Ω_a}), the four variants differing for the boundary conditions along the edges "winding up" over the torus, which can be periodic or antiperiodic in the horizontal and vertical directions. Moreover, the multipoint dimer correlations are (linear combinations of four) determinants of minors of K^{-1}. Starting from these explicit formulas, one can easily compute the limit of the dimer correlations as $L \to \infty$, thus finding, in particular, that, for any two edges e, e', letting $r(e) \equiv r$, $r(e') \equiv r'$, $x(e) \equiv x$,

$x(e') \equiv x'$:

$$\lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_e \mathbb{1}_{e'}) = K_r K_{r'} \det \begin{pmatrix} K^{-1}(x+av_r, x) & K^{-1}(x+av_r, x') \\ K^{-1}(x'+av_{r'}, x) & K^{-1}(x'+av_{r'}, x') \end{pmatrix}, \tag{3.9}$$

where $K_r = i^{r-1} t_r$, $v_1 = (0,0)$, $v_2 = (-1,0)$, $v_3 = (-1,-1)$, $v_4 = (0,-1)$, and K^{-1} is the inverse Kasteleyn matrix in the thermodynamic limit, which reads

$$K^{-1}(x,y) = \int_{[-\pi,\pi]^2} \frac{dk}{(2\pi)^2} \frac{e^{-ia^{-1}k\cdot(x-y)}}{\mu(k)}, \tag{3.10}$$

with $\mu(k) = t_1 + it_2 e^{ik_1} - t_3 e^{i(k_1+k_2)} - ie^{ik_2}$ being the "dispersion relation." Using also the fact that $\lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_e) = K_r K^{-1}(x+av_r, x)$ and $\lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_{e'}) = K_{r'} K^{-1}(x'+av_{r'}, x')$, one finds that the truncated dimer-dimer correlation reads:

$$\lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_e; \mathbb{1}_{e'}) = -K_r K_{r'} K^{-1}(x+av_r, x') K^{-1}(x'+av_{r'}, x), \tag{3.11}$$

whose large distance decay properties are dictated by those of K^{-1}. In turn, these depend on the singularity structure of $\mu(k)$: if μ has two simple zeros, denoted p_+ and p_-, a simple asymptotic computation shows that at large distances

$$K^{-1}(x,y) = \frac{a}{2\pi} \sum_{\omega=\pm} \omega \frac{e^{-ia^{-1}p_\omega\cdot(x-y)}}{\phi_\omega(x-y)} + O\big((a/|x-y|)^2\big), \tag{3.12}$$

where $\phi_\omega(x) = \beta_\omega x_1 - \alpha_\omega x_2$, with $\alpha_\omega = \partial_{k_1}\mu(p_\omega)$ and $\beta_\omega = \partial_{k_2}\mu(p^\omega)$. In view of (3.11),

$$\begin{aligned} \lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_e; \mathbb{1}_{e'}) = {} & \frac{a^2}{4\pi^2} \sum_{\omega=\pm} \frac{K_{\omega,r} K_{\omega,r'}}{(\phi_\omega(x-x'))^2} \\ & + \frac{a^2}{4\pi^2} \sum_{\omega=\pm} \frac{K_{-\omega,r} K_{\omega,r'}}{|\phi_\omega(x-x')|^2} e^{ia^{-1}(p_\omega - p_{-\omega})\cdot(x-x')} \\ & + O\big((a/|x-x'|)^{-3}\big), \end{aligned} \tag{3.13}$$

where $K_{\omega,r} = K_r e^{-ip_\omega\cdot v_r}$. Notice that both the first and second terms on the right-hand side decay at large distances (compared to the lattice spacing) as $(a/|x-x'|)^2$, but the second term behaves differently from the first because it wildly oscillates on the lattice scale.

From these formulas, via (3.2), one can compute the average height profile and the asymptotics of the height fluctuations around the average. In particular, using the expression for the dimer one-point function, we find that the two components of the average slope at $\lambda = 0$, in the sense of (3.3), are $\rho_j = \sum_{e\in C_{\eta\to\eta+a\hat{e}_j}} \sigma_e(K_{r(e)} K^{-1}(x(e)+av_{r(e)}, x(e)) - 1/4)$, with $j = 1, 2$, for any face η (here $\hat{e}_j$, is the unit coordinate vector in direction $j \in \{1,2\}$). Remarkably, ρ belongs to the region $\mathcal{C}$ where the surface tension σ is strictly convex and twice differentiable iff μ has two distinct zeros. In this case, by computing the asymptotics of the height fluctuations around the average height profile, we find, as expected, a GFF behavior: consider, e.g., four distinct points in the plane, $x_1, \dots, x_4 \in \mathbb{R}^2$; using (3.2), write the covariance of the height differences between the faces at x_1, x_2, and at x_3, x_4 as

$$\begin{aligned} & \lim_{a\to 0} \lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}\big(h(\eta_{x_1}) - h(\eta_{x_2}); h(\eta_{x_3}) - h(\eta_{x_4})\big) \\ & \quad = \lim_{a\to 0} \sum_{\substack{e\in C_{\eta_{x_1}\to\eta_{x_2}} \\ e'\in C_{\eta_{x_3}\to\eta_{x_4}}}} \sigma_e \sigma_{e'} \lim_{L\to\infty} \mathbb{E}^0_{a,\Omega}(\mathbb{1}_e; \mathbb{1}_{e'}), \end{aligned} \tag{3.14}$$

and plug the asymptotic formula for the truncated dimer–dimer correlation (3.13) on the right-hand side of this equation; using the independence of the right-hand side of (3.14) from the choice of $C_{\eta_{x_1}\to\eta_{x_2}}, C_{\eta_{x_3}\to\eta_{x_4}}$, choose these lattice paths to be well separated: by doing so, one finds that both the remainder $O((a/|x-x'|)^3)$ and the wildly oscillating terms in (3.13) give subdominant contributions to the right-hand side of (3.14), in the $a\to 0$ limit; we are then left with the contribution from the first term on the right-hand side of (3.13) and, using the remarkable fact that, for any $x\in\mathbb{R}^2$ and $j\in\{1,2\}$,

$$\sum_{e\in C_{\eta_x\to\eta_{x+ae_j}}} \sigma_e K_{\omega,r(e)} = -i\omega\partial_j\phi_\omega(x), \tag{3.15}$$

we finally get

$$\begin{aligned}&\lim_{a\to 0}\lim_{L\to\infty}\mathbb{E}^0_{a,\Omega}\big(h\big(\eta_{x_1}-h(\eta_{x_2}); h(\eta_{x_3})-h(\eta_{x_4})\big)\\ &\quad= -\frac{1}{2\pi^2}\,\mathrm{Re}\int_{\phi_+(x_1)}^{\phi_+(x_2)} dz\int_{\phi_+(x_3)}^{\phi_+(x_4)} dz'\,\frac{1}{(z-z')^2}\\ &\quad= \frac{1}{2\pi^2}\,\mathrm{Re}\log\frac{(\phi_+(x_4)-\phi_+(x_1))(\phi_+(x_3)-\phi_+(x_2))}{(\phi_+(x_4)-\phi_+(x_2))(\phi_+(x_3)-\phi_+(x_1))}.\end{aligned} \tag{3.16}$$

Similar computations can be performed for higher moments of the height fluctuations, from which one finds that, for any $n>2$ and any $2n$-ple of distinct points $x_1,\dots,x_{2n}$, $\lim_{a\to 0}\lim_{L\to\infty}\mathbb{E}^0_{a,\Omega}\,(h(\eta_{x_1})-h(\eta_{x_2});\dots;h(\eta_{x_{2n-1}})-h(\eta_{x_{2n}}))=0$. As a corollary, one finds (3.6) at $\lambda=0$, with $G_\rho(x,y)=-\frac{1}{2\pi^2}\log|\phi_+(x-y)|$. Similar results can be extended to the case of finite, simply connected, domains of arbitrary shape: in particular, the GFF behavior of the height field in the sense of (3.7)–(3.8) has been proved in [69,72].

Note that, remarkably, the prefactor in front of the logarithm on the right-hand side of (3.16) (the "stiffness" of the GFF) is independent of the slope; equivalently, $\det\sigma_{ij}(\rho)\equiv\pi^2$, irrespective of ρ, provided $\rho\in\mathcal{C}$. This is a very special property of the non-interacting model, related to the fact that the spectral curve is an algebraic Harnack curve [70], and it is not expected to be robust under the addition of interactions. More in general, one expects that the GFF behavior of the height fluctuation relies on a subtle relation between the "stiffness" coefficient of the GFF, equal to $1/(2\det\sigma_{ij}(\rho))$, and the critical exponent associated with the oscillating part of the dimer–dimer correlation. Such a connection is a restatement, in the dimer context, of a deep universality relation predicted by Kadanoff [62] and Haldane [53] for vertex models and Luttinger liquids, based on Coulomb gas and bosonization methods. In the next section I will present a rigorous statement of the Kadanoff–Haldane relation for interacting dimer models and I will discuss its role in the proof of the GFF behavior of the height field.

3.2. Interacting dimer models

Let us now consider interacting dimers, described by (3.1) with $\lambda\neq 0$. In this case, the exact solvability of the model and its underlying determinant structure break down, and no thermodynamic or correlation function can be written explicitly, in closed form. This is the same as for the Ising model, but actually, compared with the Ising case, here things are even more subtle: while in the class of non-planar perturbations of the Ising model the scaling limit is expected to be described by the *same* critical exponents as the nearest-neighbor model, the interacting dimer correlations are expected to display a complex behavior, with different decay exponents associated with their oscillatory and non-oscillatory parts; in particular, the decay exponent associated with the oscillatory part of the two-point dimer correlation (i.e., the analogue of the second term on the right-hand side of (3.13)) is expected to be *anomalous*, i.e., to depend continuously and non-trivially on λ, and to be related by a simple, universal, relation to the stiffness coefficient of the height function.

As an illustration of the non-trivial large distance behavior of the correlation functions of the interacting model, let me first state a result about the asymptotics of the two-point dimer correlation, which generalizes equation (3.13) to the case $\lambda \neq 0$. Consider, again, the case that $\Omega = \Omega_L$ is a torus of side L centered at the origin, and, given two edges e, e', let $x(e) \equiv x$, $x(e') \equiv x'$, $r(e) \equiv r$, $r(e') \equiv r'$. Then the following holds:

Theorem 3.1 ([49,50]). *Let t_1, t_2, t_3 be such that $\mu(k)$ has two distinct nondegenerate zeros, $p_\pm$. Then there exist constants $C, \lambda_0 > 0$ and functions $K^\lambda_{\omega,r}$, $H^\lambda_{\omega,r}$, α^λ_ω, β^λ_ω, p^λ_ω, $\nu(\lambda)$, analytic in λ for $|\lambda| < \lambda_0$, for which, letting $\phi^\lambda_\omega(x) = \beta^\lambda_\omega x_1 - \alpha^\lambda_\omega x_2$,*

$$\begin{aligned}\lim_{L\to\infty} \mathbb{E}^\lambda_{a,\Omega}(\mathbb{1}_e; \mathbb{1}_{e'}) = {} & \frac{a^2}{4\pi^2} \sum_{\omega=\pm} \frac{K^\lambda_{\omega,r} K^\lambda_{\omega,r'}}{(\phi^\lambda_\omega(x-x'))^2} \\ & + \frac{a^{2\nu(\lambda)}}{4\pi^2} \sum_{\omega} \frac{H^\lambda_{-\omega,r} H^\lambda_{\omega,r'}}{|\phi^\lambda_\omega(x-x')|^{2\nu(\lambda)}} e^{ia^{-1}(p^\lambda_\omega - p^\lambda_{-\omega})\cdot(x-x')} \\ & + O\Big(\Big(\frac{a}{|x-x'|}\Big)^{3-C|\lambda|}\Big). \end{aligned} \tag{3.17}$$

Moreover, $K^0_{\omega,r} = H^0_{\omega,r} = K_{\omega,r}$, $\alpha^0_\omega = \partial_{k_1}\mu(p_\omega)$, $\beta^0_\omega = \partial_{k_2}\mu(p_\omega)$, $p^0_\omega = p_\omega$, $\nu(0) = 1$,

$$\overline{\alpha^\lambda_\omega} = -\alpha^\lambda_{-\omega}, \quad \overline{\beta^\lambda_\omega} = -\beta^\lambda_{-\omega}, \quad \overline{K^\lambda_{\omega,r}} = K^\lambda_{-\omega,r}, \quad \overline{H^\lambda_{\omega,r}} = H^\lambda_{-\omega,r}, \quad p^\lambda_+ + p^\lambda_- = (\pi,\pi), \tag{3.18}$$

and, generically in the choice of the interaction V, $\nu(\lambda)$ depends nontrivially on λ, i.e., $\nu'(0) \neq 0$.

The proof of the theorem provides a constructive algorithm for computing the coefficients of the convergent power series in λ for $K^\lambda_{\omega,r}$, $H^\lambda_{\omega,r}$, etc., but does not provide closed formulas for any of them. By comparing (3.17) with (3.13), it is apparent that the interaction modifies the scaling of the oscillatory part of the dimer–dimer correlation, which acquires the “anomalous” critical exponent $\nu(\lambda)$: this may be larger or smaller than 1, depending on the sign of λ; therefore, depending on whether the dimer interaction is repulsive or attractive, the oscillatory term, in absolute value, may be dominant or subdominant at large distances with respect to the nonoscillatory term. Let us remark that an explicit computation [51] shows that, generically, not only $\nu'(0)$ is different from zero, but it also depends explicitly upon the average slope $\rho_j = \rho_j(t_1, t_2, t_3, \lambda) = \sum_{e \in C_{\eta\to\eta+a\hat{e}_j}} \sigma_e(\lim_{L\to\infty} \mathbb{E}^\lambda_{a,\Omega}(\mathbb{1}_e) - 1/4)$.

Once that the sharp asymptotics for the two-point dimer correlation is known, we can compute the variance of height fluctuations, in analogy with (3.14) and following discussion. With the same notation and assumptions as in (3.14), we write

$$\begin{aligned}&\lim_{a\to 0}\lim_{L\to\infty}\mathbb{E}^{\lambda}_{a,\Omega}\big(h(\eta_{x_1})-h(\eta_{x_2});h(\eta_{x_3})-h(\eta_{x_4})\big)\\ &\quad=\lim_{a\to 0}\sum_{\substack{e\in C_{\eta_{x_1}\to\eta_{x_2}}\\ e'\in C_{\eta_{x_3}\to\eta_{x_4}}}}\sigma_e\sigma_{e'}\lim_{L\to\infty}\mathbb{E}^{\lambda}_{a,\Omega}(\mathbb{1}_e;\mathbb{1}_{e'});\end{aligned}\tag{3.19}$$

then we plug the asymptotics (3.17) on the right-hand side of this equation, and by choosing the paths $C_{\eta_{x_1}\to\eta_{x_2}}, C_{\eta_{x_3}\to\eta_{x_4}}$ well separated, we find that the contributions to the variance from the second and third terms on the right-hand side of (3.17) vanish as $a\to 0$. So also in the interacting case, we are left with the contribution to the variance from the non-oscillating term in (3.17), which we need to evaluate in the $a\to 0$ limit. In order for the involved sums to converge to well-defined, path-independent integrals, we need the analogue of (3.15) to hold in the interacting case, too. This is very hard, if not impossible, to check directly, due to the fact that the coefficients of the convergent power series in λ for $K^{\lambda}_{\omega,r}$ and ϕ^{λ}_{ω} are defined by extremely complicated, and different, algorithms. Nevertheless, we succeeded in proving the validity of an interacting analogue of (3.15), by making use of lattice Ward Identities (WI), in combination with hidden, chiral, WI for a continuum reference model that, in an appropriate sense, describes the *infrared fixed point* of the interacting dimer model (see next section for a few additional comments on the ideas of the proof):

Theorem 3.2 ([48, 50]). *Under the same assumptions as Theorem 3.1, one has*

$$\sum_{e\in C_{\eta\to\eta+a\hat{e}_j}}\sigma_e K^{\lambda}_{\omega,r(e)}=-i\omega\sqrt{\nu(\lambda)}\,\partial_j\phi^{\lambda}_{\omega}(x),\tag{3.20}$$

where $\nu(\lambda)$ is the same as in (3.17). Consequently,

$$\begin{aligned}&\lim_{a\to 0}\lim_{L\to\infty}\mathbb{E}^{\lambda}_{a,\Omega}\big(h\big(\eta_{x_1}-h(\eta_{x_2});h(\eta_{x_3})-h(\eta_{x_4})\big)\\ &\quad=\frac{\nu(\lambda)}{2\pi^2}\operatorname{Re}\log\frac{(\phi_+(x_4)-\phi_+(x_1))(\phi_+(x_3)-\phi_+(x_2))}{(\phi_+(x_4)-\phi_+(x_2))(\phi_+(x_3)-\phi_+(x_1))}.\end{aligned}\tag{3.21}$$

An elaboration of the proof also implies that, for any $n>2$ and any $2n$-tuple of distinct points $x_1,\dots,x_{2n}$, $\lim_{a\to 0}\lim_{L\to\infty}\mathbb{E}^{\lambda}_{a,\Omega}((h(\eta_{x_1})-h(\eta_{x_2}));\dots;(h(\eta_{x_{2n-1}})-h(\eta_{x_{2n}})))=0$, from which the asymptotic GFF behavior of the height field, in the sense of (3.6), follows. The reader should not underestimate the fact that the proof of such GFF behavior comes with an exact computation of the stiffness coefficient of the GFF, which turns out to be the *same* as the critical exponent $\nu(\lambda)$ of the dimer–dimer correlation. This is a universal relation among critical exponents, equivalent to those predicted by Kadanoff and Haldane in the closely related contexts of vertex, Ashkin–Teller, and Luttinger liquid models. In particular, it is equivalent to the identity $X_p=X_e/4$ [62, EQ. (13A)] between the polarization critical exponent X_p and the energy critical exponent X_e of the Ashkin–Teller model, an elusive exact scaling relation that Kadanoff predicted on the basis of formal bosonization methods and Coulomb gas techniques. In this sense, our result is a rigorous confirmation of the predictions of bosonization in the context of interacting dimer models, and it is related to the notion of "weak universality" discussed in Baxter's book [11]; see also [48, SECTION 1] and [76] for additional discussions on the notions of bosonization and weak universality in the contexts of dimer, vertex, Ashkin–Teller and quantum spin chain models.

4. METHODS AND IDEAS BEHIND THE PROOFS

A common feature of the problems and results stated above is that they concern non-solvable 2D models in the vicinity of an exactly solvable reference model at its *free Fermi point*: this is a way of saying that the reference nearest-neighbor Ising and dimer models are exactly solvable in terms of determinants of appropriate, explicit, matrices. As well known [29, 43, 87], this allows us to express the partition and generating function of correlations of the reference, solvable, models in terms of Gaussian Grassmann integrals: in particular, for any $a > 0$ and finite Ω, the partition function of the dimer model at $\lambda = 0$ can be written as

$$Z^0_{a,\Omega} = \int D\phi e^{-(\phi^+, K\phi^-)}, \tag{4.1}$$

where K is the Kasteleyn matrix, and $\phi = \{\phi^+_x, \phi^-_x\}_{x\in\Omega_a} \equiv (\phi^+, \phi^-)$ is a collection of Grassmann variables; the Ising partition function can be written analogously, with K replaced by a different, but still explicit, matrix, and ϕ replaced by a collection of $4|\Omega_a|$ "real" Grassmann variables. Similarly, the generating functions of the dimer or energy correlations, in the two cases of dimers and Ising, can be also written as Gaussian Grassmann integrals, from which one can easily get closed formulas for the corresponding multipoint dimer or energy correlations, and prove that they satisfy an exact fermionic Wick rule at the lattice level, i.e., that they have determinant, or Pfaffian, form.

Another common feature of the dimer and Ising models discussed in this paper is that their interacting, non-solvable versions can be formulated exactly, at finite a and finite Ω, in terms of non-Gaussian Grassmann integrals [7, 44, 50]. For instance, the partition function of the interacting dimer model can be written as follows (again, the one of the non-planar Ising model admits an analogous representation):

$$Z^\lambda_{a,\Omega} = Z^0_{a,\Omega} \int P(D\phi) e^{V(\phi)}, \tag{4.2}$$

where $P(D\phi) = D\phi e^{-(\phi^+, K\phi^-)} / \int D\phi e^{-(\phi^+, K\phi^-)}$, and V is a "potential" of strength λ, which can be written as the sum of monomials in ϕ of order 2, 4, 6, etc., with kernels that are analytic in λ in a small neighborhood of the origin and decay exponentially to zero at large distances, with rate proportional to the inverse lattice spacing. The term in V that is quadratic in ϕ can be isolated from the rest of the potential and combined with the Gaussian "measure" $P(D\phi)$; after this rearrangement, the potential contains a quartic term, plus higher order subdominant terms. In this sense, both the interacting Ising and dimer models take the form of Grassmannian ϕ^4_d models in dimension $d = 2$, somewhat reminiscent of the ϕ^4_d models studied by the constructive Quantum Field Theory (QFT) community since the early 1970s [52]. Note that (4.2) provides an explicit algorithm for computing all the coefficients of the perturbative series in λ for the partition function (similar considerations hold for the free energy and generating function of correlations): it is enough to expand the exponential and compute term by term the expectation of V^n with respect to the Gaussian measure $P(D\phi)$, which can be easily done in terms of the fermionic Wick rule. What makes things nontrivial is the fact that the covariance of the reference Gaussian measure, which for dimers is the (finite volume analogue of the) inverse Kasteleyn matrix K^{-1} in (3.10), at criticality decays algebraically to zero at large distances (criticality corresponds to the condition that $\mu(k)$ has two simple zeros, for dimers; and to the condition that β is set equal to the inverse critical temperature, for Ising). This implies that the naive bounds one can easily derive on the coefficient of the perturbative series are non-uniform in a and in the system size; in order to be able to control the thermodynamic and $a \to 0$ limits, one needs to exhibit subtle cancelations, whose identification requires systematic, multiscale resummations of the perturbation series, and which are very hard, if not impossible, to prove by direct inspection of the original series.

The approach developed over the years to identify these cancelations in critical systems is based on the ideas of Wilsonian RG. More specifically, the constructive RG approach used in the proofs of Theorems 2.1 to 3.2 is that developed by Benfatto, Gallavotti, Mastropietro, and coworkers [16, 17, 19] and reviewed in, e.g., [43, 75]; see also the more recent works [7, 47, 49], which review and provide a pedagogical introduction to this method in the specific contexts of non-planar Ising models, interacting dimer models, and fermionic ϕ^4_d theories with long-range interactions, respectively. At a very coarse level, the idea is to compute (4.2) recursively, first integrating out the degrees of freedom at length scales $\ell_0 2^{-N} \simeq a$ (here ℓ_0 is the length unit and $-N = \lfloor \log_2(a/\ell_0) \rfloor$), then those at length scales $\ell_0 2^{-N+1}, \ell_0 2^{-N+2}, \dots, \ell_0 2^{-h+1}$, etc. After each integration step, we re-express the partition and generating functions in a form analogous to (4.2), with $P(D\phi)$ replaced by a Gaussian measure with covariance supported on length scales $\gtrsim \ell_0 2^{-h}$ and $V(\phi)$ replaced by an effective interaction $V^{(h)}(\phi)$ that, up to a rescaling, has a form similar to the original $V(\phi)$, with modified coupling constants in front of the quadratic, quartic, sextic, etc., contributions. A dimensional power counting shows that the terms that tend to expand under iterations (and, therefore, to produce divergences in perturbation theory) are the quadratic and quartic ones, which tend to grow linearly and logarithmically in 2^{N-h}, respectively. These are the terms to be monitored and carefully looked at, in order to identify the cancelations that, if present, allow one to define a resummed, convergent perturbation theory.

Let me describe the procedure at a slightly more technical level, focusing, for illustrative purposes, on the dimer case with $\Omega = \Omega_L$ a torus of side L, and neglecting in the following discussion finite size effects, e.g., the difference between K^{-1} and its finite-L counterpart. While the scheme described below has several similarities with that used in the case of non-planar Ising models, there are also important differences (e.g., the presence for dimers of a non-trivial effective quartic coupling, denoted λ_h in the following), which I will comment about below.

In (4.2), we first rewrite the covariance $K^{-1}(x, y)$ of the Gaussian measure as a superposition of exponentially decaying "propagators," each characterized by an exponential decay rate $\propto 2^h$, $h \leq N$, that is, recalling (3.10) and (3.12), we rewrite $K^{-1}(x, y) =$

$\sum_{\omega=\pm}\sum_{h\le N} e^{-ia^{-1}p_\omega\cdot(x-y)} g^{(h)}_\omega(x,y)$, with $g^{(h)}_\omega(x,y)\simeq 2^h g^{(0)}_\omega(2^h x, 2^h y)$, and $g^{(0)}_\omega$, $\omega=\pm$, two smooth functions, exponentially decaying to zero on scale ℓ_0. Next, we rewrite the components of the random field ϕ with reference distribution $P(D\phi)$ as $\phi^\pm_x = \sum_{\omega=\pm} e^{\pm ia^{-1}p_\omega x}\phi^\pm_{\omega,x}$, and let $\phi_\omega = \phi^{(N)}_\omega + \phi^{(\le N-1)}_\omega$ (equality to be understood in distribution), with $\phi^{(N)}_\omega$ (resp. $\phi^{(\le N-1)}_\omega$) a Grassmann Gaussian field with covariance $g^{(N)}_\omega$ (resp. $g^{(\le N-1)}_\omega = \sum_{h\le N-1} g^{(h)}_\omega$); for brevity, we shall denote by $\phi^{(N)}$ the pair of fields $\{\phi^{(N)}_+, \phi^{(N)}_-\}$, and similarly for $\phi^{(\le N-1)}$. Correspondingly, we re-express the interacting partition function in (4.2) as follows (here $V^{(N)}$ is the same as V, thought of as a function of $\phi^{(N)}+\phi^{(\le N-1)}$ rather than of ϕ):

$$\begin{aligned} Z^\lambda_{a,\Omega}/Z^0_{a,\Omega} &= \int P_{\le N-1}(D\phi^{(\le N-1)}) \int P_N(D\phi^{(N)}) e^{V^{(N)}(\phi^{(N)}+\phi^{(\le N-1)})} \\ &= e^{L^2 F_N} \int P_{\le N-1}(D\phi^{(\le N-1)}) e^{V^{(N-1)}(\phi^{(\le N-1)})}, \end{aligned} \tag{4.3}$$

where $P_N(D\phi)$ and $P_{\le N-1}(D\phi)$ are the Grassmann Gaussian integrations with covariances $\int P_N(D\phi)\phi^-_{\omega,x}\phi^+_{\omega',y} = \delta_{\omega,\omega'} g^{(N)}_\omega(x,y)$ and $\int P_{\le N-1}(D\phi)\phi^-_{\omega,x}\phi^+_{\omega',y} = \delta_{\omega,\omega'} g^{(\le N-1)}_\omega(x,y)$, respectively, and $L^2 F_N + V^{(N-1)}(\phi) = \log\int P_N(D\phi')e^{V(\phi'+\phi)}$, with $V^{(N-1)}(0)=0$; F_N is a single-scale contribution to the free energy, and $V^{(N-1)}$ is called the effective potential on scale 2^{-N+1}. Remarkably, F_N and the kernels of $V^{(N-1)}$ are *analytic* functions of λ, *uniformly* in a and L, thanks to the Grassmann nature of the theory (the key idea is that the nth order term in perturbation theory can be expressed in determinant form, thanks to a smart interpolation identity due to Battle, Brydges, Federbush, and Kennedy [8, 22, 23], and the determinants can be bounded in an optimal way, from the combinatorial point of view, thanks to the Gram–Hadamard inequality [43]); moreover, the kernels of $V^{(N-1)}$ decay exponentially to zero at large distances, with exponential rate $\propto 2^{N-1}$.

In the second line of (4.3), we isolate the quadratic terms of $V^{(N-1)}(\phi)$ from the quartic or higher-order terms, and we insert them in the reference Gaussian integration $P_{\le N-1}(D\phi)$, thus "dressing" it a little bit, the dressing corresponding to a small, $O(\lambda)$, change of the location of the zeros p_ω of $\mu(k)$, to an $O(\lambda)$ change of the "velocities" $\alpha_\omega = \partial_{k_1}\mu(p_\omega)$ and $\beta_\omega = \partial_{k_2}\mu(p_\omega)$, and to an overall rescaling by a multiplicative factor $Z_{N-1} = 1 + O(\lambda)$, which can be conveniently reabsorbed by rescaling the field ϕ by $\sqrt{Z_{N-1}}$. After the manipulation of these quadratic terms and this rescaling we are left with a modified effective interaction which includes a local quartic term, of the form $\lambda_{N-1} Z^2_{N-1} \int dx\,\phi^+_{+,x}\phi^-_{+,x}\phi^+_{-,x}\phi^-_{-,x}$, the constant λ_{N-1} playing the role of the effective interaction strength on scale 2^{-N+1}, plus a remainder, which is nonlocal, or involves monomials in ϕ of higher order than four.

We now iterate the procedure, and integrate out in the same fashion the fields on scales labeled by $N-1, N-2, \dots, h+1$, so that, for any $h\le N$, we rewrite:

$$Z^\lambda_{a,\Omega}/Z^0_{a,\Omega} = e^{L^2\sum_{h'=h+1}^N F_{h'}} \int P_{\le h}(D\phi^{(\le h)}) e^{V^{(h)}(\sqrt{Z_h}\phi^{(\le h)})}, \tag{4.4}$$

where, once again, the single-scale contributions to the free energy $F_{h'}$ and the kernels of the effective potential $V^{(h)}$ are analytic functions of λ, uniformly in a, L (but, in general, non-uniformly in $N-h$). The constant Z_h in the argument of the effective potential is the so-called wave-function renormalization, which plays the same role as the multiplicative factor Z_{N-1} introduced above, after the integration of the first scale. Moreover, $V^{(h)}(\sqrt{Z_h}\phi)$ consists of: (i) quadratic terms, which can be combined with the reference Gaussian integration $P_{\le h}(D\phi^{(\le h)})$, thus leading to an additional, iteratively defined, dressing of the effective covariance; (ii) a local quartic term, of the form $\lambda_h Z_h^2 \int dx \phi^+_{+,x}\phi^-_{+,x}\phi^+_{-,x}\phi^-_{-,x}$, with λ_h playing the role of the effective interaction strength at length scales $\propto 2^{-h}$; (iii) a remainder, including non-local interactions, or interactions of order higher than four in ϕ, called the *irrelevant* terms.

As mentioned above, while well defined at each scale, the procedure sketched above does not lead to bounds that, in general, are uniform in the number of iterations, $N-h$. Of course, in order to perform the scaling limit $a \to 0$ (which corresponds to the removal of the ultraviolet cutoff $N \to \infty$) and/or the thermodynamic limit $L \to \infty$ (which corresponds to the removal of an infrared cutoff $h_L = \lfloor \log_2(\ell_0/L) \rfloor \to -\infty$), we need to prove that the construction is well defined uniformly in $N-h$. Remarkably, it turns out that the bounds on the kernels of the effective potential are uniform in the number of iterations iff λ_h remains bounded and small, uniformly in the scale index: if this is the case, then all the irrelevant terms turn out to be bounded and small, too; in fact, they can be recasted in the form of uniformly convergent expansions in the *effective* couplings $\{\lambda_{h'}\}_{h\le h'\le N}$. In other words, all the potential sources of divergences are resummed into the scale-dependent couplings $\{\lambda_{h'}\}$ and the problem of proving bounds on the free energy and correlation functions of the dimer model that are uniform in the scale index translates into that of controlling the boundedness of the sequence of effective couplings. This is an enormous conceptual simplification because λ_h can be written as the solution to a finite difference equation, induced by the iterative integration procedure sketched above, known as the *beta function* equation, of the form $\lambda_{h-1} = \lambda_h + \beta_h(\lambda_h, \dots, \lambda_N)$, with $\beta_h(\lambda_h, \dots, \lambda_N) = c_{2,h}\lambda_h^2 +$ higher orders. A priori, the same general estimates leading to the aforementioned control on the irrelevant contributions to the effective potential tell us that $|\beta_h(\lambda_h, \dots, \lambda_N)| \le C_0 \varepsilon_h^2$, with $\varepsilon_h = \max_{h\le h'\le N} |\lambda_{h'}|$ and C_0 independent of h; therefore, using the fact that $\lambda_N = O(\lambda)$, we find $|\lambda_h| \le C|\lambda|(1+|\lambda|(N-h))$ for some h-independent constant C; the point now is to look more closely at the beta function equation and to try to identify a structure guaranteeing that λ_h behaves better than such a priori, general, bound. Explicit computations show that at second and third order $\beta_h(\lambda_h, \dots, \lambda_N)$ is bounded by (const.) $2^{h-N}\varepsilon_h^2$ and (const.) $2^{h-N}\varepsilon_h^3$, respectively. Analogous estimates at all orders would imply that $|\lambda_h| \le C|\lambda|$ uniformly in h, as desired. However, direct inspection of perturbation theory does not appear feasible, for bounding in a similar manner the general nth order contribution to the beta function.

The idea is to prove the desired cancelation via an indirect route: we introduce a reference model, which has the same beta function as the dimer model, asymptotically as $N-h \to \infty$, up to exponentially small corrections, smaller than $\varepsilon_h^2 2^{\theta(h-N)}$, for some $\theta \in (0,1)$. This reference model plays the role of the "infrared fixed point" of our Grassmann formulation of the dimer model and is a close relative of the Luttinger model, an exactly solvable model of interacting fermions in one dimension, originally solved by rigorous bosonization techniques by Mattis and Lieb [77]. The reference model we use is a variation of the same model, formulated in the Grassmann functional integral setting, differing from the original Luttinger model "just" by the choice of the ultraviolet regulator (this apparently innocent modification may a priori have serious consequences because exact integrability of the model requires a specific regularization scheme). Such a model displays additional symmetries as compared to the dimer model, most notably "local chiral gauge invariance", i.e., the model is formally (up to corrections due to the ultraviolet regulator) invariant under independent gauge transformations of the two chiral fields ϕ_ω, $\omega = \pm$. Chiral gauge invariance implies the validity of exact equations ("chiral Ward Identities") for the model's correlation functions; such equations include the so-called "anomaly terms", i.e., terms that would naively be zero if one neglected the effects of the ultraviolet regulator, which, remarkably, can be computed explicitly, in closed form. Combining such chiral Ward Identities (WI) with the so-called Schwinger–Dyson equation for the correlation functions, we are led to a closed formula for all the correlation functions of the model. Such closed formulas imply, in particular, that the effective coupling strength $\lambda_{h,\text{ref}}$ of the reference model is uniformly close to the corresponding bare coupling; and, in turn, this implies the asymptotic vanishing of the beta function both for the reference and the dimer model. They also imply that the large distance asymptotic behavior of the dimer correlation functions are the same as those of appropriate correlations of the reference model: from this we derive the asymptotic formula (3.17) and prove Theorem 3.1.

Not only that: the chiral WI imply exact identities ("scaling relations") relating different critical exponents, as well as critical exponents and the multiplicative prefactor in front of the density–density correlation of the reference model. Such exact identities, if compared and combined with the exact lattice WI satisfied by the dimer correlation functions, imply analogous scaling relations for the dimer model; in turn, the exact lattice WI of the dimer model are a consequence of the local conservation law for the number of incident dimers at each vertex. This is, at a rough level, the way in which we prove the identity (3.20), from which Theorem 3.2 and the GFF behavior of the height fluctuations follow.

Let me conclude this section by a brief discussion of how the previous strategy must be modified in order to prove Theorems 2.1 and 2.2 for non-planar Ising models. The general approach is the same: also the generating function of the energy correlations of these models can be expressed as non-Gaussian Grassmann integral, similar to (4.2). At the critical temperature, the covariance of the reference Gaussian integration decays algebraically to zero at large distances, and this implies that, in order to derive uniform bounds on the thermodynamic and correlation functions, we must appeal to a rigorous RG multiscale analysis. Therefore, also for Ising, we compute the non-Gaussian Grassmann functional integral in an iterative fashion, and we are led to the construction of a sequence of effective potentials $V^{(h)}$, in analogy with (4.4). However, a crucial difference is in the counting of the "critical" degrees of freedom of the effective theory. In the Ising case, the effective potential $V^{(h)}$ can be written as the function of a Grassmann field ϕ with two components per site rather than four (remember that in the dimer case, there were four Grassmann variables per site, $\phi^+_{+,x}$, $\phi^-_{+,x}$, $\phi^+_{-,x}$, $\phi^-_{-,x}$; in the Ising case, we have just two, $\phi_{+,x}$ and $\phi_{-,x}$, no ± label at the exponent): this implies that the local quartic term in the effective potential, the one that was so hard to control in the dimer case, is automatically zero because there is no nonvanishing quartic monomial that can be constructed (the Grassmann rule implies $\phi^2_{\omega,x} = 0$). This makes the construction of the non-planar Ising theory in the full-plane limit easier than that for interacting dimers (this also explain why Theorem 2.1, which involves non-planar Ising models in the full-plane limit, was proved already 10 years ago [44]).

The problem now is the extension to finite domains with open, or cylindrical, boundary conditions: in fact, the presence of boundaries produces an additional effective, scale-dependent, coupling, localized at the boundary, which is potentially logarithmically divergent, like the quartic effective coupling in the dimer setting. And, again, the proof that such additional boundary scale-dependent coupling remains bounded and small, uniformly in the scale index, requires to identify cancelations in its flow equation. In the Ising setting, these cancelations follow from an approximate image rule for the fermionic covariance at the boundary; once we identified this cancelation, we managed to extend the construction of the scaling limit of energy correlations to the cylindrical setting, thus proving Theorem 2.2. Our proof is currently restricted to a specific cylindrical geometry, which we need in order to identify the required boundary cancelations, and in order to obtain optimal bounds on the fermionic Green's function in the vicinity of the boundary. However, the technique itself underlying the proof of the theorem seems robust and I expect that it can be adapted, in perspective, to domains of arbitrary shape (even more, I expect that it will be capable to control the scaling limit of models in the Luttinger liquid universality class, such as interacting dimers, in finite domains). See next section for additional comments on these perspectives.

Due to space constraints, I cannot enter in more detail than this into the proofs of the main theorems presented in this paper. The purpose of this section was just to convey the main ideas we used and to highlight the strategy and main difficulties to be overcome in the proofs. For additional details, I refer the reader to the original papers [6,7,44,48–50].

5. FURTHER RESULTS, PERSPECTIVES, AND OPEN PROBLEMS

Let me conclude this review with a brief, certainly partial, discussion of related results, perspectives and open problems, whose understanding would represent in my opinion a major advance in our understanding of the scaling limit of 2D non-planar Ising models and interacting dimers models (as well as of related classes of non-integrable statistical mechanics models, such as Ashkin–Teller and vertex models). I will state explicitly only problems that are more directly connected with the results and methods reviewed in this paper. Of course, there are plenty of other challenging, extremely interesting, open problems, concerning, e.g., the scaling limit of critical interfaces [12,30], the limiting validity of Virasoro algebra for an appropriate class of "dressed" observables [55], and the construction of the massive scaling limit in the magnetic field direction [25,26] (for non-planar Ising models), or the scaling limit of vertex and monomer correlations [37], the scaling limit of the cycle-rooted spanning forest associated with the dimer configuration via the Temperley bijection

[48, SECT. 2.1.2], the validity of Cardy's formula [2,21,59] and, more generally, the computation of the subleading corrections to the free energy [27] (for interacting dimers).

5.1. Non-planar Ising models

Let us first consider the class of non-planar Ising models described above, in Section 2. There are a few extensions of the results of Theorems 2.1 and 2.2 which appear to be feasible on the basis of relatively straightforward extensions of the techniques underlying their proofs. I refer, in particular, to the computation of boundary spin correlations and boundary energy correlations (and mixed boundary spin, boundary energy, bulk energy correlations): it should be easy to show, on the basis of a mild extension of the proof of Theorem 2.2, that their scaling limit can be written as the Pfaffian of an explicit antisymmetric matrix, whose elements (involving the two-point boundary spin–spin correlations) can be written in closed form, and exhibit the expected boundary critical exponents. This would complement the results of [5], by computing explicitly the scaling limit for a wide class of nonplanar perturbations of the Ising model, not restricted to ferromagnetic pair interactions.

On the other hand, extension of Theorem 2.2, or of its expected analogue for boundary spin correlations, to domains of more general shapes than flat cylinders, appears to be much harder. Already in the $\lambda = 0$ case, the construction of the scaling limit in domains of arbitrary shape remained elusive for several decades, and has been completed in the last ten years thanks to the use of the highly nontrivial methods of discrete holomorphicity, in the form developed, among others, by Smirnov, Chelkak, Hongler, and Izyurov [31–33,56]. It would be extremely interesting to extend this construction to the interacting, non-planar case.

Open Problem 1. *Compute the scaling limit of the multipoint energy correlations of non-planar Ising models in domains $\Omega \subset \mathbb{R}^2$ of arbitrary shape, for different boundary conditions, say open, $+$ or $-$. As a corollary, prove conformal covariance of the limit.*

One possibility to attack this problem is to extend the strategy sketched in the previous section to more general domains (already the case of the rectangle is non-trivial). There are two key technical points to be understood: (1) how can we obtain sufficient control on the fermionic Green's function in situations where it cannot be diagonalized explicitly? (By "control" here I mean: define its multiscale decomposition, with optimal bounds on its asymptotic behavior in the bulk and close to the boundaries; moreover, derive a Gram representation for the single-scale Green's function, with optimal dimensional bounds on the L^∞-norm of the Gram vectors); (2) how do we prove the required cancelations on the boundary, "marginal," scale-dependent couplings? I believe that the most serious technical issue is the first. A solution may come from an effective combination of multiscale methods with those of discrete holomorphicity, which may lead to sharp bounds on the speed of convergence to the scaling limit already at the level of the $\lambda = 0$ theory.

In connection with this problem, I cannot avoid mentioning an exciting recent development due to Duminil-Copin and collaborators [38], who proved rotational invariance for the scaling limit (whenever it exists) of a wide class of 2D critical models, including Potts, 6V, and the random cluster model. The proof is based on completely different ideas than ours, and involve the coupling of different instances of these models on different isoradial graphs, characterized by different discrete rotational invariance properties, via a sequence of star-triangle transformations.

The first open problem, stated above, concerns energy correlations. Of course, analogous results for the spin correlations would also be extremely interesting. However, the study of the scaling limit of spin correlations is notoriously difficult, already in the full-plane limit, even in $\lambda = 0$ case. The reason is that the spin observable is non-local in the Grassmann representation, and, already in the integrable case, their understanding requires the use of special, sophisticated techniques. In the full-plane limit at $\lambda = 0$, one can use Szego's lemma to extract the asymptotics of the spin–spin correlations in special directions [79]; or, alternatively, one can use a set of quadratic finite difference equations, discovered by McCoy, Perk, and Wu [78], whose scaling limit is the Painlevé III equation. Multipoint spin correlations, both in the full plane limit and in finite domains, were understood much more recently, thanks to other, complementary techniques, namely discrete holomorphicity applied on a two-sheet discrete Riemann surface, associated with the original graph which the model is defined on, with cuts connecting the locations of the spin observables [31, 32, 36, 37]. It is unclear whether these ideas can be extended, and in case how, to the interacting setting, $\lambda \neq 0$.

Open Problem 2. *Compute the scaling limit of the spin correlations of non-planar Ising models, first in the full plane, then in finite domains.*

Already the case of the two-point spin correlation in the full plane limit is highly non-trivial, and its understanding would represent a breakthrough in the field, with a potential big impact on other problems that, at a heuristic level, are studied by formal bosonization and Coulomb gas techniques.

Another interesting set of open problems is related to the computation of the subleading corrections to the critical free energy of non-planar Ising models, which are expected to display subtle universality properties [27, 35]. Fix $\beta = \beta_c$, fix Ω, and compute the free energy for a small; it is expected that

$$\log Z^{\lambda}_{\beta_c;a,\Omega} = a^{-2}|\Omega| f(\lambda) + a^{-1}|\partial\Omega|\tau(\lambda) + c_\Omega(a,\lambda), \tag{5.1}$$

with $f(\lambda)$ and $\tau(\lambda)$ independent of Ω, and $c_\Omega(a,\lambda)$ of smaller order than $O(a^{-1})$. More precisely, it is expected that the behavior of this subleading term in the $a \to 0$ depends upon the Euler characteristics χ of Ω (recall that $\chi = V - E + F$ with V, E, F being the number of vertices, edges, and faces of any triangulation of Ω, e.g., $\chi = 0$ for Ω a torus or a cylinder, and $\chi = 1$ for Ω a finite, simply connected domain). If $\chi \neq 0$, it is expected that

$$\lim_{a\to 0} \frac{c_\Omega(a,\lambda)}{\log(a^{-1}|\partial\Omega|)} = -\frac{1}{12}\chi, \tag{5.2}$$

while, if $\chi = 0$, then

$$\lim_{a\to 0} c_\Omega(a,\lambda) = c^0_\Omega \quad \text{independent of } \lambda. \tag{5.3}$$

For example, if $\Omega = \Omega(\xi)$ is a torus with aspect ratio ξ,

$$c^0_\Omega = \log(\theta_2 + \theta_3 + \theta_4) - \frac{1}{3}\log(4\theta_2\theta_3\theta_4), \tag{5.4}$$

where $\theta_i = \theta_i(e^{-\pi\xi})$ are Jacobi theta functions.

Open Problem 3. *Prove* (5.2) *and* (5.4) *with an explicit expression for c^0_Ω for non-planar Ising models.*

So far, the only known rigorous result related to this conjecture is a proof of "Cardy's formula" [2,21] for toroidal domains $\Omega = \Omega(\xi)$ with aspect ratio going to infinity:

$$\lim_{\xi\to\infty}\frac{1}{\xi}\lim_{a\to 0} c_\Omega(a,\lambda) = \frac{\pi}{12}. \tag{5.5}$$

Here, the right-hand side is the $\xi \to \infty$ limit of the right-hand side of (5.4) divided by ξ. Equation (5.5) was proved in [46]. I believe that the methods developed in [6,7,50] for controlling finite size effects within the rigorous RG scheme described in Section 4 should be sufficient for proving (5.4) for the torus and the cylinder. Another story, which appears more challenging, is the case of $\chi \neq 0$. As far as I know, formula (5.2) is unproven even in the $\lambda = 0$ case (with the exception of the rectangle, in which case it was proved in [57]).

5.2. Interacting dimer models

Let us now consider the class of interacting dimer models discussed in Section 3. All the problems stated in the previous subsection in the context of non-planar Ising models have their counterparts for dimers. Due to the underlying "Luttinger liquid" nature [53] of the scaling limit, and the presence of non-trivial, anomalous, exponents, I expect that their solution will be even more challenging than the one for the corresponding Ising's problems.

Open Problem 4. *Prove the GFF nature of the scaling limit of the height fluctuations in arbitrary finite, simply connected, domains, in the sense of* (3.7)*.*

In order to prove such a statement via the multiscale methods sketched in Section 4, one will need to compute the dominant boundary corrections to the effective potentials, and, in particular, control the flow of the effective, marginal, boundary couplings (the dimer analogue of those discussed at the end of Section 4 for non-planar Ising). I expect that these boundary couplings will diverge exponentially in the limit of a large number of RG iterations, with a small, λ-dependent exponent, playing the role of an anomalous boundary critical exponent.

Open Problem 5. *Compute the asymptotic behavior, in the sense of* (3.17)*, for the two-point dimer correlation in a domain Ω with boundary, in the case in which at least one of the dimer observables is close to the boundary, and establish whether their oscillatory part exhibits an anomalous critical exponent $\nu_\partial(\lambda)$ different from the bulk one $\nu(\lambda)$; in case, compute such boundary exponent.*

Other interesting directions and open problems involve generalizations of the type of dimer interactions. For instance, rather than the class of interactions discussed in Section 3, one could imagine to break planarity of the model, by adding non-planar, non-nearest-neighbor, edges to the graph, which may be occupied by "long" dimers with small probability. Under appropriate conditions on the geometry of these nonplanar edges, e.g., if there exist lattice paths connecting faces microscopically close to any two points of the domain that never pass under the "bridges" formed by the non-planar edges[4] then it should still be possible to introduce a well-defined notion of height function. In such a situation, it would be interesting to test whether the GFF nature of the scaling limit of the height field persists, notwithstanding the loss of planarity of the model.

Open Problem 6. *Same as Open Problem 4, for weakly non-planar dimer models.*

I expect that, at least in the case of periodic models defined on a torus of side L, in the limit $L \to \infty$, a generalization of the method of proof of Theorem 3.2 will allow us to prove the convergence of a suitably defined height field to the massless GFF, in the sense of (3.6). Further extensions to different kind of dimer interactions appear more challenging. For instance, an extremely interesting problem that I propose here as my last Open Problem, is whether the GFF nature of the scaling limit of height fluctuations can be extended to the case of the interface between $+$ and $-$ phases in the 3D Ising model with "tilted" Dobrushin boundary conditions, at low enough temperatures. It is well known that the interface of the 3D Ising model with standard, flat, Dobrushin boundary conditions is rigid at low temperatures. This is not expected to be the case if the boundary conditions are assigned so that the interface has non-zero average slope. The problem of understanding the nature of fluctuations of this interface, even if apparently very different from those considered in this review, has surprisingly strict connections with that of the scaling limit of the height fluctuations for interacting dimers [28]: in fact, it is well known that the monotone height profiles of a tilted 3D Ising interface can be mapped exactly, in an invertible way, to those of the dimer model on the hexagonal lattice; moreover, under this mapping, the height distribution of the 3D Ising model at *zero* temperature is the same as that of the standard, integrable, dimer model. From this exact correspondence, the GFF nature of the height fluctuations for the 3D Ising tilted interface at zero temperature readily follows. At positive temperatures, there is no known coupling between the height distribution of the 3D Ising tilted interface with that of a dimer model; however, it is tempting to guess that the effect induced by the temperature is qualitatively the same as that of a weak, effective, interaction among dimers. If this were the case, then the methods of Theorem 3.2 would provide a possible strategy for proving the existence of a "rough phase" for the 3D Ising model.

[4] A concrete way of realizing this may be the following: consider a 2D periodic graph obtained by periodizing in two directions a planar fundamental cell G_0. Now make this non-planar by adding in each fundamental cell a number of non-planar bonds, with the restriction that they should not pass over the "corridors" between different copies of G_0. Even though the graph is non-planar, the height difference between faces in the corridors is well defined (use definition (3.2) with lattice paths passing only through the corridors).

Open Problem 7. *Prove that the fluctuations of the interface of the 3D Ising model with tilted Dobrushin boundary conditions at low temperatures converges in the scaling limit to a GFF.*

ACKNOWLEDGMENTS

I would like to thank all my collaborators on the problems of universality and scaling limits in SM and quantum many body theory, and all the colleagues who shared their ideas with me in many inspiring discussions, which were very influential for my research in this area. In particular, a special thank goes to Giovanni Antinucci, Rafael Greenblatt, Vieri Mastropietro, and Fabio Toninelli, for the enjoyable and fruitful collaborations on Ising and dimer models, which led to the results reviewed in this essay.

FUNDING

This work has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme ERC CoG UniCoSM, grant agreement no. 724939. I also acknowledge financial support from MIUR, PRIN 2017 project MaQuMA, PRIN201719VMAST01.

REFERENCES

[1] A. Abdesselam, A. Chandra, and G. Guadagni, Rigorous quantum field theory functional integrals over the p-adics I: anomalous dimensions. 2013, arXiv:1302.5971.

[2] I. Affleck, Universal term in the free energy at a critical point and the conformal anomaly. *Phys. Rev. Lett.* **56** (1986), 746–748.

[3] A. Aggarwal, Universality for Lozenge tiling local statistics. 2019, arXiv:1907.09991.

[4] M. Aizenman, D. J. Barsky, and R. Fernandez, The phase transition in a general class of Ising-type models is sharp. *J. Stat. Phys.* **47** (1987), 343–374.

[5] M. Aizenman, H. Duminil-Copin, V. Tassion, and S. Warzel, Emergent planarity in two-dimensional Ising models with finite-range Interactions. *Invent. Math.* **216** (2019), 661–743.

[6] G. Antinucci, A. Giuliani, and R. L. Greenblatt, Energy correlations of non-integrable Ising models: the scaling limit in the cylinder. 2020, arXiv:2006.04458.

[7] G. Antinucci, A. Giuliani, and R. L. Greenblatt, Non-integrable Ising models in cylindrical geometry: Grassmann representation and infinite volume limit. *Ann. Henri Poincaré* (2021). DOI 10.1007/s00023-021-01107-3.

[8] G. A. Battle and P. Federbush, A note on cluster expansions, tree graph identities, extra $1/N!$ factors!!! *Lett. Math. Phys.* **8** (1984), 55–57.

[9] R. Bauerschmidt, D. C. Brydges, and G. Slade, Scaling limits and critical behaviour of the 4-dimensional n-component $|\varphi|^4$ spin model. *J. Stat. Phys.* **157** (2014), 692–742.

[10] R. Bauerschmidt and C. Webb, The Coleman correspondence at the free fermion point. 2020, arXiv:2010.07096.

[11] R. J. Baxter, *Exactly solved models in statistical mechanics*. Academic Press Inc., London, 1989. Reprint of the 1982 original.

[12] V. Beffara, E. Peltola, and H. Wu, On the uniqueness of global multiple SLEs. *Ann. Probab.* **49** (2021), 400–434.

[13] A. A. Belavin, A. M. Polyakov, and A. B. Zamolodchikov, Infinite conformal symmetry of critical fluctuations in two dimensions. *J. Stat. Phys.* **34** (1984), 763–774.

[14] G. Benfatto, P. Falco, and V. Mastropietro, Extended scaling relations for planar lattice models. *Comm. Math. Phys.* **292** (2009), 569–605.

[15] G. Benfatto, P. Falco, and V. Mastropietro, Massless sine-Gordon and massive Thirring models: proof of Coleman's equivalence. *Comm. Math. Phys.* **285** (2009), 713–762.

[16] G. Benfatto and G. Gallavotti, Perturbation theory of the Fermi surface in quantum liquid. A general quasiparticle formalism and one-dimensional systems. *J. Stat. Phys.* **59** (1990), 541–664.

[17] G. Benfatto, G. Gallavotti, A. Procacci, and B. Scoppola, Beta function and Schwinger functions for a many fermions system in one dimension. Anomaly of the Fermi surface. *Comm. Math. Phys.* **160** (1994), 93–171.

[18] G. Benfatto and V. Mastropietro, Renormalization group, Hidden symmetries and approximate ward identities in the XYZ model. *Rev. Math. Phys.* **13** (2001), 1323–1435.

[19] G. Benfatto and V. Mastropietro, Ward identities and chiral anomaly in the Luttinger Liquid. *Comm. Math. Phys.* **258** (2005), 609–655.

[20] G. Benfatto and V. Mastropietro, Drude weight in non solvable quantum spin chains. *J. Stat. Phys.* **143** (2011), 251–260.

[21] H. W. J. Blote, J. L. Cardy, and M. P. Nightingale, Conformal invariance, the central charge and universal finite size amplitudes at criticality. *Phys. Rev. Lett.* **56** (1986), 742–745.

[22] D. Brydges and P. Federbush, A new form of the Mayer expansion in classical statistical mechanics. *J. Math. Phys.* **19** (1978), 2064.

[23] D. Brydges and T. Kennedy, Mayer expansions and the Hamilton–Jacobi equation. *J. Stat. Phys.* **48** (1987), 19–49.

[24] D. Brydges, P. Mitter, and B. Scoppola, Critical $(\phi^4)_{3,\epsilon}$. *Comm. Math. Phys.* **240** (2003), 281–327.

[25] F. Camia, C. Garban, and C. M. Newman, Planar Ising magnetization field II. Properties of the critical and near-critical scaling limits. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (2016), 146–161.

[26] F. Camia, J. Jiang, and C. M. Newman, Exponential decay for the near-critical scaling limit of the planar Ising model. *Comm. Pure Appl. Math.* **73** (2020), 1371–1405.

[27] J. L. Cardy and I. Peschel, Finite-size dependence of the free energy in two-dimensional critical systems. *Nuclear Phys. B* **300** (1988), 377–392.

[28] R. Cerf and R. Kenyon, The low-temperature expansion of the Wulff crystal in the 3D Ising model. *Comm. Math. Phys.* **222** (2001), 147–179.

[29] D. Chelkak, D. Cimasoni, and A. Kassel, Revisiting the combinatorics of the 2D Ising model. *Ann. Inst. Henri Poincaré D* **4** (2017), 309–385.

[30] D. Chelkak, H. Duminil-Copin, C. Hongler, A. Kemppainen, and S. Smirnov, Convergence of Ising interfaces to Schramm's SLE curves. *C. R. Math.* **352** (2014), 157–161.

[31] D. Chelkak, C. Hongler, and K. Izyurov, Conformal invariance of spin correlations in the planar Ising model. *Ann. of Math.* **181** (2015), 1087–1138.

[32] D. Chelkak, C. Hongler, and K. Izyurov, Correlations of primary fields in the critical Ising model. 2021, arXiv:2103.10263.

[33] D. Chelkak and S. Smirnov, Universality in the 2D Ising model and conformal invariance of fermionic observables. *Invent. Math.* **189** (2012), 515–580.

[34] J. Dimock, Nonperturbative renormalization of scalar quantum electrodynamics in $d = 3$. *J. Math. Phys.* **56** (2015), 102304.

[35] P. Di Francesco, H. Saleur, and J. B. Zuber, Critical Ising correlation functions in the plane and on the torus. *Nuclear Phys. B* **290** (1987), 527–581.

[36] J. Dubedat, Exact bosonization of the Ising model. 2011, arXiv:1112.4399.

[37] J. Dubedat, Dimers and families of Cauchy–Riemann operators, I. *J. Amer. Math. Soc.* **28** (2015), 1063–1167.

[38] H. Duminil-Copin, K. K. Kozlowski, D. Krachun, I. Manolescu, and M. Oulamara, Rotational invariance in critical planar lattice models. 2020, arXiv:2012.11672.

[39] H. Duminil-Copin and S. Smirnov, Conformal invariance of lattice models, Probability and statistical physics in two and more dimensions. *Clay Math. Proc.* **15** (2012), 213–276.

[40] P. Falco, Kosterlitz–Thouless Transition Line for the Two Dimensional Coulomb Gas. *Comm. Math. Phys.* **312** (2012), 559–609.

[41] M. E. Fisher, On the Dimer Solution of Planar Ising Models. *J. Math. Phys.* **7** (1966), 1776.

[42] R. H. Fowler and G. S. Rushbrooke, An attempt to extend the statistical theory of perfect solutions. *Trans. Faraday Soc.* **33** (1937), 1272–1294.

[43] G. Gentile and V. Mastropietro, Renormalization group for one-dimensional fermions. A review on mathematical results. *Phys. Rep.* **352** (2001), 273–343.

[44] A. Giuliani, R. L. Greenblatt, and V. Mastropietro, The scaling limit of the energy correlations in non-integrable Ising models. *J. Math. Phys.* **53** (2012), 095214.

[45] A. Giuliani and V. Mastropietro, Anomalous universality in the anisotropic Ashkin–Teller model. *Comm. Math. Phys.* **256** (2005), 681–735.

[46] A. Giuliani and V. Mastropietro, Universal Finite Size Corrections and the Central Charge in Non-solvable Ising Models. *Comm. Math. Phys.* **324** (2013), 179–214.

[47] A. Giuliani, V. Mastropietro, and S. Rychkov, Gentle introduction to rigorous Renormalization Group: a worked fermionic example. *J. High Energy Phys.* (2021), 26.

[48] A. Giuliani, V. Mastropietro, and F. Toninelli, Haldane relation for interacting dimers. *J. Stat. Mech.* (2017), 034002.

[49] A. Giuliani, V. Mastropietro, and F. Toninelli, Height fluctuations in interacting dimers. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (2017), 98–168.

[50] A. Giuliani, V. Mastropietro, and F. L. Toninelli, Non-integrable Dimers: Universal Fluctuations of Tilted Height Profiles. *Comm. Math. Phys.* **377** (2020), 1883–1959.

[51] A. Giuliani and F. Toninelli, Non-integrable dimer models: universality and scaling relations. *J. Math. Phys.* **60** (2019), 103301.

[52] J. Glimm and A. Jaffe, *Quantum Physics, A functional integral point of view. 2nd edn.* Springer, New York, 1987.

[53] F. D. M. Haldane, "Luttinger liquid theory" of one-dimensional quantum fluids. I. Properties of the Luttinger model and their extension to the general 1D interacting spinless Fermi gas. *J. Phys. C, Solid State Phys.* **14** (1981), 2585–2609.

[54] C. Hongler, *Conformal invariance of Ising model correlations*. Ph.D. thesis, Univ. Genéve, 2010.

[55] C. Hongler, F. Johansson Viklund, and K. Kytölä, Conformal field theory at the lattice level: discrete complex analysis and Virasoro structure. 2013, arXiv:1307.4104.

[56] C. Hongler and S. Smirnov, The energy density in the planar Ising model. *Acta Math.* **211** (2013), 191–225.

[57] A. Hucht, The square lattice Ising model on the rectangle II: finite-size scaling limit. *J. Phys. A: Math. Theor.* **50** (2017), 265205.

[58] C. A. Hurst and H. S. Green, New solution of the Ising problem for a rectangular lattice. *J. Chem. Phys.* **33** (1960), 1059–1062.

[59] N. S. Izmailian, V. B. Priezzhev, P. Ruelle, and C.-K. Hu, Logarithmic conformal field theory and boundary effects in the dimer model. *Phys. Rev. Lett.* **95** (2005), 260602.

[60] M. Kac and J. C. Ward, A combinatorial solution of the two-dimensional Ising model. *Phys. Rev.* **88** (1952), 1332–1337.

[61] L. P. Kadanoff, Correlations along a line in the two-dimensional Ising model. *Phys. Rev.* **188** (1969), 859–863.

[62] L. P. Kadanoff, Connections between the critical behavior of the planar model and that of the eight-vertex model. *Phys. Rev. Lett.* **39** (1977), 903–905.

[63] L. P. Kadanoff and H. Ceva, Determination of an operator algebra for the two-dimensional Ising model. *Phys. Rev. B* **3** (1971), 3918.

[64] P. W. Kasteleyn, The statistics of dimers on a lattice: I. The number of dimer arrangements on a quadratic lattice. *Physica* **27** (1961), 1209–1225.

[65] B. Kaufman and L. Onsager, Cristal statistics. III. Short range order in a binary Ising lattice. *Phys. Rev.* **76** (1949), 1244–1252.

[66] Y. Kawahigashi and R. Longo, Classification of local conformal nets. Case $c < 1$. *Ann. of Math.* **160** (2004), 493–522.

[67] R. Kenyon, Conformal invariance of domino tiling. *Ann. Probab.* **28** (2000), 759–795.

[68] R. Kenyon, Dominos and the Gaussian free field. *Ann. Probab.* **29** (2001), 1128–1137.

[69] R. Kenyon, Height fluctuations in the honeycomb dimer model. *Comm. Math. Phys.* **281** (2008), 675–709.

[70] R. Kenyon, A. Okounkov, and S. Sheffield, Dimers and amoebae. *Ann. of Math.* **163** (2006), 1019–1056.

[71] A. Kupiainen, R. Rhodes, and V. Vargas, Integrability of Liouville theory: proof of the DOZZ formula. *Ann. of Math.* **191** (2020), 81–166.

[72] B. Laslier, Central limit theorem for lozenge tilings with curved limit shape. 2021, arXiv:2102.05544.

[73] W. Lenz, Beitrag zum Verständnis der magnetischen Erscheinungen in festen Köpern. *Z. Phys.* **21** (1920), 613–615.

[74] V. Mastropietro, Ising models with four spin interaction at criticality. *Comm. Math. Phys.* **244** (2004), 595–642.

[75] V. Mastropietro, *Non-pertrubative renormalization*. World Scientific, 2008.

[76] V. Mastropietro, Universality, Phase Transitions and Extended Scaling Relations. In *Proceedings of the International Congress of Mathematicians*, Hyderabad, India, 2010.

[77] D. C. Mattis and E. H. Lieb, Exact Solution of a Many-Fermion System and Its Associated Boson Field. *J. Math. Phys.* **6** (1965), 304–312.

[78] B. M. McCoy, J. H. H. Perk, and T. T. Wu, Ising Field Theory: Quadratic Difference Equations for the n-Point Green's Functions on the Lattice. *Phys. Rev. Lett.* **46** (1981), 757.

[79] B. McCoy and T. Wu, *The two-dimensional Ising model*. Harvard Univ. Press, 1973.

[80] E. Montroll, R. Potts, and J. Ward, Correlation and spontaneous magnetization of the two dimensional Ising model. *J. Math. Phys.* **4** (1963), 308.

[81] L. Onsager, Critical statistics. A two dimensional model with an order–disorder transition. *Phys. Rev.* **56** (1944), 117–149.

[82] H. Pinson and T. Spencer, Universality and the two dimensional Ising model. Unpublished preprint.

[83] D. Poland, S. Rychkov, and A. Vichi, The conformal bootstrap: Theory, numerical techniques, and applications. *Rev. Modern Phys.* **91** (2019), 015002.

[84] J. Polchinski, Scale and conformal invariance in quantum field theory. *Nuclear Phys. B* **303** (1988), 226–236.

[85] A. M. Polyakov, Conformal symmetry of critical fluctuations. *JETP Lett.* **12** (1970), 381–383.

[86] L. Russo, A note on percolation. *Z. Wahrsch. Verw. Gebiete* **43** (1978), 39–48.

[87] S. Samuel, The use of anticommuting variable integrals in statistical mechanics. *J. Math. Phys.* **21** (1980), 2806.

[88] T. D. Schultz, D. Mattis, and E. H. Lieb, Two-dimensional Ising model as a soluble problem of many fermions. *Rev. Modern Phys.* **36** (1964), 856–871.

[89] P. D. Seymour and D. Welsh, Percolation probabilities on the square lattice. *Ann. Discrete Math.* **3** (1978), 227–245.

[90] S. Smirnov, Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *C. R. Acad. Sci., Sér. 1 Math.* **333** (2001), 239–244.

[91] S. Smirnov, Conformal invariance in random cluster models. I. Holmorphic fermions in the Ising model. *Ann. of Math.* **172** (2010), 1435–1467.

[92] T. Spencer, A mathematical approach to universality in two dimensions. *Phys. A* **279** (2000), 250–259.

[93] H. N. V. Temperley and M. E. Fisher, Dimer problem in statistical mechanics-an exact result. *Philos. Mag.* **6** (1961), 1061–1063.

[94] K. G. Wilson, The renormalization group: critical phenomena and the Kondo problem. *Rev. Modern Phys.* **47** (1975), 773–840.

[95] K. G. Wilson, The renormalization group and critical phenomena. *Rev. Modern Phys.* **55** (1983), 583–600.

[96] K. Wilson and J. B. Kogut, The Renormalization group and the epsilon expansion. *Phys. Rep.* **12** (1974), 75–199.

[97] C. N. Yang, The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.* **85** (1952), 808–816.

[98] A. B. Zamolodchikov, Irreversibility of the flux of the renormalization group in a 2D field theory. *JETP Lett.* **43** (1986), 730–732.

ALESSANDRO GIULIANI

Università degli Studi Roma Tre, Dipartimento di Matematica e Fisica, L.go S. L. Murialdo 1, 00146 Roma, Italy, and Centro Linceo Interdisciplinare *Beniamino Segre*, Accademia Nazionale dei Lincei, Palazzo Corsini, Via della Lungara 10, 00165 Roma, Italy, giuliani@mat.uniroma3.it

GAPPED QUANTUM SYSTEMS: FROM HIGHER-DIMENSIONAL LIEB–SCHULTZ–MATTIS TO THE QUANTUM HALL EFFECT

MATTHEW B. HASTINGS

ABSTRACT

We consider many-body quantum systems on a finite lattice, where the Hilbert space is the tensor product of finite-dimensional Hilbert spaces associated with each site, and where the Hamiltonian of the system is a sum of local terms. We are interested in proving uniform bounds on various properties as the size of the lattice tends to infinity. An important case is when there is a spectral gap between the lowest state(s) and the rest of the spectrum which persists in this limit, corresponding to what physicists call a "phase of matter." Here, the combination of elementary Fourier analysis with the technique of Lieb–Robinson bounds (bounds on the velocity of propagation) is surprisingly powerful. We use this to prove exponential decay of connected correlation functions, a higher-dimensional Lieb–Schultz–Mattis theorem, and a Hall conductance quantization theorem for interacting electrons with disorder.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 81V70; Secondary 82B10, 82C10

KEYWORDS

Lieb–Robinson bounds, quantum Hall effect

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4074–4095
DOI 10.4171/ICM2022/54

1. INTRODUCTION

This paper considers lattice quantum systems.[1] It is worth having some concrete examples in mind. For one specific example, consider a Hilbert space $(\mathbb{C}^2)^{\otimes L}$, i.e., the tensor product of L Hilbert spaces, each of dimension 2. Index these two-dimensional Hilbert spaces (called "spins" or "sites") with an integer i, taken periodic modulo L, and consider the Hamiltonian

$$H = J_1 \sum_i \vec{S}_i \cdot \vec{S}_{i+1} + J_2 \sum_i \vec{S}_i \cdot \vec{S}_{i+2}, \tag{1.1}$$

where $\vec{S}_i = (S_i^x, S_i^y, S_i^z)$ denotes the spin operators on the ith such Hilbert space, i.e.,

$$S^x = \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}, \quad S^y = \begin{pmatrix} 0 & \sqrt{-1}/2 \\ -\sqrt{-1}/2 & 0 \end{pmatrix}, \quad S^z = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}.$$

For $J_2 = (1/2)J_1 > 0$, the lowest eigenvalue of H is doubly degenerate for even L. A basis of ground states consists of either pairing sites $2i, 2i+1$ in a singlet for all i or pairing sites $2i, 2i-1$ in a singlet; this is called the Majumdar–Ghosh chain [16]. A slight perturbation of the Hamiltonian, taking J_2/J_1 slightly different from $1/2$, breaks the degeneracy of the lowest eigenvalue, but there is an exponentially small difference between the two lowest eigenvalues followed by a gap to the rest of the spectrum that remains nonvanishing as $L \to \infty$.

This system exemplifies some of the results that we will consider. The fact that for all J_1, J_2 the lowest eigenvalue is either degenerate or has vanishing (in the large L limit) difference to the next lowest eigenvalue is a corollary of the classic Lieb–Schultz–Mattis theorem [15]. That theorem is applicable only to one-dimensional quantum systems with periodic boundaries, meaning that sites can be arranged on a circle with short-range interactions. We will explain a more general machinery that allows us to prove a similar theorem for higher-dimensional quantum systems such as those on a two-dimensional square lattice [8].

Also, as J_1, J_2 vary, the properties of the Hamiltonian in (1.1) change, but for J_2 close to $J_1/2$, the connection correlation functions are exponentially decaying in the distance between the operators. Here a connected correlation function is $\langle AB \rangle - \langle APB \rangle$ where A, B are operators supported on some set of sites, P projects onto the two lowest (approximately) degenerate eigenvalues, and $\langle \ldots \rangle$ denotes the expectation value in an eigenvector corresponding to such an eigenvalue. This decay follows from another theorem that we will discuss, on exponential decay of correlation functions, again valid in any dimension under some assumptions on the Hamiltonian.

The Hamiltonian of (1.1) obeys a symmetry, namely the Hamiltonian commutes with the three operators $\sum_i S_i^x$, $\sum_i S_i^y$, and $\sum_i S_i^z$. Indeed, we will consider often Hamiltonians which just commute with a single operator, such as $\sum_i (S_i^z + 1/2)$; here we add a

[1] For simplicity, throughout we consider systems where the Hilbert space has a tensor product structure. It is straightforward to extend these results to the case where fermions obeying canonical anticommutation relations are present; roughly, this is done by replacing certain commutators with anticommutators as needed. We omit this for simplicity in this presentation.

factor $1/2$ so that the eigenvalues of $S_i^z + 1/2$ are integers and we can interpret this operator as a "conserved charge." Studying such Hamiltonians further in two-dimensions leads to the question of quantum Hall conductance, also discussed here.

Surprisingly, the key to many of these results is to consider the dynamical properties of the system, where we consider correlation functions of operators at different times, using the technique of Lieb–Robinson bounds. Let us begin by defining the systems we consider in some generality.

1.1. Lattice quantum systems

We consider quantum systems defined on a finite lattice. We have some finite set Λ of *sites*. The set Λ is called the lattice. Associated with each site i is some finite-dimensional Hilbert space, and the Hilbert space of the whole quantum system is the tensor product of these Hilbert spaces. There is some metric $\mathrm{dist}(i, j)$ where $i, j \in \Lambda$. In many applications, Λ may indeed be a geometric lattice in $\mathbb{R}^n$ or in $\mathbb{T}^n$ for some n; in this case, we call the system n-dimensional, and in this case the metric is inherited from the ambient space $\mathbb{R}^n$ or $\mathbb{T}^n$. However, in general Λ may be an arbitrary set with arbitrary metric.

The Hamiltonian H has the form

$$H = \sum_X h_X, \tag{1.2}$$

where the sum ranges over $X \subset \Lambda$, and each h_X is self-adjoint and is supported on set $X \subset \Lambda$.

We use $\|O\|$ to denote the operator norm (largest singular value) of O. Our interest is in local Hamiltonians; locality is expressed as some assumption on the norms $\|h_X\|$ as a function of the diameter of the sets X.

One typical assumption is that the Hamiltonian has *bounded strength and range* and that the set of sites Λ has *bounded local geometry*, meaning that all $\|h_X\| \leq J$ for some *strength* J and all sets X has $\mathrm{diam}(X) \leq R$ for some *range* R and that for all $i \in \Lambda$, we have $|\{j \in \Lambda \mid \mathrm{dist}(i, j) \leq R\}|$ bounded by some constant. Other assumptions considered include exponential decay where $\|h_X\|$ is exponentially small in $\mathrm{diam}(X)$.

From certain such locality assumptions, one can derive a so-called Lieb–Robinson bound, which can be thought of as bounding the velocity of excitations in such a lattice quantum system. For an arbitrary operator A, let

$$A(t) \equiv \exp(iHt) A \exp(-iHt) \tag{1.3}$$

denote the operator A evolved for time t under Hamiltonian H.

The first such bound was proven by Lieb and Robinson [14]. However, their proof gave bounds that depended on the dimension of the Hilbert space. In Appendix A we give a different proof that does not depend on the dimension, following the strategy in [8] as slightly modified in [9]. Indeed, rather than proving a specific bound, we give a series expansion (A.1) below which upper bounds $\|[A(t), B]\|$ and different assumptions on $\|h_X\|$ can be inserted into this series. Using this series expansion, a typical result [9] is

Lemma 1. *Suppose there are positive constants μ, s such that for all sites i we have*

$$\sum_{X \ni i} \|h_X\| |X| \exp\bigl[\mu \operatorname{diam}(X)\bigr] \le s < \infty.$$

Then, for any sets X, Y with $\operatorname{dist}(X, Y) > 0$, and any operators A, B supported on X, Y, respectively,

$$\bigl\|\bigl[A(t), B\bigr]\bigr\| \le 2\|A\|\|B\| \sum_{i \in X} \exp\bigl[-\mu \operatorname{dist}(i, Y)\bigr]\bigl[e^{2s|t|} - 1\bigr].$$

As a corollary, defining $v_{LR} = 4s/\mu$, for $\operatorname{dist}(X, Y) \ge v_{LR} t$, we have

$$\bigl\|\bigl[A(t), B\bigr]\bigr\| \le |X| \cdot \|A\|\|B\| \exp\biggl[-\frac{\mu}{2} \operatorname{dist}(i, Y)\biggr].$$

This quantity v_{LR}, called the Lieb–Robinson velocity, can be thought of as defining a "light-cone" [6], so that, up to exponentially small error, $A(t)$ is supported within distance $v_{LR} t$ of X.

Note that Lemma 1 is applicable to the case of bounded strength and range with bounded local geometry.

Remark 1. The fact that the commutator is not vanishing, but merely very small, outside the light-cone is sometimes called "leakage." In most applications of these bounds, the leakage is negligibly small compared to the other terms. Indeed, for the rest of this paper, the leakage terms will be negligible and we will avoid any detailed discussion of them. However, we emphasize that the leakage really is nonzero in every case of interest; since the commutator is an analytic function of time (this follows trivially since we have a finite-size system), if the commutator were exactly zero on some interval of time, then it would vanish for all times.

We emphasize that we consider finite-size systems, so that many properties can be defined in an elementary way. For example, the Hamiltonian is a finite-dimensional matrix; the ground state energy is simply the smallest eigenvalue of the Hamiltonian; if the smallest eigenvalue is nondegenerate, then the ground state is simply the corresponding eigenvector (up to some arbitrary phase); and correlation functions are simply the trace of the projector onto the ground state with some given other finite-dimensional matrices. This contrasts with considering systems directly in the infinite size limit where one must take some care to define an algebra of operators on an infinite system. However, although we consider finite-size systems, our interest is in bounds that are uniform in $|\Lambda|$.

1.2. Outline of results and notation

We will survey some of the results that have been obtained using these methods. A key role is played by the *spectral gap*. In this paper, unless stated otherwise, the spectral gap is defined to be the absolute value of the difference between the ground state energy of H (assumed nondegenerate) and the next smallest eigenvalue. We denote the spectral gap ΔE.

We will be loose about estimates. In many cases we will simply state that a term is small (perhaps exponentially small or some other decay), leaving the detailed proofs and precise bounds to the already-published literature. This is done to emphasize the ideas without getting too involved in the estimates. At the same time, we will give examples from physics to motivate the constructions.

We use computer-science big-O notation such as $\Omega, \mathcal{O}, \ldots$ throughout, where we implicitly consider a family of Hamiltonians defined on a family of Λ with increasing cardinality Λ. The control parameter for the big-O notation may be $|\Lambda|$ or, in the higher dimensional Lieb–Schultz–Mattis theorem later, may be some other distance scale.

We use $|\ldots|$ for the ℓ_2-norm of a state vector. We use I for the identity matrix. We use $A^\dagger$ to denote the Hermitian conjugate of an operator A.

When we refer to a quantum state, this will always be a normalized pure state (with an arbitrary phase), rather than a mixed state.

In Section 2, we sketch the proof that connected correlation functions decay exponentially in systems with a spectral gap [8,9]. This can be understood as a nonrelativistic analogue of a familiar result in relativistic quantum field theory (where the speed of light plays the role of v_{LR}) that correlation functions decay exponentially in gapped theories. In Section 3, we sketch the proof of the higher-dimensional Lieb–Schultz–Matthis theorem, proven in [8]. In Section 4, we sketch the proof of Hall conductance quantization [10]. These last two results have a certain topological flavor. In both cases, one of the physical ideas motivating the mathematical proof is that although correlation functions decay exponentially in gapped systems, it is still possible for there to be some kind of topological order in the ground state.

2. EXPONENTIAL DECAY OF CONNECTED CORRELATION FUNCTIONS

In massive relativistic quantum field theories, connected correlation functions decay exponentially. Here we consider similar results for lattice field theories. The Lieb–Robinson velocity plays some role similar to that of the speed of light in a relativistic field theory, while the spectral gap plays a role similar to that of a mass gap.

A typical result for exponential decay is

Theorem 1. *Let A_X, B_Y be supported on sets X, Y. Suppose the conditions of Lemma 1 hold. Assume there is a unique ground state with spectral gap ΔE to the rest of the spectrum. Let $\langle\ldots\rangle$ denote the expectation value in the ground state. Then,*

$$\langle A_X B_Y\rangle - \langle A_X\rangle\langle B_Y\rangle \leq \|A_X\| \cdot \|B_Y\| \left(\exp\left[-\Omega\left(\frac{\operatorname{dist}(X,Y)\Delta E}{v_{LR}}\right)\right] + \cdots\right),$$

where "…" denotes a leakage term from the Lieb–Robinson bound which is bounded by $|X| \cdot \|A_X\|\|B_Y\|$ times an exponentially decaying function of $\operatorname{dist}(X,Y)$.

Remark 2. This result can be readily generalized to the case that H has a q-fold degenerate (or almost degenerate) smallest eigenvalue and then a gap ΔE to the rest of the spectrum. Then, defining P_0 to project onto the ground state subspace and $\langle O\rangle \equiv \frac{1}{q}\operatorname{tr}(P_0 O)$, one may derive a more general bound on $\langle A_X B_Y\rangle - \langle A_X P_0 B_Y\rangle$. In this case, there is an additional term in the bound that vanishes in the limit that the q lowest eigenvalues become exactly degenerate.

Remark 3. The proof follows a general outline that we will use to derive the later results also, and, after giving the proof, we will emphasize the ideas that will be repeated later.

Proof of Theorem 1. To ease notation, replace A_X by $A_X - \langle A_X \rangle$ and B_Y by $B_Y - \langle B_Y \rangle$, so that both A_X and B_Y have vanishing expectation value in the ground state. Also, let $\ell = \operatorname{dist}(X, Y)$.

Let Ψ_n for $n = 0, 1, \ldots$ be an orthonormal basis of eigenstates of H, with eigenvalues $E_0 < E_1 \leq \cdots$. Let $\langle \psi, \phi \rangle$ denote the inner product between states ψ and ϕ. Then,

$$\begin{aligned} \big\langle \big[A_X(t), B_Y\big]\big\rangle = \sum_{n>0} \langle \psi_0, A_X \Psi_n \rangle \langle \Psi_n, B_Y \psi_0 \rangle \exp\big(-i(E_n - E_0)t\big) \\ - \sum_{n>0} \langle \psi_0, B_Y \psi_n \rangle \langle \psi_n, A_X \psi_0 \rangle \exp\big(+i(E_n - E_0)t\big). \end{aligned} \tag{2.1}$$

Thus, to compute the desired correlation function $\langle A_X(t=0) B_Y \rangle$, we want to extract the "negative frequency" part of $\langle [A_X(t), B_Y] \rangle$, meaning the first sum in (2.1), evaluated at $t = 0$. To do this, we use the following lemma [8, 9]. It shows a typical kind of technique in this subject: we have some function (in this case a step function) which has a singularity, and we construct some other function (or in this case, a limit of a family of functions) which is a good approximation to that function when the argument has absolute value $\geq \Delta E$, and we show that that approximation has a fast decaying Fourier transform.

Lemma 2. *Let $E \in \mathbb{R}$ and $\alpha > 0$. Then*

$$\begin{aligned} \lim_{T\uparrow\infty} \lim_{\varepsilon\downarrow 0} \frac{i}{2\pi} \int_{-T}^{T} \frac{e^{-iEt} e^{-\alpha t^2}}{t + i\varepsilon} \mathrm{d}t &= \frac{1}{2\pi} \sqrt{\frac{\pi}{\alpha}} \int_{-\infty}^{0} \mathrm{d}\omega \exp\big[-(\omega + E)^2/(4\alpha)\big] \\ &= \begin{cases} 1 + \mathcal{O}(\exp[-\Delta E^2/(4\alpha)]) & \textit{for } E \geq \Delta E, \\ \mathcal{O}(\exp[-\Delta E^2/(4\alpha)]) & \textit{for } E \leq -\Delta E. \end{cases} \end{aligned}$$

Using Lemma 2, one has

$$\langle \Phi, A_X B_Y \Phi \rangle = \lim_{T\uparrow\infty} \lim_{\varepsilon\downarrow 0} \frac{i}{2\pi} \int_{-T}^{T} \mathrm{d}t \frac{1}{t + i\varepsilon} \big\langle \big[A_X(t), B_Y\big] \big\rangle e^{-\alpha t^2} + \mathcal{O}\big(\exp\big[-\Delta E^2/(4\alpha)\big]\big). \tag{2.2}$$

Now we choose α and apply the Lieb–Robinson bound. Fix $\alpha = \Delta E v_{LR}/(2\ell)$. Then, $\mathcal{O}(\exp[-\Delta E^2/(4\alpha)]) = \mathcal{O}(\exp[-\ell \Delta E/(2 v_{LR})])$. This bounds the second term on the right-hand side of (2.2). To bound the first term, we break the integral over t into an integral for $|t| \leq \ell/v_{LR}$ and an integral for $|t| \geq \ell/v_{LR}$. The integral for $|t| \leq \ell/v_{LR}$ can be bounded using the Lieb–Robinson bound, giving the leakage term in the theorem. The integral for $|t| \geq \ell/v_{LR}$ is bounded by a triangle inequality by

$$2\|A_X\| \cdot \|B_Y\| \lim_{T\uparrow\infty} \lim_{\varepsilon\downarrow 0} \frac{1}{2\pi} \int_{\ell/v_{LR} \leq |t| \leq T} \mathrm{d}t \frac{1}{t + i\varepsilon} e^{-\alpha t^2},$$

which is bounded by $2\|A_X\| \cdot \|B_Y\| \cdot \mathcal{O}(\exp[-\ell \Delta E/(2 v_{LR})])$. ■

3. HIGHER-DIMENSIONAL LIEB–SCHULTZ–MATTIS

3.1. Review of one-dimensional Lieb–Schultz–Mattis theorem

One-dimensional quantum spin systems with $SU(2)$-invariant Hamiltonians exhibit different behavior depending on whether the spin is integer or half-integer. For a half-integer spin, it is found that either the ground state is degenerate or the gap vanishes in the thermodynamic limit, while for an integer spin there may be a unique ground state with a gap.

One paradigmatic example is the spin-S Heisenberg spin chain,

$$H = J \sum_i \vec{S}_i \cdot \vec{S}_{i+1},$$

where $J > 0$ and the sites are labeled by integers $i = 0, 1, \ldots, L-1$, which are periodic modulo L, and corresponding to each site there is a $(2S+1)$-dimensional Hilbert space corresponding to a spin-S representation of $SU(2)$, and where $\vec{S}_i$ is a vector of spin operators on the ith spin. For $J > 0$ and spin-$1/2$, there is a continuous spectrum of excitations. The spectral gap then vanishes polynomially in L. Another paradigmatic example is the Majumdar–Ghosh chain mentioned at the start of this paper. On the other hand, for integer spin the famous "Haldane conjecture" [7] asserts that there is a unique ground state with a spectral gap which is $\Omega(1)$, and the AKLT Hamiltonian [2] is an exactly solvable spin-1 Hamiltonian which shows this property.

This vanishing of the gap for half-integer spins is a corollary of the Lieb–Schultz–Mattis theorem [3, 15] in one-dimension. We will give the theorem in a more general setting where the Hamiltonian is $U(1)$-symmetric, without using the full $SU(2)$ symmetry, and then relate this to the case of spin chains.

Let us define some terms. Say that a system is one-dimensional with periodic boundary conditions of size L, for some integer L, if the sites in Λ correspond to vertices of a cycle graph with $|\Lambda| = L$, with the shortest path metric on the graph being the distance, and if all sites have the same Hilbert space dimension. We will label sites by integers. Define a translation operator T in the obvious way; T is a unitary operator, and conjugation by T maps the algebra of operators supported on site i to those supported on site $i+1$, and $T^L = I$. We say that a Hamiltonian is translationally invariant if $THT^{-1} = H$. We say that a system has a conserved charge Q if $Q = \sum_i q_i$, with q_i being an operator with integer eigenvalues supported on site i with $\|q_i\| \leq q_{\max}$ for some $q_{\max}$, so that $q_j = T^{j-i} q_i T^{i-j}$ and

$$[Q, H] = 0.$$

The proof that follows uses a trick of averaging over two choices of twist, to use the minimum number of assumptions; this form of the proof seems to have first appeared in [12].

Theorem 2. *Consider a one-dimensional system with periodic boundary conditions of size L with a translation invariant Hamiltonian H and a conserved charge Q. Further, assume H has strength J and range R.*

Let ψ_0 be a ground state of H with $|\psi_0| = 1$, and suppose that $\frac{\langle \psi_0, Q\psi_0 \rangle}{L}$ is not integer. Then, if the ground state is nondegenerate, the spectral gap of H is bounded by $\mathcal{O}(\frac{J q_{\max}^2 R^2}{L})$.

Proof. The proof is variational: construct another state, show that the state is orthogonal to ψ_0, and compute the expectation value of H in this state. Let

$$U_{LSM} = \exp(\pm iA),$$

where we will pick the sign later and where

$$A \equiv \sum_{j=0}^{L-1} 2\pi q_j \frac{j}{L}.$$

The variational state used is

$$\Phi = U_{LSM}\psi_0.$$

Since the ground state of H is nondegenerate and since $[T, H] = 0$, we have $T\psi_0 = z\psi_0$ for some scalar z with $|z| = 1$. Note that

$$\begin{aligned} T\Phi &= T\exp\left[\pm i \sum_{j=0}^{L-1} 2\pi q_j \frac{j}{L}\right]\psi_0 \\ &= \exp\left[\pm i \sum_{j=0}^{L-1} 2\pi \big(T q_j T^{-1}\big)\frac{j}{L}\right] T\psi_0 \\ &= z\exp\left[\pm i \sum_{j=0}^{L-1} 2\pi \big(T q_j T^{-1}\big)\frac{j}{L}\right]\psi_0 \\ &= z\exp\left[\pm i \sum_{j=0}^{L-1} 2\pi q_{j+1}\frac{j}{L}\right]\psi_0 \\ &= z\exp\left[\pm i \sum_{j=0}^{L-1} 2\pi q_j \frac{j-1}{L}\right]\psi_0 \\ &= z\exp\left[\mp i 2\pi \frac{Q}{L}\right]\Phi \\ &= z\exp\left[\mp i 2\pi \frac{\langle\psi_0, Q\psi_0\rangle}{L}\right]\Phi. \end{aligned} \tag{3.1}$$

The equality on the fifth line is a change in the index of summation, replacing j by $j-1$. Here the assumption that q_j has integer eigenvalues is used so that $\exp[i2\pi q_L \frac{L}{L}] = \exp[i2\pi q_0 \frac{0}{L}] = I$. The equality on the final line uses the assumption that $[Q, H] = 0$ so that ψ_0 is an eigenvector of Q.

Using the assumption that $\langle\psi_0, Q\psi_0\rangle/L$ is noninteger, it follows that Φ is an eigenvector of T with eigenvalue different from z, so it is orthogonal to ψ_0.

Now we estimate the energy of this state. Write $H = \sum_{i=0}^{L-1} h_i$, with $h_j = T^{j-i} h_i T^{i-j}$ and with each h_i supported on the set of sites within distance R of i, with $\|h_i\| \leq J$.

We average $\langle \Phi, H\Phi\rangle - \langle \psi, H\psi\rangle$ over the two choices of sign in U_{LSM}, giving

$$\begin{aligned}
&\Big\langle \Psi, \Big(\frac{U_{LSM}^\dagger H U_{LSM} + U_{LSM} H U_{LSM}^\dagger}{2} - H\Big)\Phi\Big\rangle \\
&\quad = L\Big\langle \Psi, \Big(\frac{U_{LSM}^\dagger h_0 U_{LSM} + U_{LSM} h_0 U_{LSM}^\dagger}{2} - h_0\Big)\Phi\Big\rangle \\
&\quad \le L\Big\| \frac{U_{LSM}^\dagger h_0 U_{LSM} + U_{LSM} h_0 U_{LSM}^\dagger}{2} - h_0 \Big\| \\
&\quad \le L\big\|[[h_0, A], A]\big\|,
\end{aligned}$$

where the averaging over signs cancels terms $[h_0, A]$. Finally, $\|[[h_0, A], A]\| = \mathcal{O}(\frac{Jq_{\max}^2 R^2}{L^2})$. ■

To apply this system to $SU(2)$-invariant spin chains with half-integer spin, we may take $q_i = 1/2 + S_i^z$, where S_i^z is the z-component of the ith spin. Then, if the ground state is nondegenerate, it has total spin 0, and hence $\langle \sum_i S_i^z\rangle = 0$, so $\langle Q\rangle/L = 1/2$, noninteger.

It is instructive to consider this variational state in the case of the Majumdar–Ghosh chain (1.1). A basis of ground states corresponds to pairing neighboring sites in singlets in one of two ways. Taking the sum of these states gives an eigenvector of T with eigenvalue $+1$. Applying U_{LSM} to this sum gives the difference of these two states, up to an error of order $\mathcal{O}(1/L)$. The difference of these states is an eigenvector T with eigenvalue -1.

3.2. Higher-dimensional extensions: physics

One might try to extend this theorem beyond one-dimensional systems. Note first that translation invariance is necessary for the theorem to hold: one can easily construct spin-$1/2$ systems without translation invariance with a unique ground state and a gap such as

$$H = \sum_i \vec{S}_{2i} \cdot \vec{S}_{2i+1}.$$

The higher-dimensional theorem will apply to n-dimensional quantum systems, for $n \ge 1$. However, we will be able to state the theorem in greater generality, which will also be convenient because it will emphasize the fact that we use translation invariance in only one direction.

We say that a system has *translation invariance in one direction with periodicity* L if the sites can be labeled by a pair (i, v), where i is an integer labeling a vertex of a cycle graph of length L and v is a vertex in some other graph G, so that the following hold. First, the metric is the shortest path metric on the graph given by the Cartesian product of that cycle graph with G. Second, the Hilbert space dimension of site (i, v) is some d_v depending only on v. Given this, define a unitary operator T such that conjugation by T maps the algebra of operators supported on site (i, v) to those supported on site $(i+1, v)$ for all v, and $T^L = I$. We say that a Hamiltonian is translationally invariant if $THT^{-1} = H$. We say that a system has a conserved charge Q if $Q = \sum_{i,v} q_{i,v}$, with $q_{i,v}$ being an operator with integer eigenvalues supported on site (i, v) with $\|q_{i,v}\| \le q_{\max}$, so that $q_{j,v} = T^{j-i} q_{i,v} T^{i-j}$ and

$$[Q, H] = 0.$$

For an n-dimensional quantum system, G may be an $(n-1)$-fold Cartesian product of cycle graphs, so that sites are labeled by n different integers, each periodic modulo some other integer.

We then have [8]

Theorem 3. *Consider a system with translation invariance in one direction with periodicity L, with a Hamiltonian with strength and range J and R, respectively, both $\mathcal{O}(1)$, and such that Λ has bounded local geometry. Assume that the number of sites is $\mathcal{O}(\mathrm{poly}(L))$. Assume that there is a conserved charge with $q_{\max} = \mathcal{O}(1)$. Assume that the ground state ψ_0 is unique with $\langle \psi_0, Q\psi_0\rangle/L$ noninteger. Then, the gap ΔE is $\mathcal{O}(\log(L)/L)$, where the constants hidden in the big-O notation depend on J, R, $q_{\max}$, on the polynomial bounding the number of sites, and on the local geometry of G.*

Note that the theorem is slightly weaker than in the one-dimensional case, with a bound $\mathcal{O}(\log(L)/L)$ rather than $\mathcal{O}(1/L)$. Also, for simplicity, we have been less explicit about the dependence of the bound on the constants.

Note also one slightly unsatisfactory feature: the theorem requires that $\langle \psi_0, Q\psi_0\rangle/L$ be noninteger. Suppose that we consider a two-dimensional system, of size L-by-L', with $\langle \psi_0, Q\psi_0\rangle/(LL') = 1/2$. This is a typical case of interest in spin systems. Then, the theorem is only applicable if L' is odd.

One might attempt to use the same proof as before. Suppose that H has bounded strength and range. Applying the same variational argument, the change in the expectation value of every term in the Hamiltonian (e.g., $\langle \Phi, h_X \Phi\rangle - \langle \psi_0, h_X \psi_0\rangle$) is still $\mathcal{O}(1/L^2)$, but the number of such terms now is not L but rather proportional to L times the number of vertices in G. As a typical application of interest, let $L_x = L$ and let G be a cycle graph of length L_y so that vertices are labeled by a pair of integers (i, j), with i periodic modulo L_x and j periodic modulo L_y. Then if the "aspect ratio" L_y/L_x is of order unity, the variational state may have energy of order unity above the ground state [1].

The problems with this approach were given in a very insightful physics article [17]. Indeed, the problem is not purely mathematical. The problem is that a two-dimensional quantum spin system can exhibit completely different behavior from a one-dimensional quantum spin system. A one-dimensional system with spin-$1/2$ can have a state like the Heisenberg chain with a polynomially small gap and power-law decaying correlations. Alternatively, it can have a state like the Majumdar–Ghosh chain. In this exactly solvable example, two choices of ground state correspond to two different ways of pairing neighbors into singlets, either pairing site $2i$ with $2i+1$ or pairing site $2i$ with $2i-1$. These two choices each break translation symmetry, though one may take symmetric and antisymmetric combinations to obtain ground states which are eigenstates of T. There is a local order parameter which distinguishes between these states,[2] i.e., indeed, there is an operator supported on a set of bounded diameter which has nonvanishing matrix elements between symmetric and

[2] See [3] for results showing that, in a sense, these are the only two possibilities for a one-dimensional system, either translational symmetry breaking or a continuous spectrum.

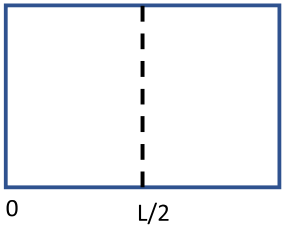


FIGURE 1
Schematic illustration of a two-dimensional system. Individual sites are not shown. Left- and right-hand sides of the system are identified to give a cylinder (or torus if top and bottom are also identified). Numbers indicate the first coordinate. The operator Q_{left} is the total charge on sites between the left-hand side and the dashed line. Twist θ twists terms near the left-hand side of the figure, while θ' twists terms near the dashed line. The twist θ' is used as a technical trick in the proof of the higher-dimensional Lieb–Schultz–Mattis theorem.

antisymmetric states. However, in two dimensions, there is a new possibility. Spins can pair into singlets (or dimers as they are called) but these dimers can enter into a quantum spin liquid state [20–22]. In this case, the physics is that there is still a "topological degeneracy," so that there is an exponentially small splitting between the two lowest eigenstates. However, there is no local order parameter: any operator supported on a set of bounded diameter is exponentially close to a scalar in the subspace of the two lowest eigenstates.

So, while these physics arguments provide some motivation to find a generalization of the Lieb–Schultz–Mattis theorem to higher-dimensional systems, they also show that the variational argument does not directly generalize. To prove the theorem, we need the tool of "quasiadiabatic continuation," described in the next subsection.

This tool is used to turn physics arguments based on an idea of twisting boundary conditions into precise results.

Assume H has finite range R, with $L \gg R$.

Given any site (i, v), its *first coordinate* is the integer i. Let

$$Q_{\text{left}} = \sum_{0 \le i < L/2} \sum_{v} q_{i,v}$$

be the total charge on sites with the first coordinate $0 \le i < L/2$. This is the "left" half of the system in Figure 1.

We say that a set has the first coordinate *near* i if it is within distance R of the set of sites with the first coordinate i, treating the first coordinate periodic modulo L. Define a two-parameter family of Hamiltonians (note the signs in the exponents)

$$H_{\theta,\theta'} \equiv \sum_{X \text{ near } 0} \exp(i\theta Q_{\text{left}}) h_X \exp(-i\theta Q_{\text{left}}) + \sum_{X \text{ near } L/2} \exp\big(-i\theta' Q_{\text{left}}\big) h_X \exp\big(i\theta' Q_{\text{left}}\big)$$
$$+ \sum_{\text{remaining } X} h_X,$$

where θ, θ' are real parameters, periodic modulo 2π. The last sum is over X not near 0 or $L/2$. Note that

$$H = H_{0,0}$$

and

$$H_{\theta+\phi,\theta'-\phi} = \exp(iQ_{\text{left}}\phi) H_{\theta,\theta'} \exp(-iQ_{\text{left}}\phi). \tag{3.2}$$

Physically, we can regard θ as defining a "gauge field" along the line of sites with the first coordinate 0, and θ' as another gauge field along the line $L/2$. Then (3.2) describes a gauge transformation relating two different "choices of gauge," with the sum $\theta + \theta'$ invariant under this. While this relation may seem trivial, it will be very useful in what follows.

Oshikawa [19] considered the effect of changing θ from 0 to 2π in an attempt to prove a higher-dimensional Lieb–Schultz–Mattis theorem. In [17], his argument was analyzed in more detail, where it was found that what it proves is that the ground state must become degenerate at some value of θ if $\langle \psi_0, Q\psi_0 \rangle / L$ is noninteger. Proof: suppose the ground state is nondegenerate for all θ. Then, the adiabatic evolution from $\theta = 0$ to 2π maps the ground state to itself. However, one may show that if the ground state is an eigenvector of the translation operator T with eigenvalue z, the adiabatically evolved state has eigenvalue $\exp(i\frac{2\pi}{L}\langle \psi_0, Q\psi_0 \rangle) z \neq z$.

This kind of argument considering a change in boundary conditions is called a flux insertion, and is related to Laughlin's argument for Hall conductance quantization [13]. While the adiabatic evolution using a flux insertion does not prove the desired result, this is still a useful physical idea. The technical tool we use to make the idea of flux insertion rigorous is called "quasiadiabatic continuation" [8].

3.3. Quasiadiabatic continuation

Suppose we have some family of Hamiltonians H_s which depend smoothly on a real parameter s. Assume that for all s the ground state is nondegenerate and the spectral gap is $\geq \Delta E$. Let $\psi_a(s)$ for $a \geq 0$ denote the (orthonormal) eigenstates of H_s, with energies $E_a(s)$, with $\psi_0(s)$ being the ground state.

Then, a familiar result of the first-order perturbation theory is that

$$\partial_s \psi_0(s) = \sum_{a>0} \frac{1}{E_0(s) - E_a(s)} \psi_a(s) \big\langle \psi_a(s), (\partial_s H_s) \psi_0(s) \big\rangle. \tag{3.3}$$

We now use a trick similar to that used in the proof of exponential decay of correlations above: we take some function of energy difference which has some singularity, i.e., in this case, $1/(E_0 - E_a)$, and we approximate that function by a smooth function which gives a good approximation when the energy difference is $\geq \Delta E$. In the case of exponential decay of correlations, the needed result is Lemma 2. Here a variety of different forms have been used, and, rather than giving details, we simply give the approach outline.

Let $f(\cdot)$ be some smooth function with some Fourier transform $\tilde{f}(\cdot)$. Assume $f(x) \approx -1/x$ for $|x| \geq \Delta$ and $f(0) = 0$ and $f(x) = -f(-x)$, where we will not be precise about the meaning of the approximation $\approx$. Then,

$$\begin{aligned}\partial_s \psi_0(s) &\approx \sum_{a>0} f(E_a - E_0)\psi_a(s)\big\langle \psi_a(s), (\partial_s H_s)\psi_0(s)\big\rangle \\ &= \sum_a \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{2\pi} \tilde{f}(t) \exp\big(i(E_a - E_0)t\big)\psi_a(s)\big\langle \psi_a(s), (\partial_s H_s)\psi_0(s)\big\rangle \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}t}{2\pi} \tilde{f}(t)\big(\exp\big(+iH_s(t)\big)(\partial_s H_s)\exp\big(-iH_s(t)\big)\big)\psi_0(s) \\ &\equiv i\mathcal{D}_s\psi_0(s), \end{aligned} \tag{3.4}$$

where the last line of the equation is interpreted as defining an operator $i\mathcal{D}_s$ called the quasiadiabatic continuation operator. The error in this approximation in (3.4) depends on the error in the approximation $f(x) \approx -1/x$ and on the norm $\|\partial_s H_s\|$, and we do not go into details here.

Since we have chosen f to be odd, $\mathcal{D}_s$ is Hermitian. We can integrate this quasiadiabatic continuation operator along a path such as $s \in [0, 1]$ to give

$$\psi_1(s) \approx \mathcal{P}\exp\left(i\int_0^1 \mathcal{D}_s \mathrm{d}s\right)\psi_0(s),$$

where $\mathcal{P}$ denotes a path-ordered exponential and $\mathcal{P}\exp(i\int_0^1 \mathcal{D}_s \mathrm{d}s)$ is a unitary.

The essential point of (3.4) is that if we choose f so that $\tilde{f}$ is sufficiently rapidly decaying in time, then (by the Lieb–Robinson bounds) the operator $\mathcal{D}_s$ enjoys certain locality properties. In particular, if $\partial_s H_s$ is supported on some given set (for example, $\partial_\theta H_{\theta,0}$ is supported within $O(1)$ of the line 0), then $\mathcal{D}_s$ can be approximated by an operator supported within some distance ℓ that set, with the error in approximation depending on the choice of $\tilde{f}$, and decreasing as ℓ is increased. Further, if $\partial_s H_s$ is a sum of operators supported on given sets then, by linearity, $\mathcal{D}_s$ can be approximated by a sum of operators supported within some distances of those sets.

In the original application of quasiadiabatic continuation [8], there were two sources of error. One came from the approximation in (3.4) since $f(x)$ was not exactly equal to $-1/x$ for $x \geq \Delta$, while the second came from the approximate locality of $\mathcal{D}_s$.

In [18], Osborne introduced a different "exact" version where $f(x)$ was exactly equal to $-1/x$ for $|x| \geq \Delta E$, and he showed that one could choose $\tilde{f}$ to decay superpolynomially in time. Using an old result in analysis [11], it is possible to improve this superpolynomial decay to an "almost exponential decay," made more precise later.

The original formulation of quasiadiabatic continuation gives tighter bounds for the higher-dimensional Lieb–Schultz–Mattis theorem. On the other hand, the "exact" quasiadiabatic continuation is more convenient for the proof of Hall conductance quantization. The exact form has the particular advantage that one may choose it so that evolution under the quasiadiabatic continuation operator also obeys a Lieb–Robinson bound.

We omit all the details of error estimates in this review.

3.4. Sketched proof of higher-dimensional Lieb–Schultz–Mattis theorem

We now sketch the proof of Theorem 3. The proof is variational, like the one-dimensional proof. It is also by contradiction. Let us assume that there is a gap ΔE, and for large enough ΔE we will derive a contradiction.

Let ψ_0 be the ground state of H. Since H is translation invariant, $T\psi_0 = z\psi_0$ for some z with $|z| = 1$.

Let U be an operator that implements a quasiadiabatic continuation of $H_{\theta,0}$ from $\theta = 0$ to 2π, with the quasiadiabatic parameter chosen appropriately (we do not go into the details) to make the following estimates work.

We emphasize that we do not make any assumption that $H_{\theta,0}$ has a gap for $\theta \neq 0$. Indeed, by the arguments above, we know that the gap closes at some θ. Nevertheless, we define U by integrating the quasiadiabatic evolution operator as if there were a gap ΔE.

Our variational state will be $\Phi = U\psi_0$. This is similar to the one-dimensional construction with the operator U_{LSM} replaced with a quasiadiabatic evolution. As in the one-dimensional proof, we prove two things: we bound $\langle \Phi, H\Phi\rangle - \langle \psi_0, H\psi_0\rangle$, and we compute $\langle \Phi, T\Phi\rangle$ to show that Φ is orthogonal to ψ_0.

To bound $\langle \Phi, H\Phi\rangle - \langle \psi_0, H\psi_0\rangle$, write

$$H = H_1 + H_2,$$

where H_1 is the sum of terms h_X such that X is closer to the set of sites with the first coordinate 0 than it is to the set of sites with the first coordinate $L/2$, and H_2 is the sum of the remaining terms:

$$H_1 \equiv \sum_{X \text{ closer to } \mathbf{0}} h_X$$

and

$$H_2 \equiv \sum_{\text{remaining } X} h_X.$$

The term H_2 commutes with U up to exponentially small (in $L\Delta E/v_{LR}$) error by locality of the quasiadiabatic evolution operator. At the same time, $\|H_2\|$ is only polynomially large in L, and so, for ΔE sufficiently large compared to $\log(L)/L$, the commutator of the second term with U is polynomially small (and indeed smaller than $\log(L)/L$).

To estimate $\langle \Phi, H_1\Phi\rangle - \langle \psi_0, H_1\psi_0\rangle$, define U' to implement quasiadiabatic evolution of $H_{0,\theta'}$ as θ' goes from 0 to -2π. Define W to implement the quasiadiabatic evolution of $H_{\theta,-\theta}$ as θ goes from 0 to 2π. Note the signs!

Since this is a sketched proof, we will write $\approx$ to indicate that something holds up to a polynomial in L times something exponentially small in $L\Delta E/v_{LR}$, so that, for ΔE sufficiently large compared to $\log(L)/L$, this $\approx$ indicates a polynomially small error. One may show the following:

$$\begin{aligned}\langle \Phi, H_1\Phi\rangle &\approx \langle U'\Phi, H_1 U'\Phi\rangle \\ &\approx \langle W\psi_0, H_1 W\psi_0\rangle \\ &\approx \langle \psi_0, H_1\psi_0\rangle, \end{aligned} \tag{3.5}$$

where the first line is by the same locality of quasiadiabatic evolution argument as we used in considering the commutator $[H_2, U]$. The second line of (3.5) is from $U'U \approx W$: this result can be understood as the quasiadiabatic evolution operator that generates W is a sum of two operators, one coming from the change in $H_{\theta,\theta'}$ with respect to θ and the other the change with respect to θ', while the operators that generate U and U', respectively, come from the change in $H_{\theta,\theta'}$ with respect to either θ or θ'. This simple argument does not fully justify the approximate equality, of course, as the evolutions are taken simultaneously in W (θ goes to 2π while θ' goes to -2π in W) and sequentially in $U'U$, but using locality one may show it is approximately true. The third line follows from (3.2): since $H_{\theta,-\theta}$ is unitarily equivalent to H, the Hamiltonian $H_{\theta,-\theta}$ must also have the same gap ΔE and so the quasiadiabatic evolution will approximately evolve the ground state of $H_{0,0}$ to the ground state $H_{2\pi,-2\pi} = H_{0,0}$, up to some phase. That is, while the gap will close for $H_{\theta,0}$, it remains open for $H_{\theta,-\theta'}$. At this point in the proof, the phase is not important, since it cancels, but in the next step a similar phase will be important.

This completes the sketch of the proof that $\langle\Phi, H\Phi\rangle - \langle\psi_0, H\psi_0\rangle$ is small. We now sketch the proof that Φ has small overlap with ψ_0. The ground state ψ_0 is an eigenvector of T with some eigenvalue z. We consider $z^{-1}\langle\Phi, T\Phi\rangle$ and bound it away from 1. We have

$$z^{-1}\langle\Phi, T\Phi\rangle = \langle U\psi_0, TUT^{-1}\psi_0\rangle \approx \langle U'U\psi_0, U'(TUT^{-1})\psi_0\rangle. \tag{3.6}$$

We have, as discussed above, $U'U\psi_0 \approx W\psi_0$, which is approximately equal to ψ_0 up to some phase. A similar result holds for $U'(TUT^{-1})$: this operator can be considered as describing the quasiadiabatic evolution in a family of Hamiltonians where instead the parameter θ describes a gauge field near the line with the first coordinate 1, rather than the first coordinate 0, while the parameter θ' still describes a gauge field near the line with the first coordinate $L/2$. So, $U'(TUT^{-1})\psi_0$ is also equal to ψ_0 up to a phase. However, analyzing this phase (which is approximately the geometric phase of some adiabatic evolution) shows that the two phases differ if $\langle\psi_0, Q\psi_0\rangle/L$ is noninteger, giving the desired result.

Remark 4. Note that if $\langle\psi_0, Q\psi_0\rangle/L$ is noninteger, then its difference from the nearest integer is $\Omega(1/L)$ since Q has integer eigenvalues. So, even if the difference from $\langle\psi_0, Q\psi_0\rangle/L$ to the nearest integer is $o(1)$, one may still bound the errors terms to show that the two phases differ.

4. HALL CONDUCTANCE QUANTIZATION

4.1. Introduction

In 1879, Edwin Hall performed an experiment in an attempt to determine the sign of the charge of charge carriers in a metal. Was current caused by negative charge carriers flowing in one direction or by positive charge carriers flowing in the opposite direction? Consider a sample of some metal, which looks like a rectangle as viewed from above. He ran a current from the left-hand side of the rectangle to the right-hand side, while applying a magnetic field into the plane. Maxwell's equations predicted that the charge carriers would experience a force, determined by their electric charge times the cross product of their velocity with the magnetic field. This force is in the plane of the rectangle, and perpendicular to the direction of current. The sign of this force is unchanged if one changes both the sign of the charge carriers and the sign of their velocity. This force is expected to lead to an accumulation of the charge carriers at the top or bottom edge of the rectangle, leading to a voltage between the top and bottom edge. This effect, that a magnetic field can lead to a voltage perpendicular to the current, is called the Hall effect.

The sign of the voltage for most materials agrees with what one would expect for charge carriers with a negative electric charge (i.e., electrons), though in some semiconductors, the sign is reversed and it is more natural to think of holes in the band as carrying the charge.

The Hall conductance has units of

$$\frac{\text{current}}{\text{voltage}} = \frac{\text{charge}}{\text{time}} \cdot \frac{\text{charge}}{\text{energy}} = \frac{\text{charge}^2}{\text{time} \cdot \text{energy}},$$

so that it has the same units as e^2/h where e is the charge of the electron and h is Planck's constant.

Surprisingly, in 1980, von Klitzing found experimentally that, in two-dimensional semiconductors at low temperatures and large magnetic field, the Hall conductance was quantized in integer multiples of e^2/h to very high accuracy.[3] The fundamental physical argument for this quantization was given by Laughlin [13] but a mathematical proof remained open.

One surprising feature is that this very accurate quantization persists even though the actual physical samples were disordered. In [5], noncommutative geometry techniques were used to prove Hall conductance quantization for free (i.e., noninteracting) electrons with disorder.

In [4], Avron and Seiler made another important advance, proving Hall conductance quantization under a certain averaging assumption. They considered a system on a torus and introduced two fluxes, θ and ϕ, on a longitude and meridian of the torus, respectively. See Figure 2. We write $H_{\theta,\phi}$ to denote a Hamiltonian as a function of these two parameters; both θ and ϕ are periodic modulo 2π. The space of parameters θ, ϕ is called the flux torus.

They assumed that $H_{\theta,\phi}$ has a unique ground state for all θ, ϕ. Consider adiabatically transporting the ground state around some infinitesimal loop in the flux torus near some given θ, ϕ. The ground state acquires some Berry phase. This Berry phase is related, by the Kubo formula, to the quantum Hall conductance at that θ, ϕ. Physically, we can understand this relation as follows. If we imagine changing one parameter (say, θ), this corresponds to a changing magnetic field which induces a voltage. This voltage induces a perpendicular current proportional to the Hall conductance, and this current will couple to the other parameter (in this case, ϕ), changing the phase of the wavefunction.

[3] There is also a fractional quantum Hall effect, where the Hall conductance is a rational multiple of e^2/h. This can occur if the ground state is (approximately) degenerate.

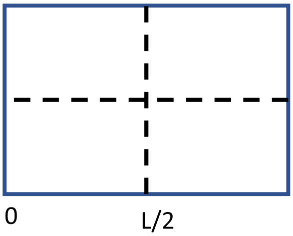


FIGURE 2
Schematic illustration of a two-dimensional system. Individual sites are not shown. Left- and right-hand sides of the figure are identified to give a cylinder, as well as top and bottom sides. Numbers indicate the first coordinate. Twists θ, ϕ twist terms near the left-hand side and top side of the figure, respectively, and are used in the averaging proof of [4]. Twists θ', ϕ' twist terms near the dashed lines. Twists θ', ϕ' are used as a technical trick in the proof of [10].

The problem then is reduced to computing the Berry phase around such an infinitesimal loop, i.e., computing a Berry curvature. The average of this curvature over the torus is quantized in integer multiples of $(2\pi)^{-1}$. Thus, Avron and Seiler proved that the average of the Hall conductance over the torus is quantized.

However, this left open the question of quantization of the Hall conductance at a specific values of θ, ϕ, assuming only a spectral gap at those θ, ϕ.

4.2. Results

We now state the results of [10], which proved Hall conductance quantization for interacting electrons without the averaging assumption. The results are quantitative, giving error bounds that decay almost exponentially fast as $L \to \infty$. A function f is called *almost-exponentially decaying*, if for all c with $0 \leq c < 1$ there is a constant C such that $f(x) \leq C \exp(-x^c)$ for all sufficiently large x, and a quantity is *almost-exponentially small* if it is bounded by an almost-exponentially decaying function.

We consider a two-dimensional quantum system, with sites on a torus T, with sites labeled by a pair (i, j) periodic modulo L for some L. We assume that there is a conserved charge $Q = \sum_v q_v$ as in the Lieb–Schultz–Mattis theorem and that the Hamiltonian has bounded strength and range. We assume that the Hamiltonian has a unique groundstate, with a spectral gap at least ΔE.

Specifically, one proves:

Theorem 4. *For any fixed, L-independent $R, J, q_{\max}$ and spectral gap $\Delta E > 0$, for any Hamiltonian satisfying the above assumptions, the difference between the Hall conductance σ_{xy} and the nearest integer multiple of e^2/h is almost-exponentially small in L, where e^2/h denotes the square of the electron charge divided by Planck's constant.*

The proof of this theorem takes several steps. First, one replaces the Berry connection used to compute the Berry phase with the quasiadiabatic evolution operator, to relate the

Hall conductance to the phase for quasiadiabatic evolution around a small loop. One considers a quasiadiabatic evolution around a small, but not infinitesimal loop, on the flux torus near $(\theta, \phi) = (0, 0)$. Keeping the loop sufficiently small (the size of the loop polynomially small in the system size), the gap remains open on evolution on this loop. Indeed, we may assume that the gap remains at least $(1 - o(1))\Delta E$, so that we may choose the function in the quasiadiabatic evolution so that it matches the adiabatic evolution on this loop. Thus, the ground state returns to itself after the quasiadiabatic evolution around this loop, up to a Berry phase. For a small loop size, this phase is proportional to the area times the Berry curvature plus higher-order corrections in loop size.

Second, one considers various other choices of $(\theta, \phi) \neq (0, 0)$. For each such choice, one defines a path from $(0, 0)$ to the given (θ, ϕ), around a small loop, and back to $(0, 0)$. Note that since we move now a large distance away from $(0, 0)$, the gap may become small, or even vanish. So, we do not have a guarantee that we return to the ground state at the end of the path. However, we can make energy estimates to show that we indeed return to the ground state. To do this, we use a similar trick to that done in the proof of the higher-dimensional Lieb–Schultz–Mattis theorem. In that proof, we introduced an additional twist θ' and used it to show that our variational state had energy close to the ground state. Here we introduce two extra twists θ', ϕ', and we use them to show that the state at the end of the path has energy close to the ground state. In this case, since we assume that the ground state is unique, this proves that we do return to the ground state up to small error. Further, we may show that the phase acquired is approximately independent of the choice of (θ, ϕ).

Next, one takes a product of the evolution over several such paths (each path going from $(0, 0)$ to some $(\theta, \phi) \neq (0, 0)$, around a loop, and back to $(0, 0)$). We do this such that the result is equivalent to evolution around a single large loop, i.e., certain segments of the paths cancel, leaving just the evolution around the large loop. This large loop starts at $(0, 0)$, then increases θ from 0 to 2π, keeping ϕ fixed. Then it increases ϕ to 2π, keeping θ fixed. Then, it decreases θ from 2π to 0, keeping ϕ fixed. Finally, it decreases ϕ from 2π to 0, keeping θ fixed. Each of those four segments of the evolution approximately returns the ground state to itself, up to some phase; this again is shown by an energy argument. However, using the 2π periodicity in the parameters θ, ϕ, the phases cancel. Thus, the combined evolution gives a phase which is approximately an integer multiple of 2π, and since this phase is approximately the product of the phases around the small loops, the phase for each small loop is approximately an integer multiple of $(2\pi)^{-1}$ times the area of the loop. This part of the proof is, of course, very similar to one way to show that the average of the Berry curvature over flux torus is quantized; essentially, it is a form of Stokes' theorem. However, since we have used a quasiadiabatic evolution so that all the small loops contribute approximately the same phase, and since we have related the phase for the small loop near $(\theta, \phi) = (0, 0)$ to the Berry curvature, it proves quantization without the averaging assumption and without any assumption of the gap remaining open.

A. LIEB–ROBINSON BOUNDS

A.1. Lieb–Robinson bound

Here we show

Lemma 3. *Given operators A supported on X and B supported on Y with $X \cap Y = \emptyset$, we have*

$$\begin{aligned}\big\|\big[A(t), B\big]\big\| \le{} & 2\|A\| \cdot \|B\| (2|t|) \sum_{Z_1 : Z_1 \cap X \neq \emptyset, Z_1 \cap Y \neq \emptyset} \|h_{Z_1}\| \\ & + 2\|A\| \cdot \|B\| \frac{(2|t|)^2}{2!} \sum_{Z_1 : Z_1 \cap X \neq \emptyset} \|h_{Z_1}\| \sum_{Z_2 : Z_2 \cap Z_1 \neq \emptyset, Z_2 \cap Y \neq \emptyset} \|h_{Z_2}\| \\ & + 2\|A\| \cdot \|B\| \frac{(2|t|)^3}{3!} \sum_{Z_1 : Z_1 \cap X \neq \emptyset} \|h_{Z_1}\| \sum_{Z_2 : Z_2 \cap Z_1 \neq \emptyset} \|h_{Z_2}\| \\ & \times \sum_{Z_3 : Z_3 \cap Z_2 \neq \emptyset, Z_3 \cap Y \neq \emptyset} \|h_{Z_3}\| \\ & + \cdots \end{aligned} \tag{A.1}$$

Remark 5. The kth term of the above series is equal to $2\|A\| \cdot \|B\| \frac{(2|t|)^k}{k!}$ times the sum over sets $Z_1, \dots, Z_k$ with $Z_1 \cap X \neq \emptyset$, $Z_j \cap Z_{j+1} \neq \emptyset$ for $0 \le j < k$, and $Z_k \cap Y \neq \emptyset$, of the product $\prod_{j=1}^{k} \|h_{Z_j}\|$.

Proof. We assume $t > 0$ because negative t can be treated in the same way. Let $\varepsilon = t/N$ with a large positive integer N, and let

$$t_n = \frac{t}{N} n \quad \text{for } n = 0, 1, \dots, N.$$

Then we have

$$\big\|\big[A(t), B\big]\big\| - \big\|\big[A(0), B\big]\big\| = \sum_{i=0}^{N-1} \varepsilon \times \frac{\|[A(t_{n+1}), B]\| - \|[A(t_n), B]\|}{\varepsilon}. \tag{A.2}$$

In order to obtain the bound (A.10) below, we want to estimate the summand in the right-hand side. To begin with, we note that the identity $\|U^* O U\| = \|O\|$ holds for any observable O and for any unitary operator U. Using this fact, we have

$$\begin{aligned} & \big\|\big[A(t_{n+1}), B\big]\big\| - \big\|\big[A(t_n), B\big]\big\| \\ & = \big\|\big[A(\varepsilon), B(-t_n)\big]\big\| - \big\|\big[A, B(-t_n)\big]\big\| \\ & \le \big\|\big[A + i\varepsilon[h_\Lambda, A], B(-t_n)\big]\big\| - \big\|\big[A, B(-t_n)\big]\big\| + \mathcal{O}\big(\varepsilon^2\big) \\ & = \big\|\big[A + i\varepsilon[I_X, A], B(-t_n)\big]\big\| - \big\|\big[A, B(-t_n)\big]\big\| + \mathcal{O}\big(\varepsilon^2\big), \end{aligned} \tag{A.3}$$

with

$$I_X = \sum_{Z : Z \cap X \neq \emptyset} h_Z, \tag{A.4}$$

where we have used

$$A(\varepsilon) = A + i\varepsilon[h_\Lambda, A] + \mathcal{O}\big(\varepsilon^2\big) \tag{A.5}$$

and the triangle inequality. Further, by using

$$A + i\varepsilon[I_X, A] = e^{i\varepsilon I_X} A e^{-i\varepsilon I_X} + \mathcal{O}\big(\varepsilon^2\big), \tag{A.6}$$

we have

$$\begin{aligned}\big\|\big[A + i\varepsilon[I_X, A], B(-t_n)\big]\big\| &\le \big\|\big[e^{i\varepsilon I_X} A e^{-i\varepsilon I_X}, B(-t_n)\big]\big\| + \mathcal{O}\big(\varepsilon^2\big)\\ &= \big\|\big[A, e^{-i\varepsilon I_X} B(-t_n) e^{i\varepsilon I_X}\big]\big\| + \mathcal{O}\big(\varepsilon^2\big)\\ &\le \big\|\big[A, B(-t_i) - i\varepsilon\big[I_X, B(-t_n)\big]\big]\big\| + \mathcal{O}\big(\varepsilon^2\big)\\ &\le \big\|\big[A, B(-t_n)\big]\big\| + \varepsilon\big\|\big[A, \big[I_X, B(-t_n)\big]\big]\big\| + \mathcal{O}\big(\varepsilon^2\big).\end{aligned} \tag{A.7}$$

Substituting this into the right-hand side in the last line of (A.3), we obtain

$$\begin{aligned}\big\|\big[A(t_{n+1}), B\big]\big\| - \big\|\big[A(t_n), B\big]\big\| &\le \varepsilon\big\|\big[A, \big[I_X, B(-t_n)\big]\big]\big\| + \mathcal{O}\big(\varepsilon^2\big)\\ &\le 2\varepsilon\|A\|\big\|\big[I_X(t_n), B\big]\big\| + \mathcal{O}\big(\varepsilon^2\big).\end{aligned} \tag{A.8}$$

Further substituting this into the right-hand side of (A.2) and using (A.4), we have

$$\begin{aligned}\big\|\big[A(t), B\big]\big\| - \big\|\big[A(0), B\big]\big\| &\le 2\|A\| \sum_{n=0}^{N-1} \varepsilon \times \big\|\big[I_X(t_n), B\big]\big\| + \mathcal{O}(\varepsilon)\\ &\le 2\|A\| \sum_{Z:Z\cap X\neq\emptyset} \sum_{n=0}^{N-1} \varepsilon \times \big\|\big[h_Z(t_n), B\big]\big\| + \mathcal{O}(\varepsilon).\end{aligned} \tag{A.9}$$

Since $h_Z(t)$ is a continuous function of the time t for a finite volume, the sum in the right-hand side converges to the integral in the limit $\varepsilon \downarrow 0$ (i.e., $N \uparrow \infty$) for any fixed finite lattice Λ. In consequence, we obtain

$$\big\|\big[A(t), B\big]\big\| - \big\|\big[A(0), B\big]\big\| \le 2\|A\| \sum_{Z:Z\cap X\neq\emptyset} \int_0^{|t|} ds \big\|\big[h_Z(s), B\big]\big\|. \tag{A.10}$$

We define

$$C_B(X, t) := \sup_{A\in\mathcal{A}_X} \frac{\|[A(t), B]\|}{\|A\|}, \tag{A.11}$$

where $\mathcal{A}_X$ is the algebra of observables supported on the set X. Then we have

$$C_B(X, t) \le C_B(X, 0) + 2 \sum_{Z:Z\cap X\neq\emptyset} \|h_Z\| \int_0^{|t|} ds\, C_B(Z, s) \tag{A.12}$$

from the above bound (A.10). Assume $\mathrm{dist}(X, Y) > 0$. Then we have $C_B(X, 0) = 0$ from the definition of $C_B(X, t)$, and note that

$$C_B(Z, 0) \le 2\|B\|, \tag{A.13}$$

for $Z \cap Y \neq \emptyset$, and

$$C_B(Z, 0)) = 0 \tag{A.14}$$

otherwise. Using these facts and the above bound (A.12) iteratively, we obtain

$$\begin{aligned}C_B(X,t) &\leq 2 \sum_{Z_1:Z_1\cap X\neq\emptyset} \|h_{Z_1}\| \int_0^{|t|} ds_1\, C_B(Z_1,s_1)\\ &\leq 2 \sum_{Z_1:Z_1\cap X\neq\emptyset} \|h_{Z_1}\| \int_0^{|t|} ds_1\, C_B(Z_1,0)\\ &\quad + 2^2 \sum_{Z_1:Z_1\cap X\neq\emptyset} \|h_{Z_1}\| \sum_{Z_2:Z_2\cap Z_1\neq\emptyset} \|h_{Z_2}\| \int_0^{|t|} ds_1 \int_0^{|s_1|} ds_2\, C_B(Z_2,s_2)\\ &\leq \cdots .\end{aligned}$$

So,

$$\begin{aligned}C_B(X,t) \leq 2\|B\|(2|t|) &\sum_{Z_1:Z_1\cap X\neq\emptyset, Z_1\cap Y\neq\emptyset} \|h_{Z_1}\|\\ &+ 2\|B\|\frac{(2|t|)^2}{2!} \sum_{Z_1:Z_1\cap X\neq\emptyset} \|h_{Z_1}\| \sum_{Z_2:Z_2\cap Z_1\neq\emptyset, Z_2\cap Y\neq\emptyset} \|h_{Z_2}\|\\ &+ 2\|B\|\frac{(2|t|)^3}{3!} \sum_{Z_1:Z_1\cap X\neq\emptyset} \|h_{Z_1}\| \sum_{Z_2:Z_2\cap Z_1\neq\emptyset} \|h_{Z_2}\|\\ &\times \sum_{Z_3:Z_3\cap Z_2\neq\emptyset, Z_3\cap Y\neq\emptyset} \|h_{Z_3}\| + \cdots .\end{aligned}$$ ■

REFERENCES

[1] I. Affleck, Spin gap and symmetry breaking in CuO$_2$ layers and other antiferromagnets. *Phys. Rev. B* **37** (1988), no. 10, 5186.

[2] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.* **59** (1987), no. 7, 799.

[3] I. Affleck and E. H. Lieb, A proof of part of Haldane's conjecture on spin chains. *Lett. Math. Phys.* **12** (1986), no. 1, 57–69.

[4] J. E. Avron and R. Seiler, Quantization of the hall conductance for general, multiparticle Schrödinger Hamiltonians. *Phys. Rev. Lett.* **54** (1985), no. 4, 259.

[5] J. Bellissard, A. van Elst, and H. Schulz-Baldes, The noncommutative geometry of the quantum Hall effect. *J. Math. Phys.* **35** (1994), no. 10, 5373–5451.

[6] S. Bravyi, M. B. Hastings, and F. Verstraete, Lieb–Robinson bounds and the generation of correlations and topological quantum order. *Phys. Rev. Lett.* **97** (2006), no. 5, 050401.

[7] F. D. M. Haldane, Continuum dynamics of the 1-d Heisenberg antiferromagnet: identification with the $O(3)$ nonlinear sigma model. *Phys. Lett. A* **93** (1983), no. 9, 464–468.

[8] M. B. Hastings, Lieb–Schultz–Mattis in higher dimensions. *Phys. Rev. B* **69** (2004), no. 10, 104431.

[9] M. B. Hastings and T. Koma, Spectral gap and exponential decay of correlations. *Comm. Math. Phys.* **265** (2006), no. 3, 781–804.

[10] M. B. Hastings and S. Michalakis, Quantization of Hall conductance for interacting electrons on a torus. *Comm. Math. Phys.* **334** (2015), no. 1, 433–471.

[11] A. Ingham, A note on Fourier transforms. *J. Lond. Math. Soc.* **1** (1934), no. 1, 29–32.

[12] T. Koma, Spectral gaps of quantum hall systems with interactions. *J. Stat. Phys.* **99** (2000), no. 1, 313–381.

[13] R. B. Laughlin, Quantized Hall conductivity in two dimensions. *Phys. Rev. B* **23** (1981), no. 10, 5632.

[14] E. H. Lieb and D. W. Robinson, The finite group velocity of quantum spin systems. In *Statistical mechanics*, pp. 425–431, Springer, 1972.

[15] E. Lieb, T. Schultz, and D. Mattis, Two soluble models of an antiferromagnetic chain. *Ann. Physics* **16** (1961), no. 3, 407–466.

[16] C. K. Majumdar and D. K. Ghosh, On next-nearest-neighbor interaction in linear chain. I. *J. Math. Phys.* **10** (1969), no. 8, 1388–1398.

[17] G. Misguich and C. Lhuillier, Some remarks on the Lieb–Schultz–Mattis theorem and its extension to higher dimensions. 2000, arXiv:cond-mat/0002170.

[18] T. J. Osborne, Simulating adiabatic evolution of gapped spin systems. *Phys. Rev. A* **75** (2007), no. 3, 032321.

[19] M. Oshikawa, Commensurability, excitation gap, and topology in quantum many-particle systems on a periodic lattice. *Phys. Rev. Lett.* **84** (2000), no. 7, 1535.

[20] N. Read and B. Chakraborty, Statistics of the excitations of the resonating-valence-bond state. *Phys. Rev. B* **40** (1989), no. 10, 7133.

[21] D. S. Rokhsar and S. A. Kivelson, Superconductivity and the quantum hard-core dimer gas. *Phys. Rev. Lett.* **61** (1988), no. 20, 2376.

[22] B. Sutherland, Systems with resonating-valence-bond ground states: Correlations and excitations. *Phys. Rev. B* **37** (1988), no. 7, 3786.

MATTHEW B. HASTINGS

Microsoft Quantum and Microsoft Research, Redmond, WA 98052, USA, mahastin@microsoft.com

BOOTSTRAP APPROACH TO 1+1-DIMENSIONAL INTEGRABLE QUANTUM FIELD THEORIES: THE CASE OF THE SINH-GORDON MODEL

KAROL KAJETAN KOZLOWSKI

ABSTRACT

1+1-dimensional integrable quantum field theories correspond to a sparse subset of quantum field theories where the calculation of physically interesting observables can be brought to explicit, closed, and manageable expressions thanks to the factorizability of the $\mathbf{S}$ matrices which govern the scattering in these models. In particular, the correlation functions are expressed in terms of explicit series of multiple integrals, this nonperturbatively for all values of the coupling. However, the question of convergence of these series, and thus the mathematical well-definiteness of these correlators, is mostly open. This paper reviews the overall setting used to formulate such models and discusses the recent progress relative to solving the convergence issues in the case of the 1+1-dimensional massive integrable Sinh-Gordon quantum field theory.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 82B23; Secondary 81Q80,45E10,33E20,30E25

KEYWORDS

Integrable quantum field theories, convergence, exactly solvable models, Riemann–Hilbert problems, concentration of measure

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4096–4118
DOI 10.4171/ICM2022/96

1. INTRODUCTION

1.1. Scattering matrices for quantum integrable field theories

It was discovered in the early 20th century that the description of matter at low-scales demands to wave-off some of the existing at the time paradigms governing the motion and very structure of particles in interactions. This led to the development of the theory of relativity on the one hand, and quantum mechanics on the other. In the latter setting, the state of a physical system is described by a vector, the wave function, belonging to some Hilbert space and supposed to encapsulate all the physical degrees of freedom of that system. On the classical level, the time evolution of particles' momenta and positions is governed by a set of generically nonlinear ordinary differential equations which can be written in the form of Hamilton's equations. In its turn, the time evolution of a wave function is governed by a first-order ordinary linear differential equation driven by the Hamiltonian operator. This operator is obtained through a quantization procedure: its symbol is given by the classical Hamiltonian of the system or, said differently, it is obtained from the classical Hamiltonian upon replacing the classical momenta and positions by operators. While the success of the approach was astonishing relatively to the amount of experiments which could have been explained, soon after the early development of the theory it became clear that in order to describe physics at even smaller scales or higher energies, one needs to develop a quantum theory of fields which would bring together the quantum and relativistic features in the setting of uncountably many degrees of freedom. In loose words, such a theory would be reached by producing operator valued generalized functions, viz. formal kernels of distributions, depending on the space-time coordinates which would satisfy analogues of nonlinear, relativistically invariant, evolution equations arising in classical field theory. While it was rather straightforward to construct the quantum theory of the free field (and nowadays such a construction is fully rigorous), the construction of interacting theories which are the sole relevant for physics appeared to be a tremendously hard task, this even on a formal level of rigor. The various approaches that were developed quickly met serious problems: the most prominent being the divergence of coefficients supposed to describe the formal perturbative expansions of physical observables around the free theories. Eventually, these problems could have been formally circumvented in certain cases by the use of the so-called renormalization procedure. The latter, while being able to produce numbers which were measured with great agreement in collider experiments, eluded for very long any attempts at making it rigorous. Some progress was eventually achieved for several instances of truly interacting, viz. nonfree, quantum field theories within the so-called constructive quantum field theory approach, see [35] for a review. While successful in rigorously showing the existence and certain overall properties of such theories, the approach did not lead yet to rigorous and manageable expressions for the correlation functions, which are the quantities measured in experiments and thus of prime interest to the theory.

Among the various alternatives to renormalization, one may single out the $\mathbf{S}$-matrix program which aimed at describing a quantum field theory directly in terms of the quantities that are measured in experiments. This led to a formulation of the theory in terms of matrix-valued functions in n complex variables, with $n = 0, 1, 2, \ldots$, that correspond to the entries of the $\mathtt{S}$-matrix between asymptotic states. The $\mathtt{S}$-matrix program was actively investigated in the 1960s and 1970s and numerous attempts were made to characterize the $\mathtt{S}$-matrix which is the central object in this approach, see, e.g., [14,17]. However, these investigations led to rather unsatisfactory results in spacial dimensions higher than one, mainly due to the incapacity of constructing viable, explicit, $\mathtt{S}$-matrices for nontrivial models.

The interest in the $\mathtt{S}$-matrix approach was revived by the pioneering work of Gryanik and Vergeles [16]. These authors set forth the first features of an integrable structure based method for determining $\mathtt{S}$-matrices for the 1+1-dimensional quantum field theories whose classical analogues exhibit an infinite set of independent local integrals of motion. Indeed, the existence of analogous conservation laws on the quantum level heavily constrains the possible form of the scattering basically by reducing it to a concatenation of two-body processes and hence making the calculations of $\mathtt{S}$-matrices feasible. The work [16] focused on the case of models only exhibiting one type of asymptotic particles, the main example being given by the quantum Sinh-Gordon model. This 1+1-dimensional quantum field theory will be taken as a guiding example from now on. It corresponds to the appropriate quantization of the classical evolution equation of a scalar field $\varphi(x,t)$ under the partial differential equation

$$\big(\partial_t^2 - \partial_x^2\big)\varphi + \frac{m^2}{g}\sinh(g\varphi) = 0, \quad (x,t) \in \mathbb{R}^2. \tag{1.1}$$

For this model, the asymptotic "in" states of the theory are described by vectors $\boldsymbol{f} = (f^{(0)}, \ldots, f^{(n)}, \ldots)$ which belong to the Fock Hilbert space

$$\mathfrak{h}_{\mathrm{in}} = \bigoplus_{n=0}^{+\infty} L^2\big(\mathbb{R}^n_>\big) \quad \text{with } \mathbb{R}^n_> = \big\{\boldsymbol{\beta}_n = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n : \beta_1 > \cdots > \beta_n\big\}. \tag{1.2}$$

This means that $f^{(n)} \in L^2(\mathbb{R}^n_>)$ has the physical interpretation of an incoming n-particle wave-packet density in rapidity space. More precisely, on physical grounds, one interprets elements of the Hilbert space $\mathfrak{h}_{\mathrm{in}}$ as parameterized by n-particles states, $n \in \mathbb{N}$, arriving, in the remote past, with well-ordered rapidities $\beta_1 > \cdots > \beta_n$ prior to any scattering which would be enforced by the interacting nature of the model.

For the 1+1-dimensional quantum Sinh-Gordon model, the $\mathtt{S}$-matrix proposed in [16] is purely diagonal and thus fully described by one scalar function of the relative "in" rapidities of the two particles:

$$\mathtt{S}(\beta) = \frac{\tanh[\frac{1}{2}\beta - \mathrm{i}\pi\mathfrak{b}]}{\tanh[\frac{1}{2}\beta + \mathrm{i}\pi\mathfrak{b}]} \quad \text{with } \mathfrak{b} = \frac{1}{2}\frac{g^2}{8\pi + g^2}. \tag{1.3}$$

This $\mathtt{S}$-matrix satisfies the unitarity $\mathtt{S}(\beta)\mathtt{S}(-\beta) = 1$ and crossing $\mathtt{S}(\beta) = \mathtt{S}(\mathrm{i}\pi - \beta)$ symmetries. These are, in fact, fundamental symmetry features of an $\mathtt{S}$-matrix and arise in many other integrable quantum field theories. Within the physical picture, throughout the flow of time, the "in" particles approach each other, interact, scatter and finally travel again as asymptotically free outgoing, viz. "out," particles. Within such a scheme, an "out" n-particle state is then parameterized by n well-ordered rapidities $\beta_1 < \cdots < \beta_n$ and can be seen as a component of a vector belonging to the Hilbert space

$$\mathfrak{h}_{\text{out}} = \bigoplus_{n=0}^{+\infty} L^2\big(\mathbb{R}^n_<\big) \quad \text{with} \quad \mathbb{R}^n_< = \big\{\boldsymbol{\beta}_n = (\beta_1, \dots, \beta_n) \in \mathbb{R}^n : \beta_1 < \cdots < \beta_n\big\}. \tag{1.4}$$

The $\mathtt{S}$-matrix will allow one to express the "out" state $\boldsymbol{g} = (g^{(0)}, \dots, g^{(n)}, \dots)$ which results from the scattering of an "in" state $\boldsymbol{f} = (f^{(0)}, \dots, f^{(n)}, \dots)$ as

$$g^{(n)}(\beta_1, \dots, \beta_n) = \prod_{a<b}^{n} \mathtt{S}(\beta_a - \beta_b) \cdot f^{(n)}(\beta_n, \dots, \beta_1). \tag{1.5}$$

Note that in this integrable setting, there is *no* particle production and that the scattering is a concatenation of two-body processes.

Over the years, it turned out to be possible to characterize thoroughly the $\mathtt{S}$-matrices for more involved quantum field theories underlying to other integrable classical field theories in 1+1 dimensions. Such models possess several types of asymptotic particles which can also form bound states. Then, the "in" Fock Hilbert space is more complicated and takes the form $\bigoplus_{n=0}^{+\infty} L^2(\mathbb{R}^n_>, \otimes^n \mathbb{C}^p)$ where the L^2-space refers to $\otimes^n \mathbb{C}^p$ valued functions on $\mathbb{R}^n_>$, with p corresponding to the number of different asymptotic particles in the given theory. The most celebrated example corresponds to the Sine-Gordon quantum field theory. Building on Faddeev–Korepin's [22] semiclassical quantization results of the solitons in the classical Sine-Gordon model, one concludes that the underlying quantum field theory possesses two distinct types of asymptotic particles of equal mass, the soliton and the antisoliton, as well as a certain number, which depends on the coupling constant, of bound states thereof. These all have distinct masses and are called breathers. Zamolodchikov argued the explicit form of the $\mathtt{S}$-matrix governing the soliton–antisoliton scattering [39] upon using the factorizability of the n-particle $\mathtt{S}$-matrix into two-particle processes, the independence of the order in which a three particle scattering process arises from a concatenation of two-particle processes as well as the fact that equal mass particles may *solely* exchange their momenta during scattering, this due to the existence of many conservation laws. This enforces that the $\mathtt{S}$ matrix satisfies the Yang–Baxter equation, which originally appeared in rather different contexts [5, 37], and strongly restricts its form. We do stress that the Yang–Baxter equation is the actual cornerstone of quantum integrability, so that it is not astonishing to recover it also in this setting. The missing pieces of the Sine-Gordon $\mathtt{S}$-matrix capturing the soliton–breather and breather–breather scattering were then proposed in [18]. Nowadays, $\mathtt{S}$-matrices of many other models have been proposed, see e.g., [1, 38].

1.2. The operator content and the bootstrap program

1.2.1. The basic operators

Having in mind the per se full construction of the quantum field theory, identifying the content in asymptotic particles, viz. the "in" particles' Hilbert space $\mathfrak{h}_{\text{in}}$, and the $\mathtt{S}$-matrix which describes their scattering only arises as the first step. Indeed, one should build, in a way that is compatible with the form of the scattering encapsulated in the $\mathtt{S}$-matrix of interest, a family $\mathtt{O}_\alpha$ of operator-valued distributions, α running through some set $\mathcal{S}$. More precisely, the $\mathtt{O}_\alpha$ should be distributions acting on smooth, compactly supported functions $d(\boldsymbol{x})$ of the Minkowskian space-time coordinate

$$\boldsymbol{x} = (x_0, x_1) \in \mathbb{R}^{1,1} \quad \text{with } \boldsymbol{x} \cdot \boldsymbol{y} = x_0 y_0 - x_1 y_1. \tag{1.6}$$

Then $\mathtt{O}_\alpha[d]$ is some densely defined operator on $\mathfrak{h}_{\text{in}}$ whose domain could, in principle, depend on d. It is useful from the point of view of connecting this picture to physics to express $\mathtt{O}_\alpha$ directly in terms of its generalized operator valued function

$$\mathtt{O}_\alpha[d] = \int_{\mathbb{R}^2} \mathrm{d}^2\boldsymbol{x}\; d(\boldsymbol{x}) \mathtt{O}_\alpha(\boldsymbol{x}). \tag{1.7}$$

In fact, in physics' terminology, it is the $\mathtt{O}_\alpha(\boldsymbol{x})$s which correspond to the quantum fields of the theory. Moreover, as will be apparent in the following, it turns out that in most handlings $\mathtt{O}_\alpha(\boldsymbol{x})$ does actually make sense as a bona fide operator valued *function* on the Minkowski space having a well-defined dense domain. Hence, unless it is mandatory so as to make an appropriate sense out of the formula, we will make use of the generalized function notation $\mathtt{O}_\alpha(\boldsymbol{x})$.

On top of being compatible with the scattering date, the operators $\mathtt{O}_\alpha(\boldsymbol{x})$ should form an algebra, viz. the product $\mathtt{O}_\alpha(\boldsymbol{x})\mathtt{O}_{\alpha'}(\boldsymbol{y})$ should be a well-defined dense operator for almost all $\boldsymbol{x}$ and $\boldsymbol{y}$, and satisfy causality, viz. that for purely Bosonic theories as the Sinh-Gordon model

$$\big[\mathtt{O}_\alpha(\boldsymbol{x}), \mathtt{O}_{\alpha'}(\boldsymbol{y})\big] \equiv \mathtt{O}_\alpha(\boldsymbol{x})\mathtt{O}_{\alpha'}(\boldsymbol{y}) - \mathtt{O}_{\alpha'}(\boldsymbol{y})\mathtt{O}_\alpha(\boldsymbol{x}) = 0 \quad \text{if } (\boldsymbol{x} - \boldsymbol{y})^2 < 0, \tag{1.8}$$

namely when $\boldsymbol{x} - \boldsymbol{y}$ is space-like. The family $\mathtt{O}_\alpha(\boldsymbol{x})$ should in particular contain the per se quantized counterparts of the classical fields arising in the original evolution equation, for instance, $\boldsymbol{\Phi}(\boldsymbol{x})$ or $\mathrm{e}^{\gamma\boldsymbol{\Phi}}(\boldsymbol{x})$ in the Sinh-Gordon quantum field theory case. Moreover, these operators should comply with the various other symmetries imposed on a quantum field theory, such as invariance under Lorentz boosts of space-time coordinates or translational invariance. In the quantum Sinh-Gordon field theory on which we shall focus from now on, the latter means that the model is naturally endowed with a unitary operator $\mathtt{U}_{\mathtt{T}_y}$ such that for any operator $\mathtt{O}(\boldsymbol{x})$

$$\mathtt{U}_{\mathtt{T}_y} \cdot \mathtt{O}(\boldsymbol{x}) \cdot \mathtt{U}_{\mathtt{T}_y}^{-1} = \mathtt{O}(\boldsymbol{x} + \boldsymbol{y}). \tag{1.9}$$

The operator $\mathtt{U}_{\mathtt{T}_y}$ acts diagonally on $\mathfrak{h}_{\text{in}}$ given in (1.2):

$$\mathtt{U}_{\mathtt{T}_y} \cdot \boldsymbol{f} = \big(\mathtt{U}_{\mathtt{T}_y}^{(0)} \cdot f^{(0)}, \dots, \mathtt{U}_{\mathtt{T}_y}^{(n)} \cdot f^{(n)}, \dots\big) \quad \text{with } \boldsymbol{f} = \big(f^{(0)}, \dots, f^{(n)}, \dots\big) \tag{1.10}$$

and where

$$\mathtt{U}_{\mathtt{T}_y}^{(n)} \cdot f^{(n)}(\boldsymbol{\beta}_n) = \exp\Bigg\{\mathrm{i}\sum_{a=1}^{n} \boldsymbol{p}(\beta_a) \cdot \boldsymbol{y}\Bigg\} f^{(n)}(\boldsymbol{\beta}_n), \tag{1.11}$$

with $\boldsymbol{p}(\beta) = (m\cosh(\beta), m\sinh(\beta))$ and $\boldsymbol{\beta}_n = (\beta_1, \dots, \beta_n)$.

Should the construction of quantum fields fulfilling to the above be achieved, the ultimate goal would consist in computing in closed and explicit form the model's vacuum-to-vacuum n-point correlation functions:

$$\big\langle \mathtt{O}_{\alpha_1}(\boldsymbol{x}_1) \cdots \mathtt{O}_{\alpha_n}(\boldsymbol{x}_n) \big\rangle = \mathrm{Tr}_{\mathfrak{h}_{\text{in}}}\big[\mathtt{P}_0 \mathtt{O}_{\alpha_1}(\boldsymbol{x}_1) \cdots \mathtt{O}_{\alpha_n}(\boldsymbol{x}_n) \mathtt{P}_0\big], \tag{1.12}$$

with $\mathtt{P}_0$ being the orthogonal projection on the 0-particle Fock space. We do stress that the above objects are still generalized functions and, as such, should be considered in an appropriate distributional interpretation. That will be made precise below.

1.2.2. The bootstrap program for the zero particle sector

By virtue of the above, in the case of the $\mathfrak{h}_{\mathrm{in}}$ Hilbert space, one may represent an operator $\mathtt{O}(\boldsymbol{x})$ as an integral operator acting on the L^2-based Fock space

$$\mathtt{O}(\boldsymbol{x}) \cdot \boldsymbol{f} = \big(\mathtt{O}^{(0)}(\boldsymbol{x}) \cdot \boldsymbol{f}, \dots, \mathtt{O}^{(n)}(\boldsymbol{x}) \cdot \boldsymbol{f}, \dots\big) \tag{1.13}$$

with $\mathtt{O}^{(n)}(\boldsymbol{x}) : \mathfrak{h}_{\mathrm{in}} \to L^2(\mathbb{R}^n_>)$. Later on, we will discuss more precisely the structure of the operators $\mathtt{O}^{(n)}(\boldsymbol{x})$ that one needs to impose so as to end up with a consistent quantum field theory. However, first, we focus our attention on the 0^{th} space operators whose action may be represented, whenever it makes sense, as

$$\mathtt{O}^{(0)}(\boldsymbol{x}) \cdot \boldsymbol{f} = \sum_{m \geq 0} \int_{\mathbb{R}^m_>} \mathrm{d}^m \beta \, \mathcal{M}^{(\mathtt{O})}_{0;m}(\boldsymbol{\beta}_m) \prod_{a=1}^{m} \big\{ \mathrm{e}^{-\mathrm{i}\boldsymbol{p}(\beta_a)\cdot\boldsymbol{x}} \big\} f^{(m)}(\boldsymbol{\beta}_m). \tag{1.14}$$

The oscillatory $\boldsymbol{x}$-dependence is a simple consequence of the translation relation (1.9) along with the explicit form of the action of the translation operator (1.11).

In order for $\mathtt{O}^{(0)}(\boldsymbol{x})$ to comply with the scattering data encoded by $\mathtt{S}$, one needs to impose a certain amount of constraints on the integral kernels $\mathcal{M}^{(\mathtt{O})}_{0;m}(\boldsymbol{\beta}_m)$. First of all, general principles of quantum field theory impose that, in order for these to correspond to kernels of quantum fields, the $\mathcal{M}^{(\mathtt{O})}_{0;m}(\boldsymbol{\beta}_m)$ have to correspond to a $+$ boundary value $\mathcal{F}^{(\mathtt{O})}_{m;+}(\boldsymbol{\beta}_m)$ on $\mathbb{R}^m_>$ of a meromorphic function $\mathcal{F}^{(\mathtt{O})}_m(\boldsymbol{\beta}_m)$ of the variables β_a belonging to the strip

$$\mathscr{S} = \big\{ z \in \mathbb{C} : 0 < \Im(z) < 2\pi \big\}. \tag{1.15}$$

Traditionally, in the physics literature, the functions $\mathcal{F}^{(\mathtt{O})}_m(\boldsymbol{\beta}_m)$ are called form factors.

Further, one imposes a set of equations on the $\mathcal{F}^{(\mathtt{O})}_m$s. These constitute the so-called form factor bootstrap program. On mathematical grounds, one should understand the form factor bootstrap program as a set of *axioms* that one imposes as a starting point of the theory given the data $(\mathfrak{h}_{\mathrm{in}}, \mathtt{S})$. Upon solving them, one has to check a posteriori that their solutions do provide one, through (1.14) and (1.16), with a collection of operators satisfying all of the requirements of the theory discussed earlier on.

The bootstrap program axioms take the form of a Riemann–Hilbert problem for a collection of functions in many variables. In the case of the Sinh-Gordon model, since there are no bound states, these take the below form.

Form Factor Axioms 1.1. *Find functions $\mathcal{F}^{(\mathtt{O})}_n$, $n \in \mathbb{N}$, such that, for each $k \in [\![1;n]\!]$ and fixed $\beta_a \in \mathscr{S}$, $a \neq k$, the maps $\beta_k \mapsto \mathcal{F}^{(\mathtt{O})}_n(\boldsymbol{\beta}_n)$ are*

- *meromorphic on $\mathscr{S}$;*
- *admit $+$, resp. $-$, boundary values $\mathcal{F}^{(\mathtt{O})}_{n;+}$ on $\mathbb{R}$, resp. $\mathcal{F}^{(\mathtt{O})}_{n;-}$ on $\mathbb{R} + 2\mathrm{i}\pi$;*
- *are bounded at infinity by $C \cdot \cosh(\ell \Re(\beta_k))$ for some n and k independent ℓ.*

The $\mathscr{F}_n^{(\mathsf{O})}$ satisfy the multivariate system of Riemann–Hilbert problems:

(i) *agreeing upon $\beta_{ab} = \beta_a - \beta_b$, one has $\mathscr{F}_n^{(\mathsf{O})}(\beta_1, \dots, \beta_a, \beta_{a+1}, \dots, \beta_n) = \mathbf{S}(\beta_{aa+1}) \cdot \mathscr{F}_n^{(\mathsf{O})}(\beta_1, \dots, \beta_{a+1}, \beta_a, \dots, \beta_n)$;*

(ii) *For $\beta_1 \in \mathbb{R}$, and given generic $\boldsymbol{\beta}'_n = (\beta_2, \dots, \beta_n) \in \mathscr{S}^{n-1}$, it holds $\mathscr{F}_{n;-}^{(\mathsf{O})}(\beta_1 + 2\mathrm{i}\pi, \boldsymbol{\beta}'_n) = \mathscr{F}_{n;+}^{(\mathsf{O})}(\boldsymbol{\beta}'_n, \beta_1) = \prod_{a=2}^{n} \mathbf{S}(\beta_{a1}) \cdot \mathscr{F}_{n;+}^{(\mathsf{O})}(\boldsymbol{\beta}_n)$;*

(iii) *The only poles of $\mathscr{F}_n^{(\mathsf{O})}$ are simple, located at $\mathrm{i}\pi$ shifted rapidities and*
$$-\mathrm{i}\mathrm{Res}\big(\mathscr{F}_{n+2}^{(\mathsf{O})}(\alpha + \mathrm{i}\pi, \beta, \boldsymbol{\beta}_n) \cdot \mathrm{d}\alpha, \alpha = \beta\big) = \left\{1 - \prod_{a=1}^{n} \mathbf{S}(\beta - \beta_a)\right\} \cdot \mathscr{F}_n^{(\mathsf{O})}(\boldsymbol{\beta}_n);$$

(iv) *$\mathscr{F}_n^{(\mathsf{O})}(\boldsymbol{\beta}_n + \theta \overline{\boldsymbol{e}}_n) = \mathrm{e}^{\theta \mathsf{s}_{\mathsf{O}}} \cdot \mathscr{F}_n^{(\mathsf{O})}(\boldsymbol{\beta}_n)$ for some number s_{O} and with $\overline{\boldsymbol{e}}_n = (1, \dots, 1)$.*

Note that the reduction occurring at the residues of $\mathscr{F}_n^{(\mathsf{O})}(\boldsymbol{\beta}_n)$ when $\beta_{ab} = \mathrm{i}\pi$ can be readily inferred from (i) and (iii).

One may already comment on the origin of the axioms. The first one illustrates the scattering properties of the model on the level of the operator's kernel. The second and third axioms may be interpreted heuristically as a consequence of the LSZ reduction [25], and locality of the operator, see, e.g., [2, 34] for heuristics on that matter. Finally, the last axiom is a manifestation of the Lorentz invariance of the theory. The number s_{O} arising in (iv) is called the spin of the operator. Moreover, the number ℓ depends on the type of operator being considered. Finally, for more complex models, one would also need to add an additional axiom which would encapsulate the way how the presence of bound states in the model governs certain additional poles in the form factors, cf. [34].

1.2.3. The bootstrap program for the multiparticle sector

It is convenient to represent the action of the operators $\mathsf{O}^{(n)}(\boldsymbol{x})$ in the form
$$\big(\mathsf{O}^{(n)}(\boldsymbol{x}) \cdot \boldsymbol{f}\big)(\boldsymbol{\gamma}_n) = \sum_{m \geq 0} \prod_{a=1}^{n} \big\{\mathrm{e}^{\mathrm{i}\boldsymbol{p}(\gamma_a)\cdot \boldsymbol{x}}\big\} \cdot \mathbf{M}_{\mathsf{O}}^{(m)}(\boldsymbol{x} \mid \boldsymbol{\gamma}_n)\big[f^{(m)}\big]. \tag{1.16}$$

There $\mathbf{M}_{\mathsf{O}}^{(m)}(\boldsymbol{x} \mid \boldsymbol{\gamma}_n)$ are distribution-valued functions which act on appropriate spaces of sufficiently regular functions in m variables. The regularity assumptions will clear out later on, once that we provide the explicit expressions (1.18) for these distributions. In fact, it is convenient, in order to avoid heavy notations, to represent their action as generalized integral operators
$$\mathbf{M}_{\mathsf{O}}^{(m)}(\boldsymbol{x} \mid \boldsymbol{\gamma}_n)\big[f^{(m)}\big] = \int_{\mathbb{R}_>^m} \mathrm{d}^m\beta\, \mathcal{M}_{n;m}^{(\mathsf{O})}(\boldsymbol{\gamma}_n; \boldsymbol{\beta}_m) \prod_{a=1}^{m} \big\{\mathrm{e}^{-\mathrm{i}\boldsymbol{p}(\beta_a)\cdot \boldsymbol{x}}\big\} f^{(m)}(\boldsymbol{\beta}_m), \tag{1.17}$$

in which one understands of the kernels $\mathcal{M}_{n;m}^{(\mathsf{O})}(\boldsymbol{\gamma}_n; \boldsymbol{\beta}_m)$ as generalized functions.

The last axiom of the bootstrap program provides one with a way to compute these kernels. Heuristically, it can be seen as a consequence of the LSZ reduction [25]:

(v)

$$\begin{aligned}
&\mathcal{M}^{(\mathsf{O})}_{n;m}(\boldsymbol{\alpha}_n;\boldsymbol{\beta}_m)\\
&\quad = \mathcal{M}^{(\mathsf{O})}_{n-1;m+1}\big(\boldsymbol{\alpha}'_n;(\alpha_1+\mathrm{i}\pi,\boldsymbol{\beta}_m)\big)\\
&\qquad + 2\pi\sum_{a=1}^{m}\delta_{\alpha_1;\beta_a}\prod_{k=1}^{a-1}\mathsf{S}(\beta_k-\alpha_1)\cdot\mathcal{M}^{(\mathsf{O})}_{n-1;m-1}\big(\boldsymbol{\alpha}'_n;(\beta_1,\dots,\widehat{\beta}_a,\dots,\beta_m)\big).
\end{aligned}$$

In the above expression, $\widehat{\beta}_a$ means that the variable β_a should be omitted and $\delta_{x;y}$ refers to the Dirac mass distribution centered at x and acting on functions of y. Finally, the evaluation at $\alpha_1+\mathrm{i}\pi$ is understood in the sense of a boundary value of the meromorphic continuation in the strip $0\le\Im(z)\le\pi$ from $\mathbb{R}$ up to $\mathbb{R}+\mathrm{i}\pi$. This axiom is to be complemented with the initialization condition $\mathcal{M}^{(\mathsf{O})}_{0;n}(\emptyset;\boldsymbol{\beta}_n)=\mathcal{F}^{(\mathsf{O})}_{n;+}(\boldsymbol{\beta}_n)$ when $\boldsymbol{\beta}_n\in\mathbb{R}^n_>$. It is direct to establish that the recursion may be solved in closed form allowing one to determine the distributional kernel $\mathcal{M}^{(\mathsf{O})}_{n;m}(\boldsymbol{\alpha}_n;\boldsymbol{\beta}_m)$ in terms of $\mathcal{F}^{(\mathsf{O})}_n(\boldsymbol{\beta}_n)$:

$$\begin{aligned}
\mathcal{M}^{(\mathsf{O})}_{n;m}(\boldsymbol{\alpha}_n;\boldsymbol{\beta}_m) = \sum_{p=0}^{\min(n,m)}\sum_{\substack{k_1<\cdots<k_p\\1\le k_a\le n}}\sum_{\substack{i_1\ne\cdots\ne i_p\\1\le i_a\le m}}\prod_{a=1}^{p}\{2\pi\delta_{\alpha_{k_a};\beta_{i_a}}\}\mathsf{S}\big(\overleftarrow{\boldsymbol{\alpha}}_n\mid\overleftarrow{\boldsymbol{\alpha}}^{(1)}_n\big)&\\
\times\,\mathsf{S}\big(\boldsymbol{\beta}^{(1)}_n\mid\boldsymbol{\beta}_n\big)\cdot\mathcal{F}_{n+m-2p;-}\big(\overleftarrow{\boldsymbol{\alpha}}^{(2)}_n+\mathrm{i}\pi\overline{\boldsymbol{e}}_{n-p},\boldsymbol{\beta}^{(2)}_m\big).&\qquad(1.18)
\end{aligned}$$

There, we have used the shorthand notations $\boldsymbol{\alpha}^{(1)}_n=(\alpha_{k_1},\dots,\alpha_{k_p})$ and $\boldsymbol{\alpha}^{(2)}_n=(\alpha_{\ell_1},\dots,\alpha_{\ell_{n-p}})$ where $\{\ell_1,\dots,\ell_{n-p}\}=[\![1;n]\!]\setminus\{k_a\}^p_1$, $\ell_1<\cdots<\ell_{n-p}$, and analogously $\boldsymbol{\beta}^{(1)}_m=(\beta_{i_1},\dots,\beta_{i_p})$ and $\boldsymbol{\beta}^{(2)}_m=(\beta_{j_1},\dots,\beta_{j_{m-p}})$ where $\{j_1,\dots,j_{m-p}\}=[\![1;m]\!]\setminus\{i_a\}^p_1$, $j_1<\cdots<j_{m-p}$. Moreover, we have introduced

$$\mathsf{S}\big(\overleftarrow{\boldsymbol{\alpha}}_n\mid\overleftarrow{\boldsymbol{\alpha}}^{(1)}_n\big)=\prod_{a=1}^{p}\prod_{\substack{b=1\\k_a>\ell_b}}^{n-p}\mathsf{S}(\alpha_{k_a}-\alpha_{\ell_b}),$$

$$\mathsf{S}\big(\boldsymbol{\beta}^{(1)}_n\mid\boldsymbol{\beta}_n\big)=\prod_{a=1}^{p}\prod_{\substack{b=1\\b<i_a}}^{m}\mathsf{S}(\beta_b-\beta_{i_a})\cdot\prod_{\substack{a>b\\i_a>i_b}}\mathsf{S}(\beta_{i_a}-\beta_{i_b}).$$

Finally, we agree upon $\overleftarrow{\boldsymbol{\gamma}}_N=(\gamma_N,\dots,\gamma_1)$ for any $\boldsymbol{\gamma}_N=(\gamma_1,\dots,\gamma_N)$.

It is clear on the level of the explicit expression (1.18) that this generalized function is well defined, even though it involves a multiplication of distributions.

1.2.4. The road towards the bootstrap program

The first calculation of certain of the operators' kernels was initiated by Weisz [36] who built on the full characterization of the S matrix of the Sine-Gordon model to argue with the help of general principles of quantum field theory an expression for the kernel $\mathcal{M}^{(\mathsf{O})}_{1;1}(\alpha;\beta)$ of the electromagnetic current operator only involving one-dimensional variables α,β. The setting up of a systematic approach allowing one to calculate all the collection of kernels characterizing an operator starting from a given model's S-matrix has been initiated by Karowski and Weisz [19] who proposed a set of equation satisfied by that model's equivalent of $\mathcal{F}^{(O)}_n(\boldsymbol{\beta}_n)$. These allowed them to provide closed-form expressions for two-particle form factors in several models. However, these equations were still far from forming the full bootstrap program as described above.

After long investigations [30, 31, 33] which revealed a deeper structure of the form factors of the Sine-Gordon model, Smirnov [32] formulated the equivalent of axioms (i)–(ii) in that model. Subsequently, Kirillov and Smirnov [20] proposed the full set of the bootstrap program axioms, exemplified in the case of the Massive Thirring model; see also [34].

2. SOLVING THE BOOTSTRAP PROGRAM

The resolution of the bootstrap program was systematized over the years and these efforts led to explicit expressions for the form factors of local operators in numerous 1+1-dimensional massive quantum field theories, see, e.g., [34]. The first expressions for the form factors were rather combinatorial in nature. Later, a substantial progress was achieved in simplifying the latter, in particular by exhibiting a deeper structure at their root. Notably, one can mention the free field based approach, also called angular quantization, to the calculation of form factors. It was introduced by Lukyanov [26] and allowed obtaining convenient representations for certain form factors solving the bootstrap program. In particular, the construction lead to closed-form and manageable expressions [10, 27] for the form factors of the exponential of the field operators in the Sinh-Gordon and the Bullough–Dodd models. Later, Babujian, Fring, Karowski, Zapetal [2] and Babujian, Karowski [3, 4] developed the more powerful $\mathcal{K}$-transform approach which will be described below on the example of the Sinh-Gordon model. The construction of [3, 4] was improved in [15, 24] so as to encompass more complicated operators, the so-called descendants of the Sinh-Gordon exponential of the field operator.

2.1. The 2-particle sector solution

The constructions of solutions to the bootstrap program starts from obtaining a specific solution to the equations (i)–(iv) when $n = 2$, i.e., for two variables. This was first achieved in [19].

Lemma 2.1 ([19]). *Let* $\mathcal{F}_2^{(O)}(\beta_1, \beta_2)$ *solve* (i)–(iv) *at* $n = 2$. *Then, there exist* $k \in \{1, \dots, \ell/2\}$, $\varkappa_a \in \mathbb{C}$, $a = 1, \dots, k$, *such that, with* $\beta_{12} = \beta_1 - \beta_2$

$$\mathcal{F}_2^{(O)}(\beta_1, \beta_2) = \mathcal{N}_O \prod_{a=1}^{k} \left\{ \sinh\left[\frac{\beta_{12} - \varkappa_a}{2} \right] \cdot \sinh\left[\frac{\beta_{12} + \varkappa_a}{2} \right] \right\} \mathrm{e}^{\frac{\mathfrak{s}_O}{2}(\beta_1 + \beta_2)} \boldsymbol{F}(\beta_{12}) \tag{2.1}$$

for some $\mathcal{N}_O \in \mathbb{C}$ *and where* $\mathbf{F}$ *is given by the integral representation valid for* $0 < \Im(\beta) < 2\pi$*:*

$$\mathbf{F}(\beta) = \exp\left\{ -4 \int_0^{+\infty} \mathrm{d}x \, \frac{\sinh(x\mathfrak{b}) \cdot \sinh(x\hat{\mathfrak{b}}) \cdot \sinh(\frac{1}{2}x)}{x \sinh^2(x)} \cos\left(\frac{x}{\pi}(\mathrm{i}\pi - \beta) \right) \right\}, \tag{2.2}$$

with $\hat{\mathfrak{b}} = \frac{1}{2} - \mathfrak{b}$.

Proof. Axiom (iv) ensures that $\mathcal{F}_2^{(O)}(\beta_1, \beta_2) = \mathrm{e}^{\frac{\mathfrak{s}_O}{2}(\beta_1 + \beta_2)} \tilde{\mathbf{F}}(\beta_1 - \beta_2)$ for some function $\tilde{\mathbf{F}}(\beta)$ that is holomorphic on the strip $0 < \Im(\beta) < 2\pi$, bounded at infinity by $C \cosh(\ell\beta)$, and such that $\tilde{\mathbf{F}}_{-}(\beta + 2\mathrm{i}\pi) = \mathbf{S}(\beta)\tilde{\mathbf{F}}_{+}(-\beta) = \tilde{\mathbf{F}}_{+}(\beta)$, $\beta \in \mathbb{R}$. One first looks for a particular solution to this scalar Riemann–Hilbert problem, namely a holomorphic function $\mathbf{F}$ in the strip $0 < \Im(\beta) < 2\pi$ which behaves as $\mathbf{F}(\beta) = 1 + \mathrm{O}(\beta^{-2})$ as $\Re(\beta) \to \pm\infty$ uniformly in $0 \leq \Im(\beta) \leq 2\pi$ and satisfies $\mathbf{F}_{-}(\beta + 2\mathrm{i}\pi) = \mathbf{S}(-\beta)\mathbf{F}_{+}(\beta) = \mathbf{F}_{+}(-\beta)$, $\beta \in \mathbb{R}$.

Starting from the below integral representation

$$\mathbf{S}(\beta) = \exp\left\{ 8 \int_{0}^{+\infty} \mathrm{d}x \, \frac{\sinh(x\mathfrak{b}) \cdot \sinh(x\hat{\mathfrak{b}}) \cdot \sinh(\frac{1}{2}x)}{x \sinh(x)} \sinh\left(\frac{x\beta}{\mathrm{i}\pi} \right) \right\}, \tag{2.3}$$

one readily checks that the solution is provided by the below $2\mathrm{i}\pi$-periodic Cauchy transform

$$\mathbf{F}(\beta) = \exp\left\{ \int_{\mathbb{R}} \frac{\mathrm{d}s}{4\mathrm{i}\pi} \coth\left[\frac{1}{2}(s - \beta) \right] \ln \mathbf{S}(s) \right\}. \tag{2.4}$$

The s integral can then be taken by means of the integral representation (2.3) for $\ln \mathbf{S}(s)$ and leads to (2.2). Now it is easy to check that the holomorphic function $\mathbf{G}(\beta) = \tilde{\mathbf{F}}(\beta)/\mathbf{F}(\beta)$ on the strip $0 < \Im(\beta) < 2\pi$ admits ± boundary values and satisfies $\mathbf{G}_{-}(\beta + 2\mathrm{i}\pi) = \mathbf{G}_{+}(\beta)$ for $\beta \in \mathbb{R}$ and is bounded by $C \cosh(\ell\Re(\beta))$ as $\Re(\beta) \to \infty$ in this strip. As a consequence, it admits a unique extension into a $2\mathrm{i}\pi$-periodic entire function bounded by $C \cosh(\ell\beta)$ and hence is of the form $P_{\ell}(\mathrm{e}^{\beta})$, where P_{ℓ} is a Laurent polynomial of maximal positive and negative degree ℓ. Since it is $2\mathrm{i}\pi$-periodic and even, $P_{\ell}(\mathrm{e}^{\beta})$ necessarily takes the form

$$P_{\ell}\big(\mathrm{e}^{\beta}\big) = \prod_{a=1}^{k} \left\{ \sinh\left[\frac{\beta - \varkappa_a}{2} \right] \cdot \sinh\left[\frac{\beta + \varkappa_a}{2} \right] \right\} \quad \text{for some } 2k \leq \ell. \tag{2.5}$$

■

2.2. The *n*-particle sector solution

Proposition 2.2. *Consider the change of unknown functions*

$$\mathcal{F}_{n}^{(\mathsf{o})}(\boldsymbol{\beta}_n) = \prod_{a<b}^{n} \mathbf{F}(\beta_{ab}) \cdot \mathcal{K}_{n}^{(\mathsf{o})}(\boldsymbol{\beta}_n) \quad \textit{with } \beta_{ab} = \beta_a - \beta_b, \tag{2.6}$$

with $\mathbf{F}$ as defined through (2.2). *Then $\mathcal{F}_{n}^{(\mathsf{o})}$ solves the bootstrap axioms* (i)–(iv) *if any only if*

(I) $\mathcal{K}_{n}^{(\mathsf{o})}$ *is a symmetric function of $\boldsymbol{\beta}_n$;*

(II) $\mathcal{K}_{n}^{(\mathsf{o})}$ *is a $2\mathrm{i}\pi$ periodic and meromorphic function of each variable taken singly;*

(III) *the only poles of $\mathcal{K}_{n}^{(\mathsf{o})}$ are simple and located at $\beta_a - \beta_b \in \mathrm{i}\pi(1 + 2\mathbb{Z})$. The associated residues are given by*

$$\begin{aligned} &\operatorname{Res}\big(\mathcal{K}_{n}^{(\mathsf{o})}(\boldsymbol{\beta}_n) \cdot \mathrm{d}\beta_1, \beta_{12} = \mathrm{i}\pi\big) \\ &\quad = \frac{\mathrm{i}}{\mathbf{F}(\mathrm{i}\pi)} \cdot \frac{1 - \prod_{a=3}^{n} \mathbf{S}(\beta_{2a})}{\prod_{a=3}^{n} \{\mathbf{F}(\beta_{2a} + \mathrm{i}\pi)\mathbf{F}(\beta_{2a})\}} \cdot \mathcal{K}_{n-2}^{(\mathsf{o})}\big(\boldsymbol{\beta}_n''\big) \end{aligned} \tag{2.7}$$

where $\boldsymbol{\beta}_n'' = (\beta_3, \ldots, \beta_n)$;

(IV) $\mathcal{K}_{n}^{(\mathsf{o})}(\boldsymbol{\beta}_n + \theta\overline{\boldsymbol{e}}_n) = \mathrm{e}^{\theta s_{\mathsf{o}}} \cdot \mathcal{K}_{n}^{(\mathsf{o})}(\boldsymbol{\beta}_n)$.

This first transformation simplifies the symmetry properties of the problem. However, the inductive reductions provided by the computation of the residues are still quite intricate. The idea is then to proceed to yet another change of unknown function, this time by means of a more involved transform. The latter will then lead to structurally much simpler, and thus easier to solve, equations satisfied by the new unknown function. As already mentioned, the dawn of this approach goes back to [10,27] and it was put in the present form in [2–4]. In particular, we refer to [3] for the proof.

Proposition 2.3 ([3]). *Let $\boldsymbol{\ell}_n \in \{0,1\}^n$ and $p_n^{(o)}(\boldsymbol{\beta}_n \mid \boldsymbol{\ell}_n)$ be a solution to the below constraints:*

(a) *$\boldsymbol{\beta}_n \mapsto p_n^{(o)}(\boldsymbol{\beta}_n \mid \boldsymbol{\ell}_n)$ is a collection of $2\mathrm{i}\pi$-periodic holomorphic functions on $\mathbb{C}$ that are symmetric in the two sets of variables jointly, viz. for any $\sigma \in \mathfrak{S}_n$ it holds $p_n^{(o)}(\boldsymbol{\beta}_n^{\sigma} \mid \boldsymbol{\ell}_n^{\sigma}) = p_n^{(o)}(\boldsymbol{\beta}_n \mid \boldsymbol{\ell}_n)$ with $\boldsymbol{\beta}_n^{\sigma} = (\beta_{\sigma(1)}, \dots, \beta_{\sigma(n)})$;*

(b) *$p_n^{(o)}(\beta_2 + \mathrm{i}\pi, \boldsymbol{\beta}'_n \mid \boldsymbol{\ell}_n) = g(\ell_1,\ell_2)\, p_{n-2}^{(o)}(\boldsymbol{\beta}''_n \mid \boldsymbol{\ell}''_n) + h(\ell_1,\ell_2 \mid \boldsymbol{\beta}'_n)$ where h does not depend on the remaining set of variables $\boldsymbol{\ell}''_n$ and*

$$g(0,1) = g(1,0) = \frac{-1}{\sin(2\pi\mathfrak{b})\boldsymbol{F}(\mathrm{i}\pi)}; \tag{2.8}$$

(c) *$p_n^{(o)}(\boldsymbol{\beta}_n + \theta\overline{\boldsymbol{e}}_n \mid \boldsymbol{\ell}_n) = \mathrm{e}^{\theta s_o} \cdot p_n^{(o)}(\boldsymbol{\beta}_n \mid \boldsymbol{\ell}_n)$.*

Then, its $\mathcal{K}$-transform

$$\mathcal{K}_n\big[p_n^{(o)}\big](\boldsymbol{\beta}_n) = \sum_{\boldsymbol{\ell}_n \in \{0,1\}^n} (-1)^{\overline{\boldsymbol{\ell}}_n} \prod_{a<b}^{n} \bigg\{1 - \mathrm{i}\frac{\ell_{ab} \cdot \sin[2\pi\mathfrak{b}]}{\sinh(\beta_{ab})}\bigg\} \cdot p_n^{(o)}(\boldsymbol{\beta}_n \mid \boldsymbol{\ell}_n), \tag{2.9}$$

in which $\overline{\boldsymbol{\ell}}_n = \sum_{a=1}^{n} \ell_k$, solves (I)–(IV).

Note that arguments were given in [15] in favor of some form of bijection between certain classes of solutions to (a)–(c) and (I)–(IV). However, we do stress that, so far, the question whether there does exist a clear cut correspondence between all solutions to (a)–(c) and (I)–(IV) is still open.

3. TOWARDS PHYSICAL OBSERVABLES AND THE CONVERGENCE PROBLEM

The resolution of the bootstrap program provides one with the expressions for the integral kernels of certain operators which are candidates for the quantum fields of the 1+1-dimensional Sinh-Gordon quantum field theory. However, for this construction to really provide one with the quantum field theory of interest, one should establish several facts. First of all, the operators so constructed should form an algebra in the sense discussed in Section 1.2.1. By virtue of the translational invariance (1.9), this means that, for any $n, m \in \mathbb{N}$, the series of multiple integrals arising in the operator product $\mathtt{U}_{\mathtt{T}_{x}}^{-1}\mathtt{P}_n\mathtt{O}_1(\boldsymbol{x})\mathtt{O}_2(\mathbf{0})\mathtt{P}_m$, where $\mathtt{P}_k$ is the orthogonal projector on the k-particle Fock space, should converge in the weak sense. Namely, for any sufficiently regular functions $\boldsymbol{\alpha}_n \mapsto f^{(n)}(\boldsymbol{\alpha}_n)$ and $\boldsymbol{\beta}_m \mapsto g^{(m)}(\boldsymbol{\beta}_m)$ belonging respectively to a dense subset of $L^2(\mathbb{R}^n_>)$ and $L^2(\mathbb{R}^m_>)$, and for any $d \in \mathcal{C}^\infty_{\mathrm{c}}(\mathbb{R}^2)$,

$$\sum_{\ell \geq 0} \int_{\mathbb{R}^\ell_>} \frac{\mathrm{d}^\ell \gamma}{(2\pi)^\ell} \left\{ \int_{\mathbb{R}^n_>} \frac{\mathrm{d}^n \alpha}{(2\pi)^n} f^{(n)}(\boldsymbol{\alpha}_n) \mathcal{M}^{(O_1)}_{n;\ell}(\boldsymbol{\alpha}_n; \boldsymbol{\gamma}_\ell) \right\} \\ \times \left\{ \int_{\mathbb{R}^2} \mathrm{d}^2 \boldsymbol{x}\, d(\boldsymbol{x}) \prod_{a=1}^{\ell} \mathrm{e}^{-\mathrm{i}\boldsymbol{p}(\gamma_a)\cdot \boldsymbol{x}} \right\} \cdot \left\{ \int_{\mathbb{R}^m_>} \frac{\mathrm{d}^n \beta}{(2\pi)^m} \mathcal{M}^{(O_2)}_{\ell;m}(\boldsymbol{\gamma}_\ell; \boldsymbol{\beta}_m) g^{(m)}(\boldsymbol{\beta}_m) \right\} \tag{3.1}$$

should converge. The simplest case corresponds to establishing the convergence of the series of multiple integrals subordinate to operator products $\mathtt{P}_0 \mathtt{O}_1(\boldsymbol{x}) \mathtt{O}_2'(\mathbf{0}) \mathtt{P}_0$, viz. for $n = m = 0$. Since the zeroth Fock space is one-dimensional, this exactly amounts to the convergence of the series of multiple integrals which represents the two-point generalized function $\langle \mathtt{O}_1(\boldsymbol{x}) \mathtt{O}_2'(\mathbf{0}) \rangle$. However, even for this specific instance, proving this property on rigorous grounds remained an open problem for a very long time. It has only recently been solved by the author [23] in the case of space-like separation between the operators, viz. $\boldsymbol{x}^2 < 0$. The scheme of proof of this result will be discussed in Section 4. From the proof's structure, it is rather clear that one can build on minor modifications of this method so as to establish convergence in the time-like regime, i.e. when $\boldsymbol{x}^2 > 0$, although this has not been done yet. Moreover, the combinatorial expressions for the kernels $\mathcal{M}^{(O)}_{n;m}(\boldsymbol{\alpha}_n; \boldsymbol{\beta}_m)$ in terms of the base form factors $\mathcal{F}^{(O)}_p$, $0 \leq p \leq m + n$ indicates that the method outlined below would also allow one to tackle the convergence problem for general multipoint correlation functions.

Once that the convergence problem is solved in full generality, hence guaranteeing that the operators $\mathtt{O}_\alpha(\boldsymbol{x})$ do form an algebra, one still needs to establish the local commutativity property of the quantum fields which ensures causality of the theory. The method for doing so is now well established. Indeed, under the hypothesis of convergence of the handled series of multiple integrals issuing from the operators products, Kirillov and Smirnov showed this property in the Sine-Gordon case in [20, 21]. Their method readily applies to the Sinh-Gordon case. Hence, convergence is the only remaining problem so as to set this construction of quantum field theories on rigorous grounds.

3.1. The well-poised series expansion for two-point functions

First of all, by translation invariance, it is enough to focus on $\langle \mathtt{O}_1(\boldsymbol{x}) \mathtt{O}_2(\mathbf{0}) \rangle$. Recall that, at least in principle, this quantity is a generalised function and should thus be understood, in the first place, as the formal integral kernel of the distribution $\langle \mathtt{O}_1 \mathtt{O}_2 \rangle$. For $d \in \mathcal{C}^\infty_{\mathrm{c}}(\mathbb{R}^2)$, provided convergence holds, one has

$$\langle \mathtt{O}_1 \mathtt{O}_2 \rangle[d] = \int_{\mathbb{R}^2} \mathrm{d}^2 \boldsymbol{x}\, d(\boldsymbol{x}) \big\langle \mathtt{O}_1(\boldsymbol{x}) \mathtt{O}_2(\mathbf{0}) \big\rangle = \sum_{n \geq 0} \frac{1}{n!} \mathcal{I}^{(\mathtt{O}_1,\mathtt{O}_2)}_n[d] \tag{3.2}$$

with

$$\mathcal{I}^{(\mathtt{O}_1,\mathtt{O}_2)}_n[d] = \int_{\mathbb{R}^n} \frac{\mathrm{d}^n \beta}{(2\pi)^n} \mathcal{F}^{(\mathtt{O}_1)}_n(\boldsymbol{\beta}_n) \mathcal{M}^{(\mathtt{O}_2)}_{n;0}(\boldsymbol{\beta}_n; \emptyset) \int_{\mathbb{R}^2} \mathrm{d}^2 \boldsymbol{x}\, d(\boldsymbol{x}) \prod_{a=1}^{n} \big\{ \mathrm{e}^{-\mathrm{i}m[t\cosh(\beta_a) - x\sinh(\beta_a)]} \big\}.$$

It is a direct consequence of the kernel reduction axiom (v) and of Lorentz invariance (iv) that

$$\mathcal{M}_{n;0}^{(\mathsf{O}_2)}(\boldsymbol{\beta}_n;\emptyset)=\mathcal{F}_n^{(\mathsf{O}_2)}(\overleftarrow{\boldsymbol{\beta}}_n+\mathrm{i}\pi\overline{\boldsymbol{e}}_n)=\mathrm{e}^{\mathrm{i}\pi\mathbf{s}_{\mathsf{O}_2}}\mathcal{F}_n^{(\mathsf{O}_2)}(\overleftarrow{\boldsymbol{\beta}}_n). \tag{3.3}$$

This identity, along with the growth bounds in each β_a of the form factors $\mathcal{F}_n^{(\mathsf{O})}(\boldsymbol{\beta}_n)$, ensures the well definiteness of the n-fold integrals since the space-time integral over $\boldsymbol{x}$ produces a decay in each β_a that is faster than any exponential $\mathrm{e}^{\pm k\beta_a}$, $\Re(\beta_a)\to\pm\infty$. By virtue of the Morera theorem, this rapid decay at infinity along with the holomorphy properties of the integrands allow one to deform, *simultaneously* for each integration variable β_a, $a=1,\dots,n$, the integration curves to $\mathbb{R}+\mathrm{i}\frac{\pi}{2}\mathrm{sgn}(x)$ when $\boldsymbol{x}$ is space-like and, when $\boldsymbol{x}$ is time-like, to $\gamma(\mathbb{R})$ where $\gamma(u)=u+\mathrm{i}\vartheta(u)$, where ϑ is smooth, $|\vartheta|<\pi/4$, and such that there exists $M>0$ large enough and $0<\varepsilon<\pi/2$ so that $\vartheta(u)=-\mathrm{sgn}(t)\mathrm{sgn}(u)\varepsilon$ when $|u|\geq M$. This operation turns the $\boldsymbol{\beta}_n$ integrals into absolutely convergent ones irrespectively of the presence of $d(\boldsymbol{x})$. In particular, for the space-like regime, one gets that

$$\mathcal{J}_n^{(\mathsf{O}_1,\mathsf{O}_2)}=\mathrm{e}^{\eta(\boldsymbol{x})}\int_{\mathbb{R}^2}\mathrm{d}^2\boldsymbol{x}\,d(\boldsymbol{x})\int_{\mathbb{R}^n}\frac{\mathrm{d}^n\beta}{(2\pi)^n}\mathcal{F}_n^{(\mathsf{O}_1)}(\boldsymbol{\beta}_n)\mathcal{F}_n^{(\mathsf{O}_2)}(\overleftarrow{\boldsymbol{\beta}}_n)\prod_{a=1}^{n}\mathrm{e}^{-mr\cosh(\beta_a)},$$

in which $r=\sqrt{x^2-t^2}$, $\tanh(\vartheta)=t/x$ while

$$\eta(\boldsymbol{x})=\mathrm{i}\pi\mathbf{s}_{\mathsf{O}_2}+(\mathrm{i}\tfrac{\pi}{2}\mathrm{sgn}(x)+\vartheta)(\mathbf{s}_{\mathsf{O}_1}+\mathbf{s}_{\mathsf{O}_2}).$$

Hence, provided convergence holds, one has the well-defined in the usual sense of numbers representation for the two-point function

$$\langle\mathsf{O}_1(\boldsymbol{x})\mathsf{O}_2(\boldsymbol{0})\rangle=\mathrm{e}^{\eta(\boldsymbol{x})}\sum_{n\geq 0}\frac{1}{n!}\int_{\mathbb{R}^n}\frac{\mathrm{d}^n\beta}{(2\pi)^n}\mathcal{F}_n^{(\mathsf{O}_1)}(\boldsymbol{\beta}_n)\mathcal{F}_n^{(\mathsf{O}_2)}(\overleftarrow{\boldsymbol{\beta}}_n)\prod_{a=1}^{n}\mathrm{e}^{-mr\cosh(\beta_a)}. \tag{3.4}$$

3.2. Convergence of series representation for two-point functions

Thus, the well-definiteness of the two-point functions boils down to providing an appropriate upper bound for the below class of N-fold integrals for $\varkappa>0$,

$$\mathcal{Z}_N(\varkappa)=\int_{\mathbb{R}^N}\mathrm{d}^N\beta\prod_{a\neq b}^{N}\mathrm{e}^{\frac{1}{2}\mathfrak{w}(\beta_{ab})}\cdot\prod_{a=1}^{N}\{\mathrm{e}^{-2\varkappa\cosh(\beta_a)}\}\mathcal{K}_N[p_N^{(\mathsf{O}_1)}](\boldsymbol{\beta}_N)\mathcal{K}_N[p_N^{(\mathsf{O}_2)}](\overleftarrow{\boldsymbol{\beta}}_N). \tag{3.5}$$

The two-body potential $\mathfrak{w}$ is defined through the relation $\mathtt{F}(\lambda)\mathtt{F}(-\lambda)=\mathrm{e}^{\mathfrak{w}(\lambda)}$.

Theorem 3.1 ([23]). *Assume that there exist C_1, C_2, and $k\in\mathbb{N}$ such that given $s\in\{1,2\}$,*

$$\left|p_N^{(\mathcal{O}_s)}(\boldsymbol{\beta}_N\mid\boldsymbol{\ell}_N)\right|\leq C_1^N\cdot\prod_{a=1}^{N}\mathrm{e}^{C_2\beta_a^k}\quad\textit{for any } \boldsymbol{\ell}_N\in\{0,1\}^N, \tag{3.6}$$

uniformly in N. Then, it holds

$$\left|\mathcal{Z}_N(\varkappa)\right|\leq\exp\left[-\frac{3\pi^2\mathfrak{b}\hat{\mathfrak{b}}\cdot N^2}{4\cdot(\ln N)^3}\left\{1+\mathrm{O}\left(\frac{1}{\ln N}\right)\right\}\right]. \tag{3.7}$$

The proof of this theorem was the goal of the author's work [23]. The proof relies on Riemann–Hilbert problem techniques for inverting singular integral operators of truncated-Wiener–Hopf type along with the Deift–Zhou nonlinear steepest descent method [12, 13], concentration of measure, and large deviation techniques which were developed for dealing with certain β-ensembles multiple integrals [7,9,28], and some generalizations thereof to the case of N-dependent integrands in N-dimensional integrals as it was developed in [8].

4. THE PROOF OF THE CONVERGENCE OF THE FORM FACTOR SERIES

In this section we shall describe the main steps of the proof. The details can be found in Proposition 3.1 of [23].

4.1. An simpler upper bound

The starting point consists in obtaining a structurally simpler upper bound on $\mathcal{Z}_N(\varkappa)$ when $\varkappa > 0$.

Proposition 4.1. *There exists $C > 0$ such that*

$$\big|\mathcal{Z}_N(\varkappa)\big| \le (C \cdot \ln N)^N \cdot \max_{p\in[\![0;N]\!]} |\mathscr{Z}_{N,p}(\varkappa)|, \tag{4.1}$$

where $\mathscr{Z}_{N,p}(\varkappa) = \int_{\mathbb{R}^{N-p}} \mathrm{d}^{N-p}\lambda \int_{\mathbb{R}^p} \mathrm{d}^p \nu \overline{\varrho}_{N,p}(\boldsymbol{\lambda}_{N-p}, \boldsymbol{\nu}_p)$ whose integrand is expressed as

$$\begin{aligned}\overline{\varrho}_{N,p}(\boldsymbol{\lambda}_{N-p}, \boldsymbol{\nu}_p) = & \prod_{a=1}^{p}\big\{\mathrm{e}^{-V_N(\nu_a)}\big\} \cdot \prod_{a=1}^{N-p}\big\{\mathrm{e}^{-V_N(\lambda_a)}\big\} \\ & \times \prod_{a<b}^{p}\big\{\mathrm{e}^{\mathfrak{w}_N(\nu_{ab})}\big\} \cdot \prod_{a<b}^{N-p}\big\{\mathrm{e}^{\mathfrak{w}_N(\lambda_{ab})}\big\} \cdot \prod_{a=1}^{p}\prod_{b=1}^{N-p}\big\{\mathrm{e}^{\mathfrak{w}_{\mathrm{tot};N}(\nu_a-\lambda_b)}\big\}. \end{aligned} \tag{4.2}$$

Above, we have used the N-dependent functions

$$V_N(\lambda) = \varkappa \cosh(\tau_N \lambda), \quad \mathfrak{w}_N(\lambda) = \mathfrak{w}(\tau_N \lambda), \quad \mathfrak{w}_{\mathrm{tot};N}(\lambda) = \mathfrak{w}_{\mathrm{tot}}(\tau_N \lambda), \tag{4.3}$$

with $\tau_N = \ln N$ and

$$\mathfrak{w}_{\mathrm{tot}}(\lambda) = \mathfrak{w}(\lambda) + \mathfrak{v}_{2\pi\mathfrak{b},0^+}(\lambda) \quad \textit{with } \mathfrak{v}_{\alpha,\eta}(\lambda) = \ln\bigg(\frac{\sinh(\lambda + \mathrm{i}\alpha)\sinh(\lambda - \mathrm{i}\alpha)}{\sinh(\lambda + \mathrm{i}\eta)\sinh(\lambda - \mathrm{i}\eta)}\bigg). \tag{4.4}$$

4.2. Energetic bounds

Proposition 4.2. *The partition function $\mathscr{Z}_{N,p}(\varkappa)$ admits the upper bound*

$$\mathscr{Z}_{N,p}(\varkappa) \le \exp\big\{-N^2 \inf\big\{\mathcal{E}_{N,\frac{p}{N}}[\mu,\nu] : (\mu,\nu) \in \mathcal{M}^1(\mathbb{R}) \times \mathcal{M}^1(\mathbb{R})\big\} + \mathrm{O}\big(N\tau_N^2\big)\big\}, \tag{4.5}$$

in which the control is uniform in $p \in [\![0; N]\!]$, and where

$$\begin{aligned}\mathcal{E}_{N,t}[\mu,\nu] = & \frac{1}{N}\bigg\{t\int V_N(s)\mathrm{d}\nu(s) + (1-t)\int V_N(s)\mathrm{d}\mu(s)\bigg\} \\ & - \frac{t^2}{2}\int \mathfrak{w}_N(s-u)\mathrm{d}\nu(s)\mathrm{d}\nu(u) - \frac{(1-t)^2}{2}\int \mathfrak{w}_N(s-u)\mathrm{d}\mu(s)\mathrm{d}\mu(u) \\ & - t(1-t)\int \mathfrak{w}_{\mathrm{tot};N}(s-u)\mathrm{d}\mu(s)\mathrm{d}\nu(u).\end{aligned}$$

One may obtain such an upper bound within the standard approach to establishing large deviation bounds for N-fold integrals as pioneered in [7], adjoined to the local regularization of the empirical distribution of the integration variables proposed in [28], and some fine bounds due to the N-dependence of the integrand which were also considered in [8]. The details can be found in Lemmata 4.2–4.3 of [23].

4.3. Characterization of the minimizer and a lower-bound minimizer

The upper bound established in Proposition 4.2 does not allow one to conclude directly on the convergence of the series. Indeed, even if one could prove that the infimum in (4.5) gives a strictly positive number, the N-dependence of the energy functional could make the infimum N-dependent and, in principle, the latter could give rise to a behaviour in N which, when multiplied by the N^2 prefactor, could turn out to be subdominant with respect to the corrections $\mathrm{O}(N\tau_N^2)$. Hence, the longest part of the proof is devoted to obtaining some sharp and explicit lower bound for the infimum which can then be computed in closed form so that one may explicitly check that the above scenario does hold.

For that purpose, one starts by showing

Proposition 4.3. *For $0 < t < 1$ $\mathcal{E}_{N,t}$ admits a unique minimizer $(\mu_{\mathrm{eq}}^{(N,t)}, \nu_{\mathrm{eq}}^{(N,t)})$ on $\mathcal{M}^1(\mathbb{R}) \times \mathcal{M}^1(\mathbb{R})$. Similarly, $\mathcal{E}_{N,0}$ and $\mathcal{E}_{N,1}$ admit unique minimizers on $\mathcal{M}^1(\mathbb{R})$.*

This is established by showing that, for $0 < t < 1$, $\mathcal{E}_{N,t}$ is lower semicontinuous and strictly convex on $\mathcal{M}^1(\mathbb{R}) \times \mathcal{M}^1(\mathbb{R})$, has compact level sets, is not identically $+\infty$, and is bounded from below. In principle, this result could be already enough to obtain sharp in N estimates for $\mathcal{E}_{N,t}[\mu_{\mathrm{eq}}^{(N,t)}, \nu_{\mathrm{eq}}^{(N,t)}]$. Indeed, by relying on the analogous to the case of β-ensembles variational characterization of the minimizers and showing that these are actually Lebesgue continuous with compact connected supports, one may establish a system of two-coupled singular linear integral equations of truncated Wiener–Hopf type depending on the large-parameter N. These may be analyzed within the method developed by Krein's school after generalizing the work [29] and solving the 4×4 associated Riemann–Hilbert problem in the large-N regime by the Deift–Zhou nonlinear steepest descent method [12,13]. However, these steps would definitely lead to an extremely cumbersome and long clamber, especially taken the minimal amount of information one needs, in the end, from such handlings. Therefore, it is more convenient to reduce the numbers of minimizers which ought to be thoroughly determined by providing a lower bound for $\mathcal{E}_{N,t}[\mu_{\mathrm{eq}}^{(N,t)}, \nu_{\mathrm{eq}}^{(N,t)}]$ whose estimation would demand less effort while still leading to the desired result.

A direct calculation shows that one has a simpler representation for $\mathcal{E}_{N,t}$ in terms of functionals only acting on one copy of a space of bounded measures:

$$\mathcal{E}_{N,t}[\mu, \nu] = \sum_{\upsilon=\pm} \mathcal{E}_N^{(\upsilon)}\big[\sigma_t^{(\upsilon)}\big] \quad \text{with } \sigma_t^{(\pm)} = t\nu \pm (1-t)\mu, \tag{4.6}$$

in which $\mathcal{E}_N^{(+)}$ is a functional on $\mathcal{M}^1(\mathbb{R})$ while $\mathcal{E}_N^{(-)}$ is a functional on $\mathcal{M}_{\mathfrak{s}}^{(2t-1)}(\mathbb{R})$, the space of signed, bounded, measures on $\mathbb{R}$ of total mass $2t-1$. These take the form

$$\mathcal{E}_N^{(+)}[\sigma] = \frac{1}{N}\int V_N(s)\,\mathrm{d}\sigma(s) - \frac{1}{2}\int \mathfrak{w}_N^{(+)}(s-u)\cdot \mathrm{d}\sigma(s)\,\mathrm{d}\sigma(u), \tag{4.7}$$

$$\mathcal{E}_N^{(-)}[\sigma] = -\frac{1}{2}\int \mathfrak{w}_N^{(-)}(s-t)\cdot \mathrm{d}\sigma(s)\,\mathrm{d}\sigma(u). \tag{4.8}$$

The two-body interactions appearing above involve $\mathfrak{w}$ and $\mathfrak{v}_{\alpha,\eta}$ introduced in (3.5) and (4.4)

$$\mathfrak{w}_N^{(\pm)}(u) = \mathfrak{w}^{(\pm)}(\tau_N u) \quad \text{with} \quad \begin{cases} \mathfrak{w}^{(+)}(u) = \mathfrak{w}(u) + \frac{1}{2}\mathfrak{v}_{2\pi\mathfrak{b},0^+}(u), \\ \mathfrak{w}^{(-)}(u) = -\frac{1}{2}\mathfrak{v}_{2\pi\mathfrak{b},0^+}(u). \end{cases} \tag{4.9}$$

By going to Fourier space, one observes that

$$\mathcal{E}_N^{(-)}[\sigma] = \frac{1}{2}\int \mathrm{d}\lambda \big|\mathcal{F}[\sigma](\lambda)\big|^2 \frac{\sinh(\pi \mathfrak{b}\lambda)\cdot\sinh(\pi\hat{\mathfrak{b}}\lambda)}{\lambda\sinh(\frac{\pi}{2}\lambda)} \geq 0, \tag{4.10}$$

where $\mathcal{F}[\sigma](\lambda)$ stands for the Fourier transform of the signed measure σ. Thus,

$$\mathcal{E}_{N,t}\big[\mu_{\mathrm{eq}}^{(N,t)}, \nu_{\mathrm{eq}}^{(N,t)}\big] \geq \mathcal{E}_N^{(+)}\big[t\nu_{\mathrm{eq}}^{(N,t)} + (1-t)\mu_{\mathrm{eq}}^{(N,t)}\big] \geq \mathcal{E}_N^{(+)}\big[\sigma_{\mathrm{eq}}^{(N)}\big]. \tag{4.11}$$

In the last line, we have used that $\mathcal{E}_N^{(+)}$ is lower-continuous, has compact level sets, is strictly convex on $\mathcal{M}^1(\mathbb{R})$, bounded from below, and not identically $+\infty$, so as to ensure the existence of a unique minimizer thereof: $\mathcal{E}_N^{(+)}[\sigma_{\mathrm{eq}}^{(N)}] = \inf\{\mathcal{E}_N^{(+)}[\sigma] : \sigma \in \mathcal{M}^1(\mathbb{R})\}$.

4.4. Singular integral equation characterization of the minimizer $\sigma_{\mathrm{eq}}^{(N)}$

By using the variational characterization of the minimizer, see e.g., [11] for an exposition in the β-ensemble case, one reduces the construction of $\sigma_{\mathrm{eq}}^{(N)}$ to finding a solution to a singular integral equation on the Sobolev space $H_s([a_N; b_N])$ driven by the operator

$$\mathcal{S}_N[\phi](\xi) = \fint_{a_N}^{b_N} \big(\mathfrak{w}^{(+)}\big)'\big[\tau_N(\xi-\eta)\big]\cdot\phi(\eta)\mathrm{d}\eta. \tag{4.12}$$

Indeed, upon introducing the effective potential subordinate to a function $\phi \in H_s([a_N; b_N])$,

$$V_{N;\mathrm{eff}}[\phi](\xi) = \frac{1}{N}V_N(\xi) - \fint_{a_N}^{b_N} \mathfrak{w}^{(+)}\big[\tau_N(\xi-\eta)\big]\cdot\phi(\eta)\mathrm{d}\eta, \tag{4.13}$$

one may formulate

Proposition 4.4. *Let $a_N < b_N$ and $\varrho_{\mathrm{eq}}^{(N)} \in H_s([a_N; b_N])$, $1/2 < s < 1$, solve*

$$\frac{1}{N\tau_N}V_N'(x) = \mathcal{S}_N\big[\varrho_{\mathrm{eq}}^{(N)}\big](x) \quad \text{on }]a_N; b_N[, \tag{4.14}$$

be subject to the conditions

$$\varrho_{\mathrm{eq}}^{(N)}(\xi) \geq 0 \quad \text{for } \xi \in [a_N; b_N], \qquad \int_{a_N}^{b_N} \varrho_{\mathrm{eq}}^{(N)}(\xi)\mathrm{d}\xi = 1, \tag{4.15}$$

and

$$V_{N;\mathrm{eff}}\big[\varrho_{\mathrm{eq}}^{(N)}\big](\xi) > \inf\big\{V_{N;\mathrm{eff}}\big[\varrho_{\mathrm{eq}}^{(N)}\big](\eta) : \eta \in \mathbb{R}\big\} \quad \text{for any } \xi \in \mathbb{R}\setminus[a_N; b_N]. \tag{4.16}$$

Then, the equilibrium measure $\sigma_{\mathrm{eq}}^{(N)}$ is supported on the segment $[a_N; b_N]$ and continuous in respect to Lebesgue's measure with density $\varrho_{\mathrm{eq}}^{(N)}$. Moreover, the density takes the form

$$\varrho_{\mathrm{eq}}^{(N)}(\xi) = \sqrt{(b_N-\xi)(\xi-a_N)}\cdot h_N(\xi) \quad \text{with } h_N \in \mathcal{C}^\infty\big([a_N; b_N]\big). \tag{4.17}$$

The above proposition thus provides one with the following strategy for determining the equilibrium measure. One starts by solving the singular integral equation (4.14) for *any* endpoints a_N and b_N. The inversion should be carried out in an appropriate functional space which is dictated by the local structure (4.17) of the equilibrium measure's density, as can be inferred from an analysis of the systems of loop equations associated with the probability measure on $\mathbb{R}^N$ naturally subordinate to the energy functional $\mathcal{E}_N^{(+)}$. The fact that $\mathcal{S}_N$ should be inverted on $H_s([a_N; b_N])$, $0 < s < 1$, imposes a constraint on a_N and b_N. A second constraint is obtained from the fact that the equilibrium measure has unit mass (4.15). This is still not enough so as to be sure that the solution constructed in this way provides one with the equilibrium measure. For that to happen, one still needs to verify that the two positivity constraints (4.15)–(4.16) are fulfilled. The realization of such a program demands to have a thorough control on the inversion of $\mathcal{S}_N$. The latter may be reached within the scheme developed in [29], by solving an auxiliary 2×2 Riemann–Hilbert problem.

4.5. The Riemann–Hilbert based inversion of the operator

In the following, we adopt the shorthand notations

$$\overline{a}_N = \tau_N a_N, \quad \overline{b}_N = \tau_N b_N, \quad \overline{x}_N = \tau_N (b_N - a_N). \tag{4.18}$$

Consider the Riemann–Hilbert problem for a 2×2 matrix function $\chi \in \mathcal{M}_2(\mathcal{O}(\mathbb{C} \setminus \mathbb{R}))$:

- χ has continuous $\pm$-boundary values on $\mathbb{R}$;
- there exist constant matrices $\chi^{(a)}$ with $\chi^{(1)}_{12} \neq 0$ such that when $\lambda \to \infty$,

$$\chi(\lambda) = \begin{cases} \mathcal{P}_{L;\uparrow}(\lambda) \cdot \begin{pmatrix} -\mathfrak{s}_\lambda \cdot \mathrm{e}^{\mathrm{i}\lambda \overline{x}_N} & 1 \\ -1 & 0 \end{pmatrix} \cdot \dfrac{(-\mathrm{i}\lambda)^{\frac{3}{2}\sigma_3}}{\mathrm{e}^{-\mathrm{i}\frac{3\pi}{2}\sigma_3}} \\ \qquad \times \left(I_2 + \frac{\chi^{(1)}}{\lambda} + \frac{\chi^{(2)}}{\lambda^2} + \mathrm{O}(\lambda^{-3})\right) \cdot \mathsf{Q}(\lambda), & \lambda \in \mathbb{H}^+, \\ \mathcal{P}_{L;\downarrow}(\lambda) \cdot \begin{pmatrix} -1 & \mathfrak{s}_\lambda \cdot \mathrm{e}^{-\mathrm{i}\lambda \overline{x}_N} \\ 0 & 1 \end{pmatrix} \cdot (\mathrm{i}\lambda)^{\frac{3}{2}\sigma_3} \\ \qquad \times \left(I_2 + \frac{\chi^{(1)}}{\lambda} + \frac{\chi^{(2)}}{\lambda^2} + \mathrm{O}(\lambda^{-3})\right) \cdot \mathsf{Q}(\lambda), & \lambda \in \mathbb{H}^-, \end{cases}$$

  in which the matrix Q takes the form

$$\mathsf{Q}(\lambda) = \begin{pmatrix} 0 & -\chi^{(1)}_{12} \\ \{\chi^{(1)}_{12}\}^{-1} & \mathfrak{q}_1 + \lambda \end{pmatrix} \quad \text{with } \mathfrak{q}_1 = (\chi^{(1)}_{11}\chi^{(1)}_{12} - \chi^{(2)}_{12}) \cdot \{\chi^{(1)}_{12}\}^{-1};$$

- $\chi_+(\lambda) = G_\chi(\lambda) \cdot \chi_-(\lambda)$ for $\lambda \in \mathbb{R}$ where

$$G_\chi(\lambda) = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\lambda\overline{x}_N} & 0 \\ \frac{1}{\mathrm{i}\pi} \cdot R(\lambda) & -\mathrm{e}^{-\mathrm{i}\lambda\overline{x}_N} \end{pmatrix}$$

$$\text{with } R(\lambda) = 2\frac{\sinh(\pi\mathfrak{b}\lambda) \cdot \sinh(\pi\hat{\mathfrak{b}}\lambda) \cdot \sinh(\frac{\pi}{2}\lambda)}{\cosh^2(\frac{\pi}{2}\lambda)}.$$

Here $\mathfrak{s}_\lambda = \operatorname{sgn}(\Re\lambda)$, $\mathcal{O}(A)$ stands for the ring of holomorphic functions on A, while the O remainder appearing in matrix equalities should be understood entrywise. Moreover, we point out that the matrix Q appearing in the asymptotic expansion for χ is chosen such that χ has the large-λ behavior

$$\chi(\lambda) = \chi^{(\infty)}_{\uparrow/\downarrow}(\lambda) \cdot (\mp\mathrm{i}\lambda)^{\frac{1}{2}\sigma_3}, \quad \lambda \in \mathbb{H}^\pm, \tag{4.19}$$

with $\chi^{(\infty)}_{\uparrow/\downarrow}(\lambda)$ bounded at ∞.

The Deift–Zhou nonlinear steepest descent method [12, 13] allows one to reduce the above Riemann–Hilbert problem into one that is uniquely solvable by the singular integral equation method of [6], provided that N is large enough and $b_N - a_N > c > 0$ uniformly in N.

The solution χ then provides one with a full description of the inverse of $\mathcal{S}_N$.

Proposition 4.5. *Let $0 < s < 1$. The operator $\mathcal{S}_N : H_s([a_N; b_N]) \to H_s(\mathbb{R})$ is continuous and invertible on its image:*

$$\mathfrak{X}_s(\mathbb{R}) = \left\{ H \in H_s(\mathbb{R}) : \int_{\mathbb{R}+\mathrm{i}\varepsilon'} \chi_{12}(\mu)\mathcal{F}[H](\tau_N\mu)\mathrm{e}^{-\mathrm{i}\mu\overline{b}_N} \cdot \frac{\mathrm{d}\mu}{(2\mathrm{i}\pi)^2} = 0 \right\}. \tag{4.20}$$

More specifically, one has the left and right inverse relations

$$\mathcal{W}_N \circ \mathcal{S}_N = \mathrm{id} \quad \text{on } H_s\big([a_N; b_N]\big) \quad \text{and} \quad \mathcal{S}_N \circ \mathcal{W}_N[H](\xi) = H(\xi) \quad \text{a.e. on } [a_N; b_N]$$

for any $H \in \mathfrak{X}_s(\mathbb{R})$. The operator $\mathcal{W}_N : \mathfrak{X}_s(\mathbb{R}) \to H_s([a_N; b_N])$ is given, whenever it makes sense, as an encased oscillatorily convergent Riemann integral transform

$$\mathcal{W}_N[H](\xi) = \frac{\tau_N^2}{\pi} \int_{\mathbb{R}+2\mathrm{i}\varepsilon'} \frac{\mathrm{d}\lambda}{2\mathrm{i}\pi} \int_{\mathbb{R}+\mathrm{i}\varepsilon'} \frac{\mathrm{d}\mu}{2\mathrm{i}\pi} \mathrm{e}^{-\mathrm{i}\tau_N\lambda(\xi-a_N)} W(\lambda,\mu)\mathrm{e}^{-\mathrm{i}\mu\overline{b}_N}\mathcal{F}[H](\tau_N\mu), \tag{4.21}$$

where $\varepsilon' > 0$ is small enough. The integral kernel

$$W(\lambda,\mu) = \frac{1}{\mu-\lambda}\left\{ \frac{\mu}{\lambda} \cdot \chi_{11}(\lambda)\chi_{12}(\mu) - \chi_{11}(\mu)\chi_{12}(\lambda) \right\} \tag{4.22}$$

is expressed in terms of the entries of the matrix χ.

These pieces of information, along with the explicit, uniform on $\mathbb{C}$, large-N expansion of the solution χ to the above Riemann–Hilbert problem and several technical estimates which allow one to check that (4.15)–(4.16) hold, allow one to formulate

Theorem 4.6. *Let $N \geq N_0$ with N_0 large enough. Then the unique minimizer $\sigma_{\mathrm{eq}}^{(N)}$ of the functional $\mathcal{E}_N^{(+)}$ introduced in (4.7) is absolutely continuous in respect to the Lebesgue measure with density $\varrho_{\mathrm{eq}}^{(N)}$ and is supported on the segment $[a_N; b_N]$. The endpoints are the unique solutions to the equations*

$$a_N + b_N = 0 \quad \text{and} \quad \vartheta \cdot \frac{(\overline{b}_N)^2\mathrm{e}^{\overline{b}_N}}{N} \cdot \mathfrak{t}(2\overline{b}_N) \cdot \left\{1 + \mathrm{O}\big((\overline{b}_N)^5\mathrm{e}^{-2\overline{b}_N(1-\varepsilon)}\big)\right\} = 1,$$

for any $1 > \varepsilon > 0$, and the remainder is smooth and differentiable in $\overline{b}_N$. Above, one has

$$\vartheta = \frac{2\varkappa}{3(2\pi)^{\frac{5}{2}}} \cdot \frac{\Gamma(\mathfrak{b},\hat{\mathfrak{b}})}{\mathfrak{b}^{\mathfrak{b}}\hat{\mathfrak{b}}^{\hat{\mathfrak{b}}}},$$

while, upon using the constants w_k introduced below in (4.24),

$$\mathfrak{t}(\overline{x}_N) = \frac{6}{(\overline{x}_N)^2}\left\{2 + w_2 - w_1 - \frac{w_1 w_3}{w_2}\right\} \underset{\overline{x}_N \to +\infty}{\sim} 1 + \mathrm{O}\left(\frac{1}{\overline{x}_N}\right). \tag{4.23}$$

In particular, $\overline{b}_N$ is uniformly away from zero and admits the large-N expansion

$$\overline{b}_N = \ln N - 2\ln\ln N - \ln\vartheta + \mathrm{O}\left(\frac{\ln\ln N}{\ln N}\right).$$

Finally, the density $\varrho_{\mathrm{eq}}^{(N)}$ of the equilibrium measure is expressed in terms of the integral transform of the potential $\varrho_{\mathrm{eq}}^{(N)} = \mathcal{W}_N[V_N']/(N\tau_N)$.

In the statement of the theorem, we made use of the coefficients w_k arising in the $\lambda \to 0$ expansion below

$$2\mathrm{i}\frac{\mathfrak{b}^{2\mathrm{i}\mathfrak{b}\lambda}\hat{\mathfrak{b}}^{2\mathrm{i}\hat{\mathfrak{b}}\lambda}2^{\mathrm{i}\lambda}}{\lambda^3\mathfrak{b}\hat{\mathfrak{b}}\mathrm{e}^{\mathrm{i}\lambda\overline{x}_N}}\Gamma^2\begin{pmatrix}\frac{1}{2}+\mathrm{i}\frac{\lambda}{2}\\ \frac{1}{2}-\mathrm{i}\frac{\lambda}{2}\end{pmatrix}\Gamma\begin{pmatrix}1-\mathrm{i}\mathfrak{b}\lambda, 1-\mathrm{i}\hat{\mathfrak{b}}\lambda, 1-\mathrm{i}\frac{\lambda}{2}\\ \mathrm{i}\mathfrak{b}\lambda, \mathrm{i}\hat{\mathfrak{b}}\lambda, \mathrm{i}\frac{\lambda}{2}\end{pmatrix} = \sum_{\ell=0}^{3}\frac{(-\mathrm{i})^{\ell}w_{\ell}}{\lambda^{3-\ell}} + \mathrm{O}(\lambda). \tag{4.24}$$

4.6. Estimation of the minimum

The closed-form expression for $\sigma_{\mathrm{eq}}^{(N)}$ in terms of the solution χ to the above Riemann–Hilbert problem and the close relation between the two-body interaction in the potential and the $\mathcal{S}_N$ operator's kernel allow one to exploit the system of jumps for χ so as to recast $\mathcal{E}_N^{(+)}[\sigma_{\mathrm{eq}}^{(N)}]$ only in terms of N, $\overline{b}_N$, and χ evaluated at special points:

$$\begin{aligned}\mathcal{E}_N^{(+)}\big[\sigma_{\mathrm{eq}}^{(N)}\big] = \frac{\varkappa}{2N}\cosh(\overline{b}_N) + \frac{\varkappa^2\mathrm{e}^{2\overline{b}_N}}{8\pi N^2}\{\chi_{12}^2(\mathrm{i}) + 2\big[\chi_{12}(\mathrm{i})\chi_{11}'(\mathrm{i}) - \chi_{11}(\mathrm{i})\chi_{12}'(\mathrm{i})\big]\}\\ - \frac{\varkappa\mathrm{e}^{\overline{b}_N}}{4N}\{1+\mathrm{e}^{-\overline{x}_N} + \chi_{22;-}(0)\big[2\chi_{11}(\mathrm{i}) + \mathrm{i}\chi_{12}(\mathrm{i})\big] - 2\chi_{21;-}(0)\chi_{12}(\mathrm{i})\}.\end{aligned}$$

Once that one arrives to the above closed expression, it is a matter of direct calculations which build on the uniform on $\mathbb{C}$ large-N asymptotic expansion for χ provided by the nonlinear steepest descent so as to infer the large-N asymptotics

Proposition 4.7. *One has the large-N asymptotic behavior*

$$\mathcal{E}_N^{(+)}\big[\sigma_{\mathrm{eq}}^{(N)}\big] = \frac{3\pi^4\mathfrak{b}\hat{\mathfrak{b}}\tilde{w}_1}{4(\overline{b}_N)^3\tilde{w}_2\mathfrak{t}(2\overline{b}_N)} + \frac{9\pi^4\mathfrak{b}\hat{\mathfrak{b}}}{8(\overline{b}_N)^4\mathfrak{t}^2(2\overline{b}_N)}\bigg\{1 - \frac{2\tilde{w}_1}{\overline{b}_N\tilde{w}_2}\bigg\} + \mathrm{O}\big(\mathrm{e}^{-2\overline{b}_N(1-\varepsilon)}\big), \tag{4.25}$$

where $\mathfrak{t}$ is as introduced in Theorem 4.6 and we have rescaled the w_k variables:

$$w_1 = 2\overline{b}_N\tilde{w}_1, \quad w_2 = 2(\overline{b}_N)^2\tilde{w}_2, \quad \textit{with } \tilde{w}_k = 1 + \mathrm{O}\bigg(\frac{1}{\overline{b}_N}\bigg) \quad \textit{as } N \to +\infty. \tag{4.26}$$

Together with Propositions 4.2–4.3 and the lower bound in (4.11), the above theorem yields Theorem 3.1.

5. CONCLUSION

In this paper we reviewed the bootstrap program approach to the rigorous construction of 1+1-dimensional integrable quantum field theories arising as appropriate quantizations of integrable classical evolution equations of 1+1-dimensional field theory. This was done on the example of the Sinh-Gordon quantum field theory which is the simplest and nontrivial instance of such model. The approach starts by proposing an appropriate Hilbert space on which such a model is realized. Then, it produces the form of the $\mathbf{S}$-matrix which governs the scattering in such a case. This $\mathbf{S}$-matrix arises as a solution of certain symmetry constrains on the scattering in a relativistically-invariant theory along with the requirement of the factorizability of scattering into a concatenation of two-particle processes. Then, the quantum fields, which are operator-valued distributions on functions of the space-time variables, are constructed as integral operators whose integral kernels satisfy a set of equations, the bootstrap program axioms (i)–(v), which should be taken as the basic axioms of the theory. These axiomatic equations strongly depend on the form of the **S**-matrix for the given theory. It turns out that the bootstrap program equations can be solved explicitly with the help of the algebraic setting provided by the quantum integrability of the model and, in particular, the Yang–Baxter equation satisfied by the **S**-matrix. Once one ends up with the set of explicit solutions to (i)–(v), it remains to check the consistency of the whole construction, in particular, that the so-constructed quantum fields do form an algebra and that they commute at space-like separations. The latter requirement is crucial for guaranteeing the causality of the so-constructed theory and thus it being viable as a per se quantum field theory. To check these last steps of the construction, one must show that the series of multiple integrals resulting from the integral operator's multiplications do converge. This was a long standing open question in this field and its solution [23], in the simplest case scenario, was discussed by the author in the last section of this paper.

There are still numerous open questions related to these topic: first of all, to implement the method of [23], for establishing the convergence of form factor expansions for the time-like separated two-point functions as well as the multipoint correlation functions in all possible regimes of separation between the operators. These questions definitely seem to be manageable within a finite time. Further, one would like to extend the methods of proving the convergence to more challenging but also more physically relevant models such as the 1+1-dimensional integrable Sine-Gordon quantum field theory. There, the multitude of asymptotic particles, along with the presence of bound states and equal mass asymptotic particles, will definitely be a challenging, but hopefully surmountable task.

Last but not least, one should provide a thorough description of the correlation functions in the infrared limit, viz. when the Minkowski separation between the operator approaches zero. In the case of the two-point function given in (3.4) that would correspond to extracting the $r \to 0^+$ limit.

ACKNOWLEDGMENTS

We thank M. Karowski, M. Lashkevich, and F. Smirnov for stimulating discussions on various aspects of integrable quantum field theories.

FUNDING

This work was supported by CNRS. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1928930 while the author participated in a program hosted by the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2021 semester.

REFERENCES

[1] A. E. Arinshtein, V. A. Fateev, and A. B. Zamolodchikov, Quantum S-matrix of the (1+1) dimensional Todd chain. *Phys. Lett. B* **87** (1979), 389–392.

[2] H. Babujian, A. Fring, M. Karowski, and A. Zapletal, Exact form factors in integrable quantum field theories: the sine-Gordon model. *Nuclear Phys. B* **538** (1999), 535–586.

[3] H. Babujian and M. Karowski, Exact form factors in integrable quantum field theories: the sine-Gordon model (II). *Nuclear Phys. B* **620** (2002), 407–455.

[4] H. Babujian and M. Karowski, Sine-Gordon breather form factors and quantum field equations. *J. Phys. A* **35** (2002), 9081–9104.

[5] R. J. Baxter, Partition function of the eight vertex lattice model. *Ann. Phys.* **70** (1972), 193–228.

[6] R. Beals and R. R. Coifman, Scattering and inverse scattering for first order systems. *Comm. Pure Appl. Math.* **37** (1984), 39–90.

[7] G. Ben Arous and A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields* **108** (1997), 517–542.

[8] G. Borot, A. Guionnet, and K. K. Kozlowski, *Asymptotic expansion of a partition function related to the sinh-model*. Math. Phys. Stud., Springer, 2016.

[9] A. Boutet de Monvel, L. Pastur, and M. Shcherbina, On the statistical mechanics approach in the random matrix theory: Integrated density of states. *J. Stat. Phys.* **79** (1995), no. 3–4, 585–611.

[10] V. Brazhnikov and S. Lukyanov, Angular quantization and form factors in massive integrable models. *Nuclear Phys. B* **512** (1998), 616–636.

[11] P. A. Deift, *Orthogonal polynomials and random matrices: a Riemann–Hilbert approach*. Courant Lect. Notes 3, New York University, 1999.

[12] P. A. Deift and X. Zhou, A steepest descent method for oscillatory Riemann–Hilbert problems. *Bull. Amer. Math. Soc.* **26** (1992), no. 1, 119–123.

[13] P. A. Deift and X. Zhou, A steepest descent method for oscillatory Riemann–Hilbert problems. Asymptotics of the mKdV equation. *Ann. of Math.* **137** (1993), 297–370.

[14] R. J. Eden, P. V. Landshoff, D. I. Olive, and J. C. Polkinghorne, *The analytic S matrix*. Cambridge University Press, 1966.

[15] B. Feigin and M. Lashkevich, Form factors of descendant operators: free field construction and reflection relations. *J. Phys. A: Math. Theor.* **42** (2009), 304014.

[16] V. M. Gryanik and S. N. Vergeles, Two-dimensional quantum field theories having exact solutions. *J. Nucl. Phys.* **23** (1976), 1324–1334.

[17] D. Iagolnitzer, *The S matrix*. North Holland Publishing Company, Amsterdam, New York, Oxford, 1978.

[18] M. Karowski and H. J. Thun, Complete S-matrix of the massive Thirring model. *Nuclear Phys. B* **130** (1978), 295–308.

[19] M. Karowski and P. Weisz, Exact form factors in (1+1)-dimensional field theoretic models with soliton behaviour. *Nuclear Phys. B* **139** (1978), 455–476.

[20] A. N. Kirillov and F. A. Smirnov, A representation of the current algebra connected with the SU(2)-invariant Thirring model. *Phys. Rev. B* **198** (1987), 506.

[21] A. N. Kirillov and F. A. Smirnov, Form-factors in the SU(2)-invariant Thirring model. *J. Sov. Math.* **47** (1989), 2423–2450.

[22] V. E. Korepin and L. D. Faddeev, Quantisation of solitons. *Theoret. Math. Phys.* **25** (1975), 1039–1049.

[23] K. K. Kozlowski, On convergence of form factor expansions in the infinite volume quantum Sinh-Gordon model in 1+1 dimensions. 2020, arXiv:2007.01740.

[24] M. Lashkevich and Y. Pugai, Form factors of descendant operators: resonance identities in the sinh-Gordon model. *J. High Energy Phys.* **2014** (2014), 112.

[25] K. S. H. Lehmann and W. Zimmerman, Zür Formulierung quantisierter Feldtheorien. *Nuovo Cimento* **1** (1955), 205–225.

[26] S. Lukyanov, Free field representation for massive integrable models. *Comm. Math. Phys.* **167** (1995), 183–226.

[27] S. Lukyanov, Form-factors of exponential fields in the sine-Gordon model. *Modern Phys. Lett. A* **12** (1997), 2543–2550.

[28] M. Maïda and E. Maurel-Segala, Free transport-entropy inequalities for non-convex potentials and application to concentration for random matrices. *Probab. Theory Related Fields* **159** (2014), no. 1–2, 329–356.

[29] V. Yu. Novokshenov, Convolution equations on a finite segment and factorization of elliptic matrices. *Mat. Zametki* **27** (1980), 449–455.

[30] F. A. Smirnov, Quantum Gelfand–Levitan–Marchenko equations and form factors in the sine-Gordon model. *J. Phys. A: Math. Gen.* **17** (1984), L873–L878.

[31] F. A. Smirnov, Quantum Gelfand–Levitan–Marchenko equations for the sine-Gordon model. *Theoret. Math. Phys.* **60** (1984), 871–880.

[32] F. A. Smirnov, The general formula for solitons form factors in sine-Gordon model. *J. Phys. A* **19** (1986), L575–578.

[33] F. A. Smirnov, Solution of quantum Gel'fand–Levitan–Marchenko equations for the sine-Gordon model in the soliton sector for $\gamma = \pi/\nu$. *Theoret. Math. Phys.* **67** (1986), 344–351.

[34] F. A. Smirnov, *Form factors in completely integrable models of quantum field theory*. Adv. Ser. Math. Phys. 14, World Scientific, 1992.

[35] S. J. Summers, A perspective on constructive quantum field theory. 2016, arXiv:1203.3991.

[36] P. H. Weisz, Exact quantum sine-Gordon soliton form factors. *Phys. Rev. B* **67** (1977), 179.

[37] C. N. Yang, Some exact results for the many-body problem in one dimension with repulsive delta-function interaction. *Phys. Rev. Lett.* **19** (1967), 1312–1315.

[38] A. B. Zamolodchikov and Al. B. Zamolodchikov, Factorized S-matrices in two dimensions as the exact solutions of certain relativistic quantum field theory models. *Ann. Phys.* **120** (1979), 253–291.

[39] Al. B. Zamolodchikov, Exact two-particle S-matrix of quantum sine-Gordon solitons. *Comm. Math. Phys.* **55** (1977), 183–186.

KAROL KAJETAN KOZLOWSKI

Univ Lyon, ENS de Lyon, Univ Claude Bernard Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France, karol.kozlowski@ens-lyon.fr

SINGULARITIES IN GENERAL RELATIVITY

JONATHAN LUK

ABSTRACT

We survey some recent mathematical progress in understanding singularities arising in solutions to the Einstein equations. After some quick discussions of background material, we focus on the following three topics:

- constructions of singular solutions to the Einstein vacuum equations,
- the singularity structure in the interior of generic dynamical black holes and the relation to the strong cosmic censorship conjecture,
- the formation of trapped surfaces, instabilities for the Einstein vacuum equations, and the relation to singularities.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 83-02; Secondary 83C57, 83C75, 35Q75, 35Q76, 35B44

KEYWORDS

General relativity, singularities, cosmic censorship conjectures, black holes, gravitational collapse, weak null singularities

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4120–4141
DOI 10.4171/ICM2022/9

1. INTRODUCTION

We study the Cauchy problem for the celebrated Einstein equations for a (3 + 1)-dimensional Lorentzian manifold $(\mathcal{M}, g)$ with appropriate matter fields:

$$\mathrm{Ric}(g) - \frac{1}{2}S(g)g + \Lambda g = 8\pi T, \tag{1.1}$$

where $\mathrm{Ric}(g)$ and $S(g)$ are, respectively, the Ricci- and scalar-curvature tensors of g, $\Lambda \in \mathbb{R}$ is the cosmological constant, and T is the stress–energy–momentum tensor describing the matter content in $\mathcal{M}$. Equation (1.1) is already highly nontrivial in vacuum, i.e., when $T \equiv 0$, and with vanishing cosmological constant $\Lambda = 0$, in which case (1.1) reduces to

$$\mathrm{Ric}(g) = 0. \tag{1.2}$$

A fascinating feature of solutions to (1.2), or more generally (1.1), is the presence of *singularities*, which can arise even from regular initial data. The most well-known singularities are those occurring at the big bang or in the interior of black holes, though more exotic singularities are known. Viewing (1.1) as a system of partial differential equations, it is desirable to give a complete description of all possible singularities, a goal which at present seems far out of reach.

In this article, we instead survey some recent mathematical progress in the following specific physically interesting settings:

(i) We first discuss some local constructions of different types of singular solutions to (1.2) (Section 2).

(ii) We then turn to the discussion of singularities in the interior of dynamical black holes. This is closely related to the strong cosmic censorship conjecture, stated as Conjecture 1.3 below (Section 3).

(iii) Finally, we discuss how trapped surfaces form dynamically in solutions to (1.1). As we will see below, the formation of trapped surfaces is closely related to black holes and singularities (Section 4).

Before we turn to these topics, we first give some further context regarding singularities in general relativity in the remainder of the introduction.

1.1. The Cauchy problem in general relativity

Any discussion of the Cauchy problem in general relativity begins with the following fundamental theorem (see also the earlier [30]):

Theorem 1.1 (Choquet-Bruhat–Geroch [12]). *Let $(\Sigma, \hat{g})$ be a Riemannian 3-manifold and $\hat{k}$ be a symmetric 2-tensor. Suppose $(\hat{g}, \hat{k})$ are sufficiently regular and satisfy the constraint equations. Then there exists a unique maximal Cauchy development $(\mathcal{M}, g)$ such that*

(1) *the metric g solves (1.2),*

(2) *$(\Sigma, \hat{g}) \hookrightarrow (\mathcal{M}, g)$ isometrically, and $\hat{k}$ is the induced second fundamental form,*

(3) *any other development* $(\mathcal{M}', g')$ *satisfying* (1) *and* (2) *embeds into* $(\mathcal{M}, g)$ *isometrically.*

In general, Theorem 1.1 does not guarantee the maximal Cauchy development to be geodesically complete. Thus, from the point of view of PDE theory, Theorem 1.1 should be viewed as a *local existence* result.

Under suitable *smallness* assumptions, the Choquet-Bruhat–Geroch theorem can be extended to a *global* result. More precisely, if the initial data are close to that of Minkowski spacetime, then the maximal Cauchy development is geodesically complete and converges to Minkowski for large times. This is the monumental *stability of Minkowski* theorem by Christodoulou–Klainerman [19].

In general, however, one must face the possibility of *singularities*. In particular, as we will see, singularities can arise from complete asymptotically flat initial data sets.

1.2. Singularities and black hole spacetimes

The simplest example of formation of singularity for (1.2) can be found in the Schwarzschild solution $(\mathcal{M}_{M,0}, g_{M,0})$, where $M > 0$ is the mass parameter, $\mathcal{M}_{M,0} = \mathbb{R}^2 \times \mathbb{S}^2$, and in a local coordinate system, $g_{M,0}$ is given by

$$g_{M,0} = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2\gamma_{\mathbb{S}^2(1)},$$

where $\gamma_{\mathbb{S}^2(1)}$ denotes the round metric on $\mathbb{S}^2(1)$. The Schwarzschild solution is depicted by the Penrose diagram in Figure 1.

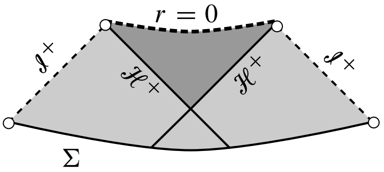


FIGURE 1
Schwarzschild as the maximal future Cauchy development of Σ.

Despite having smooth asymptotically flat initial data, the maximal future Cauchy development of Schwarzschild data has *singularity* inside the black hole region, depicted as the $\{r = 0\}$ surface. What is a singularity? There are a few inequivalent ways to capture the "singular nature" of $\{r = 0\}$ of Schwarzschild:

(i) (Geodesic incompleteness) Any causal geodesic entering the black hole must be incomplete and reach $\{r = 0\}$ in finite time.

(ii) (Blowup of curvature) The curvature invariant $R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu} \to \infty$ as $r \to 0$.

(iii) (Infinitude of tidal deformation) Any observer heading towards the singularity will be infinitely torn apart.

1.3. Trapped surfaces and Penrose's incompleteness theorem

At first, one may hope that the Schwarzschild singularity only arises because Schwarzschild data are very special (e.g., because it is spherically symmetric). This was initially supported by the heuristics of Lifshitz–Khalatnikov [56]: they considered a class of asymptotically Kasner singularities (of which the Schwarzschild singularity is a particular example) and showed that they have one fewer functional degree of freedom compared to the Cauchy problem, which should mean that these singularities are highly nongeneric.

However, in a breakthrough work, Penrose [72] proved that singularities – at least in the sense of geodesic incompleteness – is a stable phenomenon. More precisely, he proved

Theorem 1.2 (Penrose). *If Σ is noncompact, and the maximal Cauchy development $(\mathcal{M}, g)$ contains a compact trapped surface, then $(\mathcal{M}, g)$ is future causally geodesically incomplete.*

Since trappedness is a stable condition, and the Schwarzschild solution contains many compact trapped surfaces, Theorem 1.2 implies that given any sufficiently small perturbations of Schwarzschild data, the corresponding maximal Cauchy development must be future causally geodesically incomplete.

It should be noted that Penrose's fundamental theorem (Theorem 1.2) only asserts the geodesic incompleteness of the spacetime; indeed, one important goal of the subject is to understand when the incompleteness is tied to stronger senses of singularities of curvature or tidal deformation. Already for small perturbations of Schwarzschild data, the geodesic incompleteness can look very different from Schwarzschild! To see this, one needs not look further than the explicit Kerr family of solutions $(\mathcal{M}_{M,a} = \mathbb{R}^2 \times \mathbb{S}^2, g_{M,a})$ for $|a| \leq M$, $M > 0$. When $a = 0$, this reduces to the Schwarzschild subfamily. However, when $0 < |a| < M$, the black hole region terminates with a smooth Cauchy horizon (see Figure 2); in particular, the solution remains completely smooth despite being geodesically incomplete!

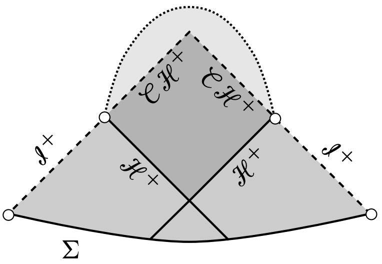


FIGURE 2
Kerr as the maximal future Cauchy development of Σ, with a non-unique extension.

1.4. The cosmic censorship conjectures

The further mathematical study of singularities is guided by two important conjectures of Penrose known as the cosmic censorship conjectures. In a sense, both conjectures assert that some desirable features of the Schwarzschild singularity should be *generic*.

As we discussed above, the interior of the Kerr black hole does not have any singularities. This poses a challenge to the deterministic nature of Einstein's theory as it reflects a breakdown of global uniqueness: the maximal Cauchy development of Kerr data (when $0 < |a| < M$) can be further extended (see Figure 2) – in infinitely many inequivalent ways – as a solution to the Einstein vacuum equations (1.2) beyond the smooth Cauchy horizon.

From this point of view, therefore, the Schwarzschild singularity is preferable to the smooth Kerr Cauchy horizon. Indeed, the first cosmic censorship conjecture asserts that the Schwarzschild case – as opposed to the Kerr case – should be generic.

Conjecture 1.3 (Strong cosmic censorship conjecture [17, 76]). *Maximal Cauchy developments of generic asymptotically flat initial data sets are inextendible as suitably regular Lorentzian manifolds.*

(See also [82, 83] for interesting works on an analogous conjecture for cosmological, i.e., compact, spacetimes. They will not be further discussed here.)

Conjecture 1.3, if true, would resolve the breakdown of determinism. In particular, the smooth Kerr Cauchy horizon would be nongeneric. From the point of view of PDE theory, Conjecture 1.3 can be viewed as a *global uniqueness* conjecture.

At this point, the formulation of Conjecture 1.3 is quite general: in the process of proving the conjecture, one must make precise the notions of "genericity" and "suitable regularity." The regularity class in which the solution is inextendible can be thought of as a convenient way to measure the strength of the singularity. We will refer to "the C^k formulation of Conjecture 1.3" when we mean to impose C^k-inextendibility of the metric. Note that C^2-inextendibility is related to curvature blowup, while C^0-inextendibility can be thought of as a more severe blowup, related to the infinitude of the tidal deformation seen in Schwarzschild. As we will see later (see Section 3), we must carefully distinguish the different formulations in order to capture the precise nature of the singularity in the interior of generic dynamical black holes.

Another preferable feature of the Schwarzschild singularity is that it is hidden behind an event horizon, and thus not visible to far-away observers. A mathematical reformulation of this fact without explicitly referring to the singularities is to say that null infinity of Schwarzschild is complete. In fact, the full Kerr family of black holes, not just the Schwarzschild subfamily, possess a complete null infinity. This is conjectured to be generic:

Conjecture 1.4 (Weak cosmic censorship conjecture [17, 73]). *Maximal Cauchy developments of generic asymptotically flat initial data sets possess a* complete *null infinity.*

Conjecture 1.4 can be viewed as a conjecture on *global existence* in the large; indeed, this is the best notion of global existence one can hope for in view of Theorem 1.2.

2. CONSTRUCTION OF SINGULARITIES

The first step towards understanding singularities in general relativity is to construct specific classes of singular solutions. Explicit singular solutions (including Schwarzschild and Kasner) have, of course, been known for a long time. There are also many results where singularities are constructed using simplifying assumptions of symmetry and analyticity. However, more general constructions of singularities have only been achieved quite recently.

2.1. Spacelike singularities

While perhaps Schwarzschild or Kasner singularities are the simplest to write down, Lifshitz–Khalatnikov (see Section 1.3) argued that such singularities depend only on three functional degrees of freedom (i.e., one fewer than that for the Cauchy problem) and are thus nongeneric. Nonetheless, one can construct the full class of such singular solutions:

Theorem 2.1 (Fournodavlos–Luk [31]). *There exists a class of asymptotically Kasner singular solutions to* (1.2) *parametrized by three functional degrees of freedom.*

See [44,51] and references therein for earlier works with symmetry and/or analyticity assumptions.

The key realization here is that the Einstein vacuum equations are, in fact, locally well-posed in a Gaussian coordinate system, i.e., in a gauge such that

$$g = -dt^2 + {}^{(3)}g_{ij}dx^i dx^j$$

for some Riemannian metric ${}^{(3)}g$, which is realized by considering the wave equation for the second fundamental form and appropriate renormalizations. In this gauge, we can carry out a Fuchsian-type analysis to construct an approximate solution, and then upgrade the construction to a bona fide solution by performing singular energy estimates.

The singularities constructed in Theorem 2.1 are not expected to be stable. Nonetheless, these singularities are stable after *restricting in suitable symmetry class*:

Theorem 2.2 (Alexakis–Fournodavlos [1], Founodavlos–Rodnianski–Speck [32]). *The singularities of Schwarzschild* [1] *and Kasner* [32] *are respectively stable under polarized axisymmetry and polarized* $\mathbb{U}(1)$ *symmetry.*

Note that in these symmetry classes, the Cauchy data depend only on two functional degrees of freedom. On the other hand, [32] treats a much more general case – the so-called subcritical regime – which includes a large class of Kasner singularities (a) in vacuum in high dimensions and (b) with matter fields in $(3 + 1)$ dimensions, without any symmetry assumptions.

It should be remarked that the influential paper [4] suggests that there should be a large class of spacelike singularities which are *oscillatory* (unlike the asymptotically Kasner singularities). Some progress has been made for a class of spatially homogeneous solutions [81]. However, its relevance in the spatially nonhomogeneous setting remains unclear.

2.2. Null singularities

It turns out that the Einstein vacuum equations admit a class of very different singular solutions which are much more stable! In contrast to Section 2.1, the singular hypersurfaces are null in these spacetimes. These solutions were first discovered in the context of the study of strong cosmic censorship conjecture for various matter models; see Section 3 below. Such singularities are often called "weak" null singularities, as the metric can be extended continuously beyond the singularities and the tidal deformation remains finite. However, they should also be thought of as "essential" singularities, since (at least conjecturally) they cannot be extended in $W^{1,p}$ for any $p > 1$.

Theorem 2.3 (Luk [57]). *A class of stable weak null singularities exist for the vacuum equation* (1.2) *without any symmetry assumptions.*

Analytic examples were previously constructed by Ori–Flanagan [71].

Like in the proof of Theorem 2.1 (and Theorem 2.2), the choice of coordinates lies at the heart of the proof. The proof uses a local coordinate system $(u, \underline{u}, \theta^1, \theta^2)$ adapted to a double null foliation, i.e., the metric takes the form

$$g = -4\Omega^2 du d\underline{u} + \gamma_{AB}(d\theta^A - b^A du)(d\theta^B - b^B du), \tag{2.1}$$

and the singular null hypersurface is a constant-u or constant-$\underline{u}$ hypersurface (or both in the case of a bifurcate null singularity). The Einstein equations in this gauge have remarkable – both linear and nonlinear – structure. First, by introducing appropriately "renormalized" curvature components, one can recast the Einstein equations in the gauge (2.1) as a coupled system of hyperbolic–elliptic–transport system which avoids the most singular (non-L^1) components of curvature. Moreover, the system of equations have important *nonlinear null structure* so that the potentially most dangerous singular terms do not appear.

The proof of Theorem 2.3 was inspired by earlier works [61,62] by Luk–Rodnianski on the propagation and interaction of *impulsive gravitational waves* without symmetry assumptions. These solutions to (1.2), first discovered in symmetry classes (see [43,74]), contain null singularities which are weaker so that a local well-posedness theory still holds. (In this context, note also the more recent work [65,66] which considers the interaction of *three* impulsive gravitational waves, for which one needs geometric constructions beyond (2.1).)

2.3. κ-self-similar singularities and naked singularities

Self-similar singularities play an important role in many evolutionary PDEs. For the Einstein vacuum equations (1.2), Rodnianski–Shlapentokh-Rothman recently constructed a class of (what they called) κ-self-similar singularities. In fact, the singularities they constructed are *naked singularities*, i.e., spacetimes with incomplete future null infinities (cf. Conjecture 1.4).

Theorem 2.4 (Rodnianski–Shlapentokh-Rothman [84]). *The Einstein vacuum equations* (1.2) *admit solutions with naked singularities.*

If Conjecture 1.4 is true, then the naked singularities in Theorem 2.4 (and indeed any naked singularities) would be unstable. Nevertheless, Theorem 2.4 shows that in order to resolve Conjecture 1.4, one must come to terms with understanding "genericity."

A closely related construction was previously achieved by Christodoulou for the Einstein–scalar field system in spherical symmetry [15]. There are also numerical evidence of other regimes of (discretely) self-similar singularities [11,53].

3. BLACK HOLE INTERIORS AND THE STRONG COSMIC CENSORSHIP CONJECTURE

We now turn to the singularities that arise in black hole interiors. We have already seen the example of the singularity in a Schwarzschild black hole. We will soon also encounter black hole interiors with null singularities, just like those constructed in Section 2.2. However, unlike in Section 2, our main concern here is not only the *local* structure of the singularities (as in Sections 2.1, 2.2), but instead we are interested in what singularities are *formed* in dynamical evolution inside black holes.

In particular, we will be interested in the question of strong cosmic censorship (see Section 1.4), i.e., whether black hole interiors are indeed generically singular as in the Schwarzschild case.

3.1. Spherically symmetric model problems

The first results concerning the issues of black hole interiors and strong cosmic censorship were obtained under the assumption of spherical symmetry. The spherical symmetry assumption rules out the Kerr solution; nevertheless, if one couples the Einstein equations with a Maxwell field, the two-parameter family (parametrized by the mass and charge M, Q) of the Reissner–Nordström solution (when $0 < |Q| < M$) also has a Penrose diagram given by Figure 2. In particular, these solutions have a smooth global bifurcate Cauchy horizon which can be extended nonuniquely as solutions to the Einstein–Maxwell system.

The early breakthroughs [37,78,79] concerned the Einstein–Maxwell–null dust system in spherical symmetry. In these works of Hiscock, Poisson–Israel, it was already shown that both stability and instability aspects are present: the perturbed solution still has a Cauchy horizon, and the metric remains continuous up to the Cauchy horizon; it is only the higher derivatives, for instance, the Hawking mass, that blow up.

A more satisfactory spherically symmetry model, which involves a wave-type dynamical degree of freedom, is the Einstein–Maxwell–scalar field system:

$$\begin{aligned}\mathrm{Ric}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R &= 2\big(T^{(\mathrm{sf})}_{\mu\nu} + T^{(\mathrm{em})}_{\mu\nu}\big),\\ T^{(\mathrm{sf})}_{\mu\nu} &= \partial_\mu\phi\partial_\nu\phi - \frac{1}{2} g_{\mu\nu}(g^{-1})^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi,\\ T^{(\mathrm{em})}_{\mu\nu} &= (g^{-1})^{\alpha\beta} F_{\mu\alpha}F_{\nu\beta} - \frac{1}{4} g_{\mu\nu}(g^{-1})^{\alpha\beta}(g^{-1})^{\gamma\sigma} F_{\alpha\gamma}F_{\beta\sigma},\end{aligned} \tag{3.1}$$

where ϕ is a real-valued scalar function and F is a 2-form satisfying

$$\Box_g \phi = 0, \quad dF = 0, \quad \nabla_\nu F^{\mu\nu} = 0. \tag{3.2}$$

It turns out that spherical symmetry breaks the supercriticality of the problem, and, in fact, one can study the structure of the black hole interior for large data (i.e., not only those which are small perturbations of Reissner–Nordström). For this model, it was proven that the Cauchy horizon is a generic feature!

Theorem 3.1 (Dafermos [20], Dafermos–Rodnianski [26]). *Given any asymptotically flat, spherically symmetric, admissible data on $\Sigma = \mathbb{R} \times \mathbb{S}^2$, if the initial charge is not identically* 0*, then the solution to* (3.1) *and* (3.2) *satisfies the following:*

(1) *each component of the black hole exterior converges to Reissner–Nordström,*

(2) *the black hole interior has a (null) Cauchy horizon as (at least) part of the boundary,*

(3) *the solution is extendible up to the Cauchy horizon with a continuous metric.*

In particular, the C^0-formulation of the strong cosmic censorship conjecture is false.

In fact, it can be shown [21,52] based on Theorem 3.1 that when the initial charge is nonvanishing, then the solution either has the Penrose diagram of Reissner–Nordström, or else the boundary of the black hole interior has both null and spacelike components, as indicated in Figure 3. Put differently, Theorem 3.1 shows that when the charge is nonvanishing, the black hole interior, at least near timelike infinity i^+, looks more like Reissner–Nordström than Schwarzschild. This phenomenon is due to a subtle interplay between the amplification effect in the black hole interior and the decay in the black hole exterior. On the one hand, the local *blue-shift* effect present at the Reissner–Nordström Cauchy horizon [86] causes an exponential growth of waves. On the other hand, [26] established that waves in the black hole exterior decay with at least an inverse polynomial rate (as predicted by the linear heuristics of Price [80]), by understanding the dispersion of waves in the far-away region and the red-shift effect near the black hole event horizon. This decay competes with the exponential growth induced by the blue-shift effect, resulting in a black hole interior which is still C^0-extendible.

The C^0-extendibility result in Theorem 3.1, however, is not the end of the story. While the solution is extendible for *all* data with nonvanishing charge, the result is consistent with spacetime metrics arising from *generic* data having derivatives that blow up at the

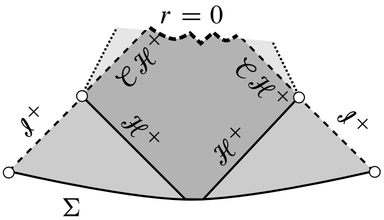


FIGURE 3

A possible Penrose diagram for Theorem 3.1.

Cauchy horizon. In fact, a conditional result was proven in [20], showing that the derivatives of the metric indeed blow up *assuming* some pointwise inverse polynomial lower bound.

More recently, it was proven that in fact the following version of the C^2-formulation of the strong cosmic censorship conjecture holds (unconditionally):

Theorem 3.2 (Luk–Oh [59,60]). *There exists an open and dense subset of the set of initial data in Theorem 3.1 such that the maximal future Cauchy development is future <u>in</u>extendible as time-oriented Lorentzian manifold with a C^2-metric.*

Like Theorem 3.1, the blowup in the interior proven in Theorem 3.2 results from an interplay of the decay in the exterior and the growth in the interior. Indeed, the proof of Theorem 3.2 proceeds by first showing that *generically*, waves in the black hole exterior obey an inverse polynomial *lower bound* (slightly different from that in [20]), and then proving that the solution is C^2-future inextendible whenever such a lower bound holds. In the course of the proof, a condition at null infinity is identified: we define a functional $\mathfrak{L}$, which can be computed only in terms of the radiation field and the Bondi mass at null infinity, such that $\mathfrak{L} \neq 0$ generically, and $\mathfrak{L} \neq 0$ implies the desired inverse polynomial lower bound.

From the point of view of PDE theory, one may even hope that generic solutions are inextendible in $W^{1,2}_{\mathrm{loc}}$, so as to exclude the possibility of any extension as weak solutions to the Einstein equations [18]. The estimates in Theorem 3.2 indeed suggest that this may be true, though such a geometric statement is still unknown. Very recently, Sbierski [85] proved that generic solutions as in Theorem 3.2 are C^1-inextendible.

3.2. C^0-stability of the Kerr Cauchy horizon

While the above results completed the story for (3.1) in the spherically symmetric setting, it should be noted that the strong cosmic censorship conjecture concerns *generic* data. Spherically symmetric data are, of course, by definition far from generic!

In order to make progress towards the strong cosmic censorship conjecture *without any symmetry assumptions*, we investigate a perturbative regime near the Kerr solution. It has been shown that the presence of Cauchy horizons is a generic feature even outside of symmetry!

Theorem 3.3 (Dafermos–Luk [24]). *Consider general vacuum initial data corresponding to the expected induced geometry of a dynamical black hole settling down to Kerr (with parameters $0 < |a| < M$) on a suitable spacelike hypersurface Σ_0 in the black hole interior. Then the maximal future development spacetime $(\mathcal{M}, g)$ corresponding to Σ_0 is globally covered by a double null foliation and has a nontrivial Cauchy horizon* CH^+ *across which the metric is continuously extendible.*

If the Kerr exterior is stable – as is widely expected (see [23,50]) – then Theorem 3.3 in particular implies that any small perturbations of 2-ended Kerr initial data lead to a black hole interior with a Cauchy horizon across which the metric is continuously extendible. In fact, assuming stability of Kerr exterior, it can be proven that for small perturbations of two-ended Kerr data, the maximal future Cauchy development has a global bifurcate Cauchy horizon, as in given by Figure 2 [25].

However, the implications of Theorem 3.3 go beyond small perturbations of Kerr data. Indeed, it is sometimes conjectured – in the so-called final state conjecture [77] – that generic solutions settle down to finitely many Kerr exterior solutions moving away from each other. Moreover, one expects the asymptotically Schwarzschild (or asymptotically extremal; see Section 3.3.3) solutions to occur only for nongeneric data [23]. If this is true, then Theorem 3.3 would in fact apply to generic black hole interior near timelike infinity.

It should be noted that Theorem 3.3 does not indicate whether the Cauchy horizon is actually singular. In fact, Theorem 3.3 is proven as a *stability* theorem. The proof, however, relies only on very weak norms, consistent with the Cauchy horizon possibly being a weak null singularity.

One of the challenges of the proof of Theorem 3.3 is to control solutions to the Einstein vacuum equation using only weak norms consistent with the solution being only – at least when measured in the worst direction – $C^0 \cap W^{1,1}$. This is way below the threshold for well-posedness for the Einstein equations. Instead, the proof relies on the estimates developed in the construction of local weak null singularities without symmetry assumptions (cf. Theorem 2.3). At the same time, Theorem 3.3 requires an understanding of the decay towards timelike infinity in order to close the global problem. In particular, one needs to extend ideas in Theorem 3.1 to a setting without symmetry assumptions.

Even though Theorem 3.3 by itself does not show any blowup, on the basis of Theorem 3.2 and some model linear problems [3, 27, 35, 64], it seems reasonable to expect that with generic data, the Cauchy horizon is a weak null singularity as in Theorem 2.3:

Conjecture 3.4. Generic *small perturbations of two-ended Kerr data lead to a maximal Cauchy development where the global bifurcate Cauchy horizon is a weak null singularity.*

In a similar manner to Theorem 3.2, one expects that the key to Conjecture 3.4 is to understand the precise rates of convergence in the black hole exterior.

3.3. Further problems concerning black hole interiors

While Theorem 3.3 (and Conjecture 3.4) gives the structure of the interior of generic dynamic asymptotically flat black hole near timelike infinity, we survey some other situations here, where the black hole interior is expected to be different. At the moment these are only understood under spherical symmetry or even just in a linear setting.

3.3.1. Breakdown of weak null singularities

For astrophysical gravitational collapse, the initial hypersurface does not have two ends. Instead, a black hole is expected to form from initial data on $\Sigma = \mathbb{R}^3$ (cf. Section 4). These solutions are in particular not globally close to Kerr, though as discussed in Section 3.2, Kerr is still relevant as they may arise as the asymptotic state for the black hole exterior. In this case, Theorem 3.3 still applies to show that the metric is close to Kerr in C^0

near timelike infinity. However, there are regions in the black hole which are far away from Kerr and cannot be treated by perturbative arguments.

To gain some insight, one returns to spherical symmetry: a convenient model to simultaneously study gravitational collapse and the (in)stability of the Cauchy horizons in black hole interior in spherical symmetry is the Einstein–Maxwell–charged scalar field system, i.e., unlike (3.1), the scalar field is complex-valued and charged.

With the above model, Van de Moortel considered that problem where the initial data are posed on $\mathbb{R}^3$. He proved that if a black hole forms and converges to Reissner–Nordström in the exterior with appropriate rates, then the Cauchy horizon in the black hole interior is a weak null singularity as in Theorems 3.1 and 3.2 [89]. Even more interestingly, he also proved that the null boundary in the black hole interior must break down [90]! Conjecturally, this would mean that the singular boundary in the black hole has a null component and a spacelike component. See also the numerical works [7,10].

3.3.2. Other singularities in the presence of matter

Other singularities can occur in the interior of black holes; some of these singularities may be specific to the matter models involved. Some examples include highly oscillatory null singularities arising in the interior of black holes when a charged scalar field oscillates in the black hole exterior [42], and violent nonlinear spacelike singularities in the interior of hairy black holes [34,91].

3.3.3. Extremal black holes

When $|a| = M > 0$ for Kerr or $|Q| = M > 0$ for Reissner–Nordström, these black hole spacetimes are known to be *extremal*. Though their black hole interiors have a somewhat different global structure, they have a smooth Cauchy horizon as in the subextremal case. Unlike their subextremal counterparts, however, the local blue shift at the Cauchy horizon degenerates in the extremal case.

The degeneration of the local blue shift suggests that a dynamical black hole settling down to an extremal black hole may in fact have a Cauchy horizon so that the solution is not only C^0-extendible, but also extendible as a weak solution to the Einstein equations. For a spherically symmetric model, this was studied numerically in [69] and has been later proven by Gajic–Luk [33]. Amusingly, in order to go beyond symmetry, even though the nonlinear theory is expected to be simpler than Theorem 3.3 in view of the weaker singularity, the linear theory appears to be more complicated.

Notice that this phenomenon, by itself, does not pose a threat to strong cosmic censorship, since asymptotically extremal black holes are only expected to arise from a nongeneric set of data!

3.3.4. Nonvanishing cosmological constant

When the cosmological constant $\Lambda \neq 0$ (but still in vacuum), (1.1) admits the Kerr–(anti-)de Sitter black hole solutions, which (when $|a| \neq 0$) like Kerr, admit smooth Cauchy horizons in the interior of the black hole. However, the stability properties of these Cauchy horizons may turn out to be quite different from the $\Lambda = 0$ case in Theorem 3.3 and Conjecture 3.4!

When $\Lambda > 0$, at least when $|a|$ is sufficiently small, the nonlinear stability of Kerr–de Sitter has been established by Hintz–Vasy [36] and (unlike the Kerr case) is no longer a conjecture. Moreover, [36] shows that perturbations of Kerr–de Sitter data lead to solutions that converge *exponentially fast* back to Kerr–de Sitter. Thus the proof of Theorem 3.3 applies, mutatis mutandis, to show that the Kerr–de Sitter Cauchy horizon is C^0-stable.

However, the rapid exponential decay has the possibility to make Conjecture 3.4 false! To understand whether this happens seems to require determining the precise exponential rate of decay. This problem has attracted much heuristic and numerical works; see [6,9,28] and references therein.

When $\Lambda < 0$, the situation is difficult (and interesting) for a different reason: even linear waves on Kerr–anti-de Sitter spacetime decay only logarithmically [39]. Kehle [41] has made some interesting progress for the linear (in)stability of the Cauchy horizon, showing that the stability properties depend on the Diophantine properties of the black hole parameters in a subtle way.

4. GRAVITATIONAL COLLAPSE, FORMATION OF TRAPPED SURFACES, AND THE WEAK COSMIC CENSORSHIP CONJECTURE

As discussed in Section 1.3, Penrose's theorem (Theorem 1.2) shows that geodesic incompleteness is intimately related to the presence of trapped surfaces. In this final section, we discuss how trapped surfaces are formed dynamically from initial data without trapped surfaces. This is particularly relevant in gravitational collapse where black holes form. Finally, in Section 4.4, we will discuss how trapped surface formation relates to the weak cosmic censorship conjecture.

4.1. Formation of trapped surfaces by focussing of gravitational radiation

In the explicit Schwarzschild and Kerr solutions, either a (marginally) trapped surface or an antitrapped surface is present in any initial hypersurface. Physically, one expects that trapped surfaces may form dynamically in gravitational collapse, i.e., they may arise even if the initial hypersurface has trivial topology and is far from having a trapped surface.

Examples of formation of trapped surfaces in the presence of matter have been constructed very early on [70,75]. The problem is much harder for the vacuum equations since any such construction is necessarily large data and (by Birkhoff's theorem) outside spherical symmetry. In a monumental breakthrough, Christodoulou constructed a large set of (stable) solutions in vacuum where trapped surfaces form dynamically from dispersed data via focussing of gravitational radiation.

Theorem 4.1 (Christodoulou [18]). *Consider the characteristic initial value problem with data on two intersecting null hypersurfaces H_0 and $\underline{H}_0$ such that*

(1) *the data on $\underline{H}_0$ is that of an incoming cone in Minkowski space;*

(2) *the data on H_0 is given in a region where $0 \leq \underline{u} \leq \delta$ and the initial shear $\hat{\chi}$ obeys the upper bound*

$$\sum_{i+j \leq 10} \delta^{\frac{1}{2}} \left\| \nabla_4^i \nabla^j \hat{\chi} \right\|_{L^\infty} \leq C,$$

(3) *the initial $\hat{\chi}$ on H_0 obeys the lower bound*

$$\inf_{\vartheta \in S^2} \int_0^\delta |\hat{\chi}|^2(\underline{u}', \vartheta)\, d\underline{u}' \geq c > 0.$$

Then, for $\delta > 0$ sufficiently small, a trapped surface forms in the causal domain of the data.

The significance of the breakthrough work of Christodoulou goes beyond the trapped surface formation problem, as it is also the first large data long time result regarding the dynamics of the Einstein vacuum equation without any symmetry assumptions. What allowed Christodoulou to handle a large data regime was a novel idea of "short pulse": the incoming radiation is concentrated in a region with a short length scale δ, so that despite the largeness, the nonlinear structure of the equations allowed Christodoulou to propagate a hierarchy of large and small estimates quantified by δ and to close all the estimates.

As was observed later in [63], as $\delta \to 0^+$, the spacetimes constructed by Christodoulou limit to a spacetime in which a null dust shell (i.e., the null dust is a delta measure on a null hypersurface) collapses and trapped surfaces form (thus the limit metric solves the Einstein equations with matter, even though for each $\delta > 0$ the spacetime is vacuum). In other words, after understanding that solutions to the Einstein–null dust system can, in fact, arise as limits of vacuum solutions [8, 40], the Christodoulou construction can be conceptually thought of as an approximation of the trapped surface formation examples with matter in [75, 87].

There are many subsequent simplifications and extensions of Theorem 4.1; see, for instance, [48, 54, 55]. We record two results that strengthen Theorem 4.1. The first improvement allows the focussing to occur only in some (as opposed to all) directions:

Theorem 4.2 (Klainerman–Luk–Rodnianski [46]). *Suppose* (1) *and* (2) *of Theorem* 4.1 *hold, and the* inf *in* (3) *is replaced by a* sup, *i.e.,*

$$\sup_{\vartheta \in S^2} \int_0^\delta |\hat{\chi}|^2(\underline{u}', \vartheta)\, d\underline{u}' \geq c > 0,$$

then a trapped surface forms in the causal domain of the data.

Theorem 4.2 is achieved by combining the existence theorem in [18] with a deformation argument, which identifies a trapped surface by solving an elliptic inequality.

The second improvement allows the incoming radiation to be much weaker: on the one hand, the incoming radiation is only required to be large in a scale-invariant norm; on the other hand, in some situations the required lower bound can be much smaller than the upper bound:

Theorem 4.3 (An–Luk [2]). *Assume* (1) *of Theorem* 4.1 *and replace* (2) *and* (3) *by*

(2) *the data on H_0 is given in a region where $0 \leq \underline{u} \leq \delta$ and the initial $\hat{\chi}$ obeys the upper bound*
$$\sum_{i+j\leq 10} \delta^{\frac{1}{2}} \left\| \nabla_4^i \nabla^j \hat{\chi} \right\|_{L^\infty} \leq a^{\frac{1}{2}},$$

(3) *the initial $\hat{\chi}$ on H_0 obeys the lower bound*
$$\inf_{\vartheta\in S^2} |\hat{\chi}|^2(\underline{u}', \vartheta)\, d\underline{u}' \geq 4ba^{\frac{1}{2}}\delta,$$

where $b \leq a$, $\delta a^{\frac{1}{2}} b < 1$, and $b \geq b_0$ for some universal large constant b_0. Then there exists a trapped surface in the causal domain of the data.

The scale-invariant results in Theorem 4.3 are proven using weighted estimates capturing the precise growth rate of the geometric quantities close to the vertex of $\underline{H}_0$.

4.2. Instability of anti-de Sitter spacetime

In Theorem 4.1, Christodoulou arranged gravitational waves to focus so that the nonlinear effect on the geometry causes a trapped surface to form dynamically. In spectacular recent works, Moschidis has demonstrated – albeit only in a spherically symmetric setting – a new trapped surface formation mechanism that goes beyond mere focussing of waves: he showed that nonlinear interaction of waves can enhance the focussing effect, which finally leads to trapped surface formation.

Moschidis' work is in the context of the AdS stability problem. The anti-de Sitter (AdS) spacetime ($M_{\mathrm{AdS}} = \mathbb{R}^{3+1}, g_{\mathrm{AdS}}$), with g_{AdS} given by
$$g_{\mathrm{AdS}} = -\left(1 - \frac{\Lambda r^2}{3}\right) dt^2 + \left(1 - \frac{\Lambda r^2}{3}\right)^{-1} dr^2 + r^2 \gamma_{\mathbb{S}^2(1)},$$
is a solution to the Einstein vacuum equations with $\Lambda < 0$. Since it is not globally hyperbolic, one needs to impose boundary conditions to study its stability properties. AdS is conjectured [22] to be unstable under reflective boundary conditions, and this has been studied heuristically and numerically [5]. (Nevertheless, it is expected to be stable under maximally dissipative conditions; see [38].)

In a series of remarkable recent works, Moschidis resolved the AdS instability conjecture for various matter models in spherical symmetry, showing that the AdS spacetime is unstable against *trapped surface formation*.

Theorem 4.4 (Moschidis). *There exist arbitrarily small spherically symmetric perturbations of AdS data for*

(1) *the Einstein–null dust system with an inner mirror* [67]

(2) *the Einstein–massless Vlasov system* [68]

(3) *the Einstein–scalar field system* [Moschidis, in preparation]

with reflective boundary conditions such that a trapped surface forms dynamically.

In particular, viewed as a solution to the above Einstein–matter systems with reflective boundary conditions, the AdS solution is unstable.

The proof of Theorem 4.4 constructs small perturbations of AdS consisting of many spherically symmetric matter beams with judiciously chosen widths and amplitudes. The basic underlying instability mechanism is as follows: whenever two such matter beams interact, the energy of the incoming beam is concentrated. Due to the reflective boundary condition, these beams interact many times, so much so that the nonlinear interaction eventually causes a trapped surface to form.

4.3. The bounded L^2-curvature theorem and beyond

Returning to the Einstein vacuum equations, we now know that the Moschidis instability – first established in spherical symmetry for various matter models in the AdS instability problem – can be adapted in the vacuum case *without symmetry assumptions*.

To provide some context for this instability in vacuum, we recall the celebrated bounded L^2-curvature theorem (first conjectured in [45]):

Theorem 4.5 (Klainerman–Rodnianski–Szeftel [49]). *There exists $\epsilon_0 > 0$ such that if the initial data have H^2-norm $\leq \epsilon_0$, then the solution has H^2-norm $O(\epsilon_0)$ up to time $O(1)$.*

As pointed out in [47, 49], H^2 is sharp for estimating the null conjugacy radius, which is an important step in the construction of the parametrix. It turns out that not only the techniques cannot be extended below H^2, but the result itself also cannot be improved:

Theorem 4.6 (Luk–Moschidis, in progress). *There exists $\delta > 0$ such that the following holds for every $s \in [2-\delta, 2)$: For any $\epsilon > 0$, there exist initial data such that the initial data have H^s-norms of size ϵ, but the H^s-norms at time $O(1)$ are $\gtrsim \epsilon^{-1}$.*

Notice that this is not an ill-posedness result in a *fixed gauge*. For instance, it has been previously proven in [29] that the Einstein vacuum equations *in wave coordinates* are ill-posed in H^2. In contrast, in Theorem 4.6 we proved that the H^s-norm becomes large in *any* coordinate system so that the metric remains C^0-close to the Minkowski metric.

Theorem 4.6 is an instability result in a Sobolev space *above scaling*. (The scaling-invariant norm would be $H^{3/2}$.) In particular, a corresponding instability result is false for model problems such as wave maps [88] or for the Einstein–scalar field system in spherical symmetry [14]. The underlying instability mechanism, which is based on quasilinear interaction of gravitational "wave packets," is quasilinear and anisotropic, and is inspired by the Moschidis mechanism used in Theorem 4.4.

Of course, Theorem 4.6 is only an instability result, and it does not say anything about the formation of trapped surfaces per se. We remark, however, that it is not so difficult to show that if smallness is imposed on the H^s norm for $s > \frac{3}{2}$, then the initial hypersurface does not contain any trapped surfaces [58]. Extending the ideas in Theorem 4.6, one may imagine a scenario where the data are small in H^s (for $s \in (\frac{3}{2}, 2)$), but then there is an evolutionary formation of trapped surfaces associated with the growth in the H^s-norm:

Problem 4.7. Is it possible to dynamically form a trapped surface with localized initial data that are small in H^s for some $s \in [\frac{3}{2}, 2)$?

If the answer to Problem 4.7 is positive, then there would be a trapped surface formation mechanism for data even weaker than in Theorems 4.2 and 4.3.

4.4. Weak cosmic censorship conjecture

In [17], Christodoulou proposed a program to tackle the weak cosmic censorship conjecture (Conjecture 1.4). This program in particular relates weak cosmic censorship to the formation of trapped surfaces.

The strategy suggested in [17] is inspired by the spectacular work [16] which resolved the weak cosmic censorship conjecture for the Einstein–scalar field system in spherical symmetry. This latter work is, in fact, so far the only mathematical work which gives us some insights as to why naked singularities should be nongeneric. The strategy in [16] combines two ingredients: (i) a sharp trapped surface formation criterion [13], and (ii) a scale-invariant breakdown criterion [14]. Using (ii), Christodoulou further showed that small perturbations of naked singularities must be blue-shifted so that, using (i), he showed that trapped surfaces must form arbitrarily close by in the perturbed spacetime. As a result, for generic data, first singularities are preceded by trapped surfaces arbitrarily close by. From this, Christodoulou deduced that weak cosmic censorship holds for this model.

To tackle the weak cosmic censorship conjecture in vacuum without symmetry assumptions, Christodoulou introduced the following conjecture:

Conjecture 4.8 (Trapped surface conjecture, Christodoulou [17]). *For generic asymptotically flat vacuum data on Σ, the maximal globally hyperbolic development has the following property: Given a terminal indecomposable past set $\mathcal{P}$, if $\mathcal{P} \cap \Sigma$ has compact closure, then for every $\Sigma \supset U \supset \overline{\mathcal{P} \cap \Sigma}$, the domain of dependence of U contains a closed trapped surface.*

As pointed out in [17], Conjecture 4.8 has the advantage of being formulated locally, without referring to future null infinity (as in Conjecture 1.4).

At present, Conjecture 4.8 is far out of reach. In spherical symmetry, a sharp trapped surface formation result [13] has turned out to play a fundamental role for the analogue of Conjecture 4.8. The analysis outside symmetry is of course fundamentally more difficult, but one may hope that understanding the mechanisms for trapped surface formation (Sections 4.1–4.3) will likewise be relevant for Conjecture 4.8.

ACKNOWLEDGMENTS

I thank my many collaborators for their patience, inspiration, and friendship. I am particularly grateful to Igor Rodnianski, who, in addition, introduced me to this beautiful subject. I also thank Mihalis Dafermos for the figures in this article.

FUNDING

This work was partially supported by a Terman fellowship and the NSF grant DMS-2005435.

REFERENCES

[1] S. Alexakis and G. Fournodavlos, Stable space-like singularity formation for axi-symmetric and polarized near-Schwarzschild black hole interiors. 2020, arXiv:2004.00692.

[2] X. An and J. Luk, Trapped surfaces in vacuum arising dynamically from mild incoming radiation. *Adv. Theor. Math. Phys.* **21** (2017), no. 1, 1–120.

[3] Y. Angelopoulos, S. Aretakis, and D. Gajic, Late-time tails and mode coupling of linear waves on Kerr spacetimes. 2021, arXiv:2102.11884.

[4] V. A. Belinski, E. M. Lifshitz, and I. Khalatnikov, Oscillatory approach to the singular point in relativistic cosmology. *Sov. Phys., Usp.* **13** (1971), no. 6, 745.

[5] P. Bizoń and A. Rostworowski, Weakly turbulent instability of anti-de Sitter spacetime. *Phys. Rev. Lett.* **107** (2011), no. 3, 031102.

[6] P. R. Brady, I. G. Moss, and R. C. Myers, Cosmic censorship: as strong as ever. *Phys. Rev. Lett.* **80** (1998), no. 16, 3432.

[7] P. R. Brady and J. D. Smith, Black hole singularities: a numerical approach. *Phys. Rev. Lett.* **75** (1995), no. 7, 1256–1259.

[8] G. A. Burnett, The high-frequency limit in general relativity. *J. Math. Phys.* **30** (1989), no. 1, 90–96.

[9] V. Cardoso, J. a. L. Costa, K. Destounis, P. Hintz, and A. Jansen, Quasinormal modes and strong cosmic censorship. *Phys. Rev. Lett.* **120** (2018), 031103.

[10] P. M. Chesler, R. Narayan, and E. Curiel, Singularities in Reissner–Nordström black holes. *Classical Quantum Gravity* **37** (2020), no. 2, 025009.

[11] M. W. Choptuik, Universality and scaling in gravitational collapse of a massless scalar field. *Phys. Rev. Lett.* **70** (1993), 9–12.

[12] Y. Choquet-Bruhat and R. P. Geroch, Global aspects of the Cauchy problem in general relativity. *Comm. Math. Phys.* **14** (1969), 329–335.

[13] D. Christodoulou, The formation of black holes and singularities in spherically symmetric gravitational collapse. *Comm. Pure Appl. Math.* **44** (1991), no. 3, 339–373.

[14] D. Christodoulou, Bounded variation solutions of the spherically symmetric Einstein–scalar field equations. *Comm. Pure Appl. Math.* **46** (1993), no. 8, 1131–1220.

[15] D. Christodoulou, Examples of naked singularity formation in the gravitational collapse of a scalar field. *Ann. of Math. (2)* **140** (1994), no. 3, 607–653.

[16] D. Christodoulou, The instability of naked singularities in the gravitational collapse of a scalar field. *Ann. of Math.* **149** (1999), 183–217.

[17] D. Christodoulou, On the global initial value problem and the issue of singularities. *Classical Quantum Gravity* **16** (1999), A23–A35.

[18] D. Christodoulou, *The formation of black holes in general relativity*. EMS Monogr. Math., European Mathematical Society (EMS), Zürich, 2009.

[19] D. Christodoulou and S. Klainerman, *The global nonlinear stability of the Minkowski space*. Princeton Math. Ser. 41, Princeton University Press, Princeton, 1993.

[20] M. Dafermos, The interior of charged black holes and the problem of uniqueness in general relativity. *Comm. Pure Appl. Math.* **58** (2005), no. 4, 445–504.

[21] M. Dafermos, Black holes without spacelike singularities. *Comm. Math. Phys.* **332** (2014), no. 2, 729–757.

[22] M. Dafermos and G. Holzegel, Dynamic instability of solitons in 4 + 1-dimensional gravity with negative cosmological constant. 2006, https://www.dpmms.cam.ac.uk/~md384/ADSinstability.pdf.

[23] M. Dafermos, G. Holzegel, I. Rodnianski, and M. Taylor, The non-linear stability of the Schwarzschild family of black holes. 2021, arXiv:2104.08222.

[24] M. Dafermos and J. Luk, The interior of dynamical vacuum black holes I: the C^0-stability of the Kerr Cauchy horizon. 2017, arXiv:1710.01722.

[25] M. Dafermos and J. Luk, The interior of dynamical vacuum black holes III: the C^0-stability of the bifurcation sphere of the Kerr Cauchy horizon. In preparation.

[26] M. Dafermos and I. Rodnianski, A proof of Price's law for the collapse of self-gravitating scalar field. *Invent. Math.* **162** (2005), 381–457.

[27] M. Dafermos and Y. Shlapentokh-Rothman, Time-translation invariance of scattering maps and blue-shift instabilities on Kerr black hole spacetimes. *Comm. Math. Phys.* **350** (2017), no. 3, 985–1016.

[28] O. J. C. Dias, F. C. Eperon, H. S. Reall, and J. E. Santos, Strong cosmic censorship in de Sitter space. *Phys. Rev. D* **97** (2018), no. 10, 104060.

[29] B. Ettinger and H. Lindblad, A sharp counterexample to local existence of low regularity solutions to Einstein equations in wave coordinates. *Ann. of Math. (2)* **185** (2017), no. 1, 311–330.

[30] Y. Foures-Bruhat, Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires. *Acta Math.* **88** (1952), no. 1, 141–225.

[31] G. Fournodavlos and J. Luk, Asymptotically Kasner-like singularities. 2020, arXiv:2003.13591.

[32] G. Fournodavlos, I. Rodnianski, and J. Speck, Asymptotically Kasner-like singularities. 2020, arXiv:2012.05888.

[33] D. Gajic and J. Luk, The interior of dynamical extremal black holes in spherical symmetry. *Pure Appl. Anal.* **1** (2019), no. 2, 263–326.

[34] S. A. Hartnoll, G. T. Horowitz, J. Kruthoff, and J. E. Santos, Gravitational duals to the grand canonical ensemble abhor Cauchy horizons. *J. High Energy Phys.* **(10):102, 23** (2020).

[35] P. Hintz, A sharp version of Price's law for wave decay on asymptotically flat spacetimes. 2020, arXiv:2004.01664.

[36] P. Hintz and A. Vasy, The global non-linear stability of the Kerr–de Sitter family of black holes. *Acta Math.* **220** (2018), no. 1, 1–206.

[37] W. A. Hiscock, Evolution of the interior of a charged black hole. *Phys. Rev. Lett. A* **83** (1981), 110–112.

[38] G. Holzegel, J. Luk, J. Smulevici, and C. Warnick, Asymptotic properties of linear field equations in anti-de Sitter space. *Comm. Math. Phys.* **374** (2020), no. 2, 1125–1178.

[39] G. Holzegel and J. Smulevici, Quasimodes and a lower bound on the uniform energy decay rate for Kerr–AdS spacetimes. *Anal. PDE* **7** (2014), no. 5, 1057–1090.

[40] C. Huneau and J. Luk, High-frequency backreaction for the Einstein equations under polarized $U(1)$-symmetry. *Duke Math. J.* **167** (2018), no. 18, 3315–3402.

[41] C. Kehle, Diophantine approximation as Cosmic Censor for Kerr–AdS black holes. 2020, arXiv:2007.12614.

[42] C. Kehle and M. Van de Moortel, Strong Cosmic Censorship in the presence of matter: the decisive effect of horizon oscillations on the black hole interior geometry. 2021, arXiv:2105.04604.

[43] K. A. Khan and R. Penrose, Scattering of two impulsive gravitational plane waves. *Nature* **229** (1971), 185–186.

[44] S. Kichenassamy and A. D. Rendall, Analytic description of singularities in Gowdy spacetimes. *Classical Quantum Gravity* **15** (1998), no. 5, 1339–1355.

[45] S. Klainerman, PDE as a unified subject. In *GAFA 2000 number special volume. Part I*, pp. 279–315. 2000 (Tel Aviv, 1999).

[46] S. Klainerman, J. Luk, and I. Rodnianski, A fully anisotropic mechanism for formation of trapped surfaces in vacuum. *Invent. Math.* **198** (2014), no. 1, 1–26.

[47] S. Klainerman and I. Rodnianski, Causal geometry of Einstein–vacuum spacetimes with finite curvature flux. *Invent. Math.* **159** (2005), no. 3, 437–529.

[48] S. Klainerman and I. Rodnianski, On the formation of trapped surfaces. *Acta Math.* **208** (2012), no. 2, 211–333.

[49] S. Klainerman, I. Rodnianski, and J. Szeftel, The bounded L^2 curvature conjecture. *Invent. Math.* **202** (2015), no. 1, 91–216.

[50] S. Klainerman and J. Szeftel, *Global nonlinear stability of Schwarzschild spacetime under polarized perturbations*. Ann. of Math. Stud. 210, Princeton University Press, Princeton, 2020.

[51] P. Klinger, A new class of asymptotically non-chaotic vacuum singularities. *Ann. Physics* **363** (2015), 1–35.

[52] J. Kommemi, The global structure of spherically symmetric charged scalar field spacetimes. *Comm. Math. Phys.* **323** (2013), no. 1, 35–106.

[53] L. Lehner and F. Pretorius, Black strings, low viscosity fluids, and violation of cosmic censorship. *Phys. Rev. Lett.* **105** (2010), no. 10, 101102.

[54] J. Li and J. Liu, Instability of spherical naked singularities of a scalar field under gravitational perturbations. 2017, arXiv:1710.02422.

[55] J. Li and P. Yu, Construction of Cauchy data of vacuum Einstein field equations evolving to black holes. *Ann. of Math. (2)* **181** (2015), no. 2, 699–768.

[56] E. M. Lifshitz and I. M. Khalatnikov, Investigations in relativistic cosmology. *Adv. Phys.* **12** (1963), 185–249.

[57] J. Luk, Weak null singularities in general relativity. *J. Amer. Math. Soc.* **31** (2018), no. 1, 1–63.

[58] J. Luk and G. Moschidis. In preparation.

[59] J. Luk and S.-J. Oh, Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data I. The interior of the black hole region. *Ann. of Math. (2)* **190** (2019), no. 1, 1–111.

[60] J. Luk and S.-J. Oh, Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data II: the exterior of the black hole region. *Ann. PDE* **5** (2019), no. 1, Paper No. 6, 194.

[61] J. Luk and I. Rodnianski, Local propagation of impulsive gravitational waves. *Comm. Pure Appl. Math.* **68** (2015), no. 4, 511–624.

[62] J. Luk and I. Rodnianski, Nonlinear interaction of impulsive gravitational waves for the vacuum Einstein equations. *Camb. J. Math.* **5** (2017), no. 4, 435–570.

[63] J. Luk and I. Rodnianski, High-frequency limits and null dust shell solutions in general relativity. 2020, arXiv:2009.08968.

[64] J. Luk and J. Sbierski, Instability results for the wave equation in the interior of Kerr black holes. *J. Funct. Anal.* **271** (2016), no. 7, 1948–1995.

[65] J. Luk and M. Van de Moortel, Nonlinear interaction of three impulsive gravitational waves I: Main result and the geometric estimates. 2020, arXiv:2101.08353.

[66] J. Luk and M. Van de Moortel, Nonlinear interaction of three impulsive gravitational waves II: The wave estimates. 2020, arXiv:2106.05479.

[67] G. Moschidis, A proof of the instability of AdS for the Einstein–null dust system with an inner mirror. 2017, arXiv:1704.08681.

[68] G. Moschidis, A proof of the instability of AdS for the Einstein–massless Vlasov system. 2018, arXiv:1812.04268.

[69] K. Murata, H. S. Reall, and N. Tanahashi, What happens at the horizon(s) of an extreme black hole? *Classical Quantum Gravity* **30** (2013), no. 23, 235007.

[70] J. R. Oppenheimer and H. Snyder, On continued gravitational contraction. *Phys. Rev.* **56** (1939), no. 5, 455.

[71] A. Ori and E. E. Flanagan, How generic are null spacetime singularities? *Phys. Rev. D* **53** (1996), 1754–1758.

[72] R. Penrose, Gravitational collapse and space-time singularities. *Phys. Rev. Lett.* **14** (1965), 57–59.

[73] R. Penrose, Gravitational collapse: The role of general relativity. *Riv. Nuovo Cim.* **1** (1969), 252–276.

[74] R. Penrose, The geometry of impulsive gravitational waves. In *General relativity (papers in honour of J. L. Synge)*, pp. 101–115, Clarendon Press, Oxford, 1972.

[75] R. Penrose, Naked singularities. *Ann. N.Y. Acad. Sci.* **224** (1973), no. 1, 125–134.

[76] R. Penrose, Gravitational collapse. In *Gravitational radiation and gravitational collapse*, edited by C. Dewitt-Morette, pp. 82–91, Symp., Int. Astron. Union 64, Springer, Berlin, 1974.

[77] R. Penrose, Some unsolved problems in classical general relativity. In *Seminar on differential geometry*, edited by S.-T. Yau, pp. 631–668, Ann. of Math. Stud. 102, Princeton University Press, Princeton, 1982.

[78] E. Poisson and W. Israel, Inner-horizon instability and mass inflation in black holes. *Phys. Rev. Lett.* **63** (1989), 1663–1666.

[79] E. Poisson and W. Israel, Internal structure of black holes. *Phys. Rev. D* **41** (1990), 1796–1809.

[80] R. H. Price, Nonspherical perturbations of relativistic gravitational collapse. I. Scalar and gravitational perturbations. *Phys. Rev. D* **3** (1972), no. 5, 2419–2438.

[81] H. Ringström, The Bianchi IX attractor. *Ann. Henri Poincaré* **2** (2001), no. 3, 405–500.

[82] H. Ringström, *The Cauchy problem in general relativity*. ESI Lect. Math. Phys. European Mathematical Society (EMS), Zürich, 2009.

[83] H. Ringström, Strong cosmic censorship in T^3-Gowdy spacetimes. *Ann. of Math. (2)* **170** (2009), no. 3, 1181–1240.

[84] I. Rodnianski and Y. Shlapentokh-Rothman, Naked singularities for the Einstein vacuum equations: The exterior solution. 2019, arXiv:1912.08478.

[85] J. Sbierski, On holonomy singularities in general relativity and the $C^{0,1}_{\mathrm{loc}}$-inextendibility of spacetimes. 2020, arXiv:2007.12049.

[86] M. Simpson and R. Penrose, Internal instability in a Reissner–Nordström black hole. *Internat. J. Theoret. Phys.* **7** (1973), 183–197.

[87] J. L. Synge, A model in general relativity for the instantaneous transformation of a massive particle into radiation. *Proc. Roy. Irish Acad. Sect. A* **59** (1957), 1–13.

[88] T. Tao, Global regularity of wave maps. II. Small energy in two dimensions. *Comm. Math. Phys.* **224** (2001), no. 2, 443–544.

[89] M. Van de Moortel, Stability and instability of the sub-extremal Reissner–Nordström black hole interior for the Einstein–Maxwell–Klein–Gordon equations in spherical symmetry. *Comm. Math. Phys.* **360** (2018), no. 1, 103–168.

[90] M. Van de Moortel, The breakdown of weak null singularities inside black holes. 2019, arXiv:1912.10890.

[91] M. Van de Moortel, Violent nonlinear collapse in the interior of charged hairy black holes. 2021, arXiv:2109.10932.

JONATHAN LUK

Department of Mathematics, Stanford University, 450 Jane Stanford Way, Stanford,
CA 94305-2125, USA, jluk@stanford.edu

CLASSIFICATION OF GAPPED GROUND STATE PHASES IN QUANTUM SPIN SYSTEMS

YOSHIKO OGATA

ABSTRACT

Recently, classification problems of gapped ground state phases attracted a lot of attention in quantum statistical mechanics. We explain our operator algebraic approach to these problems.

MATHEMATICS SUBJECT CLASSIFICATION 2020

Primary 81R15; Secondary 82B10

KEYWORDS

Quantum spin systems, ground states

INTERNATIONAL CONGRESS
OF MATHEMATICIANS
2022 JULY 6–14

Proc. Int. Cong. Math. 2022, Vol. 5, pp. 4142–4161
DOI 10.4171/ICM2022/29

1. INTRODUCTION

In quantum mechanics, physical models are determined in terms of some self-adjoint operators called Hamiltonians. Recently, Hamiltonians whose spectrum has a gap between the lowest eigenvalue (which coincides with the infimum of the spectrum) and the rest of the spectrum attracted a lot of attention. Physically, these models are considered to be in normal phases, where no critical phenomena occur. Despite that, it has turned out that the structure of these normal gapped phases is actually mathematically interesting when we introduce some equivalence relation to them. Roughly speaking, we say that two models are equivalent if we can connect them smoothly within those normal phases. In spacial dimensions higher than one, it is believed (and partially proven) that there are multiple phases with respect to such classifications. If we further introduce some symmetry to the game, we obtain interesting mathematical structures, even in one dimension. In this paper, we explain the operator-algebraic approach to those problems.

2. FINITE-DIMENSIONAL QUANTUM MECHANICS

In order to motivate us for the operator algebraic framework of quantum statistical mechanics, we first recall finite-dimensional quantum mechanics in this section. In finite-dimensional quantum mechanics, physical observables are represented by elements of M_n, the algebra of $n \times n$-matrices. Each positive matrix ρ with $\operatorname{Tr} \rho = 1$ (called a density matrix) defines a physical state by

$$\omega_\rho : \mathrm{M}_n \ni A \mapsto \operatorname{Tr}(\rho A) \in \mathbb{C}.$$

We call this map ω_ρ a state. Clearly, it is positive, i.e., $\omega_\rho(A^* A) \geq 0$ and normalized $\omega_\rho(\mathbb{I}) = 1$. This corresponds to the procedure of taking expectation values of each physical observables $A \in \mathrm{M}_n$, in the physical state ω_ρ. Note that the set of all states forms a convex compact set. Its extremal points are called pure states. A state ω_ρ is pure if and only if ρ is a rank-one projection.

Time evolution (Heisenberg dynamics) is given by a self-adjoint matrix H, called a Hamiltonian, via the formula

$$\mathrm{M}_n \ni A \mapsto \tau_t(A) := e^{itH} A e^{-itH}, \quad t \in \mathbb{R}. \tag{2.1}$$

Let p be the spectral projection of H corresponding to the lowest eigenvalue. A state $\omega_\rho(A) := \operatorname{Tr} \rho A$ on M_n is said to be a ground state of H if the support of ρ is under p. The ground state is unique if and only if p is a rank one projection, i.e., if the lowest eigenvalue of H is nondegenerate. In this case, the unique ground state is of the form $\omega_p(A) := \operatorname{Tr} pA$, and it is pure because p has rank one.

Sometimes we consider time-dependent Hamiltonians $H(t)$. Then the time evolution of an observable $A \in \mathrm{M}_n$ is given by a solution $\tau_t(A)$ of the differential equation

$$\frac{d}{dt}\tau_t(A) = i\big[H(t), \tau_t(A)\big], \quad \tau_0(A) = A, \quad A \in \mathrm{M}_n.$$

When the Hamiltonian is time-dependent $H(t) = H$, this reduces to the above Heisenberg dynamics $e^{itH} A e^{-itH}$.

Symmetry plays an important role in physics. Let G be a finite group and suppose that there is a group action $\beta : G \to \mathrm{Aut}(\mathrm{M}_n)$ given by unitaries V_g, $g \in G$,

$$\beta_g(A) := \mathrm{Ad}(V_g)(A), \quad A \in \mathrm{M}_n,\ g \in G.$$

Here and thereafter, $\mathrm{Aut}(\mathcal{A})$ for a $*$-algebra $\mathcal{A}$ denotes the automorphism group of $\mathcal{A}$. If a Hamiltonian H satisfies $\beta_g(H) = H$ for all $g \in G$, we say that H is β-invariant. If a β-invariant Hamiltonian H has a unique ground state $\omega_p(A) := \mathrm{Tr}\, pA$, then this unique ground state ω_p is β-invariant $\omega_p(\beta_g(A)) = \omega_p(A)$, $A \in \mathrm{M}_n$, because the spectral projection p is β-invariant, i.e., $\beta_g(p) = p$.

3. QUANTUM SPIN SYSTEMS

Operator-algebraic framework of quantum statistical mechanics allows us to extend the framework of finite-dimensional quantum mechanical systems to infinite dimensions. Let $2 \le d \in \mathbb{N}$ and $\nu \in \mathbb{N}$ be fixed. Physically, $\frac{d-1}{2}$ denotes the size of on-site spin (spin quantum number) and ν denotes the spacial dimension. We denote by $\mathfrak{S}_{\mathbb{Z}^\nu}$ the set of all finite subsets of $\mathbb{Z}^\nu$. To each finite subset $\Lambda \in \mathfrak{S}_{\mathbb{Z}^\nu}$ we associate a finite-dimensional C^*-algebra

$$\mathcal{A}_\Lambda := \bigotimes_\Lambda \mathrm{M}_d.$$

Here, M_d is the algebra of $d \times d$-matrices. The ν-dimensional quantum spin system $\mathcal{A}_{\mathbb{Z}^\nu}$ is the C^*-inductive limit of this inductive net, given by the natural inclusion. For each infinite subset Γ, we may define $\mathcal{A}_\Gamma$ in exactly the same manner. The C^*-algebra $\mathcal{A}_\Gamma$ can be naturally regarded as a C^*-subalgebra of $\mathcal{A}_{\mathbb{Z}^\nu}$. We say that an element A has support in Γ if it belongs to $\mathcal{A}_\Gamma$. If an automorphism α acts trivially on $\mathcal{A}_{\Gamma^c}$ for some $\Gamma \subset \mathbb{Z}^\nu$, we say that α has support in Γ. The set of all elements in $\mathcal{A}_{\mathbb{Z}^\nu}$ with finite support is called a local algebra and denoted by $\mathcal{A}_{\mathrm{loc}}$.

A state ω on $\mathcal{A}_\Gamma$ is defined to be a linear functional on $\mathcal{A}_\Gamma$ with $\omega(\mathbb{I}) = 1$ which is positive in the sense that $\omega(A^*A) \ge 0$ for any $A \in \mathcal{A}_\Gamma$. The map $\mathcal{A}_\Gamma \ni A \mapsto \omega(A) \in \mathbb{C}$ corresponds to the procedure of taking the expectation value of a physical observable A in our physical state ω. The set of all states on $\mathcal{A}_\Gamma$ forms a convex weak$*$-compact set. Its extremal points are called pure states. By the Krein–Milman theorem, the set of states is the weak$*$-closure of the convex envelope of pure states. See [6] for more details.

For each state, we can associate a representation of $\mathcal{A}_\Gamma$ essentially uniquely.

Theorem 3.1 (GNS representation). *For each state ω on $\mathcal{A}_\Gamma$, there exist a representation π_ω of $\mathcal{A}_\Gamma$ on a Hilbert space $\mathcal{H}_\omega$ and a unit vector $\Omega_\omega \in \mathcal{H}_\omega$ such that*

$$\omega(A) = \langle \Omega_\omega, \pi_\omega(A)\Omega_\omega \rangle, \quad A \in \mathcal{A}_\Gamma, \quad \text{and} \quad \mathcal{H}_\omega = \overline{\pi_\omega(\mathcal{A}_\Gamma)\Omega_\omega}. \tag{3.1}$$

Here, $\bar{\cdot}$ denotes the norm closure. It is unique up to unitary equivalence.

The triple $(\mathcal{H}_\omega, \pi_\omega, \Omega_\omega)$ is called the GNS triple of ω. We frequently consider the commutant or bicommutant of $\pi_\omega(\mathcal{A}_\Gamma)$. For a $*$-algebra $\mathcal{M}$ acting on a Hilbert space $\mathcal{H}$, we denote by $\mathcal{M}'$ the set of all elements in $\mathcal{B}(\mathcal{H})$ (the set of all bounded operators on $\mathcal{H}$) commuting with every element in $\mathcal{M}$. The algebra $\mathcal{M}'$ is called a commutant of $\mathcal{M}$, and the commutant of $\mathcal{M}'$ is called bicommutant and denoted by $\mathcal{M}''$.

For a pure state ω, it is known that π_ω is irreducible (i.e., there is no nontrivial closed subspace of $\mathcal{H}_\omega$ invariant under $\pi_\omega(\mathcal{A}_\Gamma)$) and $\pi_\omega(\mathcal{A}_\Gamma)$ is dense in $\mathcal{B}(\mathcal{H}_\omega)$ with respect to the strong operator topology. This property can be rephrased as $\pi_\omega(\mathcal{A}_\Gamma)'' = \mathcal{B}(\mathcal{H}_\omega)$.

Given GNS representations, we can introduce some equivalence relation between states. We say that two states ω, φ on $\mathcal{A}_\Gamma$ are equivalent (denoted $\omega \simeq \varphi$) if and only if the corresponding GNS representations are unitarily equivalent. For a state ω and an automorphism α on $\mathcal{A}_\Gamma$, if ω and $\omega \circ \alpha$ are equivalent, then there is a unitary u on the GNS Hilbert space $\mathcal{H}_\omega$ implementing α in the sense

$$\mathrm{Ad}(u) \circ \pi_\omega = \pi_\omega \circ \alpha. \tag{3.2}$$

This is because $\pi_\omega \circ \alpha$ is a GNS representation of $\omega \circ \alpha$. In our context of quantum spin systems, we can see that two states ω, φ are equivalent if they can be approximated by a local perturbation of each other. More precisely, ω can be approximated arbitrarily well in the norm topology of $\mathcal{A}_{\mathbb{Z}^2}^*$ by states of the form $\varphi(A^* \cdot A)$, with $A \in \mathcal{A}_{\mathrm{loc}}$, and vice versa. Physically, it means that ω and φ are macroscopically the same.

There is yet another equivalence relation between states, which is called quasiequivalence. Two states ω, φ are said to be quasiequivalent if there is a $*$-isomorphism $\iota : \pi_\omega(\mathcal{A}_\Gamma)'' \to \pi_\varphi(\mathcal{A}_\Gamma)''$ such that $\pi_\varphi(A) = \iota \circ \pi_\omega(A)$, for all $A \in \mathcal{A}_\Gamma$. Note that if two states are equivalent, they are quasiequivalent. The converse is not true in general, but if the states are pure, it is true.

In the operator-algebraic framework of quantum spin systems, physical models are specified with a map called interaction. An interaction Φ is a map $\Phi : \mathfrak{S}_{\mathbb{Z}^\nu} \to \mathcal{A}_{\mathrm{loc}}$ satisfying

$$\Phi(X) = \Phi(X)^* \in \mathcal{A}_X$$

for all $X \in \mathfrak{S}_{\mathbb{Z}^\nu}$. Physically, this $\Phi(X)$ indicates an interaction term between spins inside of X.

The easiest type of interaction is an on-site interaction, satisfying

$$\Phi(X) = 0 \quad \text{if } |X| \neq 1. \tag{3.3}$$

It means that the only possibly nonzero interaction terms are of the form $\Phi(\{\mathbf{x}\})$, with $\mathbf{x} \in \mathbb{Z}^\nu$. (Here and thereafter, $|X|$ indicates the number of elements in X.) Note that all interaction terms commute with each other for such interactions.

Physically, we are more interested in interactions that have nonzero interaction terms between different sites of $\mathbb{Z}^\nu$. For example, let $\{S_j\}_{j=1,2,3}$ be generators of the irreducible representation of $\mathfrak{su}(2)$ on $\mathbb{C}^d$. Then an interaction of $\mathcal{A}_{\mathbb{Z}}$ given by

$$\Phi\big(\{x, x+1\}\big) = \sum_{j=1}^{3} S_j^{(x)} S_j^{(x+1)}, \quad x \in \mathbb{Z}, \tag{3.4}$$

is called the antiferromagnetic Heisenberg chain, which has been extensively studied.

Now, given an interaction, we would like to define a dynamics on $\mathcal{A}_{\mathbb{Z}^\nu}$ out of it. For this, we need to assume that Φ is "suitably local." The simplest condition among such is the condition of the uniform boundedness and finite range. An interaction is of finite range if there exists an $m \in \mathbb{N}$ such that $\Phi(X) = 0$ for X with a diameter larger than m. It is uniformly bounded if it satisfies $\sup_{X \in \mathfrak{S}_{\mathbb{Z}^\nu}} \|\Phi(X)\| < \infty$. We can relax this restriction extensively. More generally, we define norms on interactions and consider interactions with finite norms; see [40].

Given a suitably local interaction, we may define a C^*-dynamics, i.e., strongly continuous one-parameter group of automorphisms on $\mathcal{A}_{\mathbb{Z}^\nu}$. For an interaction Φ and a finite set $\Lambda \subset \mathbb{Z}^\nu$, we define the local Hamiltonian on Λ by

$$(H_\Phi)_\Lambda := \sum_{X \subset \Lambda} \Phi(X). \tag{3.5}$$

Then we consider the Heisenberg dynamics given by the local Hamiltonian $e^{it(H_\Phi)_\Lambda} A e^{-it(H_\Phi)_\Lambda}$ and take the thermodynamic limit. If our interaction Φ is suitably local, for example, if it is a uniformly bounded finite-range interaction, the limit

$$\tau_\Phi^t(A) = \lim_{\Lambda \to \mathbb{Z}^\nu} e^{it(H_\Phi)_\Lambda} A e^{-it(H_\Phi)_\Lambda}, \quad t \in \mathbb{R},\ A \in \mathcal{A}_{\mathbb{Z}^\nu} \tag{3.6}$$

exists and defines a dynamics τ_Φ on $\mathcal{A}_{\mathbb{Z}^\nu}$. The reason why we consider the dynamics τ_Φ instead of Hamiltonians is because there is no mathematically meaningful limit of local Hamiltonians $(H_\Phi)_\Lambda$ as $\Lambda \to \mathbb{Z}^\nu$, while the limit (3.6) makes sense. For this reason, in the operator-algebraic framework of quantum statistical mechanics, we talk about dynamics instead of Hamiltonians.

For the same reason, a ground state is defined in terms of the dynamics τ_Φ. Let δ_Φ be the generator of τ_Φ. A state ω on $\mathcal{A}_{\mathbb{Z}^\nu}$ is called a τ_Φ-ground state if the inequality

$$-i\omega\big(A^* \delta_\Phi(A)\big) \geq 0 \tag{3.7}$$

holds for any element A in the domain $\mathcal{D}(\delta_\Phi)$ of δ_Φ. We occasionally say a ground state of Φ instead of a τ_Φ-ground state. We denote by $\mathcal{G}_\Phi$ the set of all ground states of Φ. Clearly, $\mathcal{G}_\Phi$ is a weak$*$-compact convex set, and it is known that its extremal points $\operatorname{ex} \mathcal{G}_\Phi$ consists of pure states (see [7, THEOREM 5.3.37]).

Let $(\mathcal{H}_\omega, \pi_\omega, \Omega_\omega)$ be the GNS triple of a τ_Φ-ground state ω. Then there exists a unique positive operator $H_{\omega,\Phi}$ on $\mathcal{H}_\omega$ such that $e^{itH_{\omega,\Phi}} \pi_\omega(A)\Omega_\omega = \pi_\omega(\tau_\Phi^t(A))\Omega_\omega$, for all $A \in \mathcal{A}_{\mathbb{Z}^\nu}$ and $t \in \mathbb{R}$. We call this $H_{\omega,\Phi}$ the bulk Hamiltonian associated with ω. Note that Ω_ω is an eigenvector of $H_{\omega,\Phi}$ with eigenvalue 0 (see [7, PROPOSITION 5.3.19]).

Let us consider the corresponding condition for a finite quantum system M_n with dynamics given by a Hamiltonian H (2.1). Let p be the spectral projection of H corresponding to the lowest eigenvalue E_0. Recall that a state ω on M_n is given by a density matrix ρ with the formula $\omega(A) = \operatorname{Tr} \rho A$. Let $s(\rho)$ be the support projection of this ρ. Then one can check that ω is a τ-ground state if and only if $s(\rho)$ satisfies $s(\rho) \leq p$. Recall that the last condition is the very definition of the ground state in finite-dimensional quantum mechanics. In fact, note that the generator δ of τ in (2.1) is $\delta(A) = i[H, A]$. If $s(\rho) \leq p$, then we have

$$-i\omega\big(A^*\delta(A)\big) = \omega\big(A^*(H - E_0)A\big) \geq 0, \quad A \in \mathrm{M}_n,$$

hence ω is a τ-ground state. Conversely, suppose that ω is a τ-ground state. For any unit eigenvectors ξ, η of H with $H\xi = E_0\xi$, $H\eta = E\eta$, for $E > E_0$, set $A \in \mathrm{M}_n$ to be a matrix satisfying $A\zeta = \langle \eta, \zeta \rangle \xi$ for any $\zeta \in \mathbb{C}^n$. Substituting this A, we get

$$0 \leq -i\omega\big(A^*\delta(A)\big) = (E_0 - E)\langle \eta, \rho\eta \rangle.$$

Because $E_0 - E < 0$, this means that $\langle \eta, \rho\eta \rangle = 0$ for any such η. Hence we conclude that $p\rho p = \rho$, namely, $s(\rho) \leq p$. It means that our definition in the operator-algebraic framework can be regarded as a natural generalization of the usual definition of a ground state to infinite systems.

Note, in general, that there can be many states satisfying condition (3.7). Namely, the ground state need not be unique. If the ground state is unique, it is automatically an extremal point of $\mathcal{G}_\Phi$. As a result, it is pure.

The systems we are interested in, in this paper, are those with gapped ground states.

Definition 3.1. We say that Φ has gapped ground states in the bulk if the following hold:

(i) The bulk Hamiltonian $H_{\omega,\Phi}$ of any pure τ_Φ-ground state ω has 0 as its nondegenerate eigenvalue.

(ii) There exists a constant $\gamma > 0$ such that

$$\sigma(H_{\omega,\Phi}) \setminus \{0\} \subset [\gamma, \infty), \tag{3.8}$$

for any pure τ_Φ-ground state ω. Here $\sigma(H_{\omega,\Phi})$ denotes the spectrum of $H_{\omega,\Phi}$.

We denote by $\mathcal{P}$ the set of all uniformly bounded finite-range interactions with gapped ground states in the bulk.

An interaction Φ is said to have a unique gapped ground state if its ground state is unique and gapped in the sense of Definition 3.1; see [1, 17, 18, 42–44] for examples of such models. If we consider the corresponding condition for a finite system M_n with dynamics (2.1). This condition corresponds to the situation that "the lowest eigenvalue of H is nondegenerate and the difference between the lowest eigenvalue and the second-lowest eigenvalue is at least γ." One remarkable property of the unique gapped ground state is the exponential decay of correlation functions.

Theorem 3.2 ([22, 37, 39]). *Let Φ be a uniformly bounded finite-range interaction with a unique gapped ground state ω_Φ. Then the correlation functions of ω_Φ decay exponentially fast: there exist constants $\mu > 0$ and $C > 0$ such that for all $A \in \mathcal{A}_X$, $B \in \mathcal{A}_Y$, with finite $X, Y \subset \mathbb{Z}^\nu$,*

$$\big|\omega_\Phi(AB) - \omega_\Phi(A)\omega_\Phi(B)\big| \leq C\|A\|\|B\||X|e^{-\mu d(X,Y)}$$

holds. Here $d(X, Y)$ denotes the distance between X and Y.

This means ω_Φ is "almost like a product state."

4. PATHS OF AUTOMORPHISMS GENERATED BY TIME-DEPENDENT INTERACTIONS

In the previous section, we considered time-independent interactions, and derived a C^*-dynamics out of them. The same procedure can be carried out for time-dependent interactions to derive strongly continuous paths of automorphisms. (Recall that in finite-dimensional quantum mechanics, we also considered time-dependent Hamiltonians.) Let $\Phi : [0,1] \ni t \to \Phi_t = (\Phi(X;t))$ be a piecewise-continuous path of interactions. Namely, for each finite X, the matrix-valued function $[0,1] \ni t \to \Phi(X;t) \in \mathcal{A}_X$ is piecewise continuous. We then define the path of local Hamiltonians $(H_{\Phi_t})_\Lambda := \sum_{X \subset \Lambda} \Phi(X;t)$ for each finite subset Λ of $\mathbb{Z}^\nu$ and consider the solution $\alpha_{\Phi,t,\Lambda}(A)$ of the differential equation

$$\frac{d}{dt}\alpha_{\Phi,t,\Lambda}(A) = i\big[(H_{\Phi_t})_\Lambda, \alpha_{\Phi,t,\Lambda}(A)\big], \quad \alpha_{\Phi,0,\Lambda}(A) = A.$$

If the interactions along this path are suitably local, analogous to those considered in the previous section, then the thermodynamic limit

$$\alpha_{\Phi,t}(A) = \lim_{\Lambda \to \mathbb{Z}^\nu} \alpha_{\Phi,t,\Lambda}(A), \quad A \in \mathcal{A}_{\mathbb{Z}^\nu}$$

exists and defines a strongly continuous path of automorphisms $\alpha_{\Phi,t}$. We denote by $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ the set of all automorphisms $\alpha = \alpha_{\Phi,t}$ generated by some time-dependent interactions Φ in this manner. It forms a subgroup of the automorphism group $\mathrm{Aut}(\mathcal{A}_{\mathbb{Z}^\nu})$ on $\mathcal{A}_{\mathbb{Z}^\nu}$.

Due to the fact that $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ is given out of local interactions, it shows some nice locality properties. The most famous one is the Lieb–Robinson bound, which has been extensively studied and used [4,22,37,39,40]. It gives an estimate on $\|[\alpha(A), B]\|$ for $A \in \mathcal{A}_X$, $B \in \mathcal{A}_Y$, which decays as the distance between finite subsets X and Y goes to infinity.

The other property that is satisfied by $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ is the factorization property. It basically says that we can split α into two along any cut of the system modulo some error terms localized around the boundary. For example, in one-dimensional systems, if we cut the system into two parts at the origin, we have

$$\alpha = \mathrm{Ad}(v) \circ (\alpha_L \otimes \alpha_R), \tag{4.1}$$

where α_L is an automorphism on the left infinite chain $\mathcal{A}_L := \mathcal{A}_{(-\infty,-1]\cap\mathbb{Z}}$, while α_R is an automorphism on the right infinite chain $\mathcal{A}_R := \mathcal{A}_{[0,\infty)\cap\mathbb{Z}}$. The term $\mathrm{Ad}(v)$ is an inner automorphism given by some unitary v in $\mathcal{A}_\mathbb{Z}$, which corresponds to the "error around the boundary." In a two-dimensional system, for example, we have the following when we cut the system into two by the y-axis. For $0 < \theta < \frac{\pi}{2}$, we define a double cone C_θ by

$$C_\theta := \big\{(x,y) \in \mathbb{Z}^2 \mid |y| \le \tan\theta \cdot |x|\big\}. \tag{4.2}$$

Furthermore, H_L, H_R, H_U, H_D denotes left/right and upper/lower half-planes, and $C_{\theta,L} := C_\theta \cap H_L$, $C_{\theta,R} := C_\theta \cap H_R$. For any $0 < \theta < \frac{\pi}{2}$, there is $\alpha_L \in \mathrm{Aut}\,\mathcal{A}_{H_L}$, $\alpha_R \in \mathrm{Aut}\,\mathcal{A}_{H_R}$, and $\Theta \in \mathrm{Aut}\,\mathcal{A}_{(C_\theta)^c}$ such that

$$\alpha = \mathrm{Ad}(v)(\alpha_L \otimes \alpha_R) \circ \Theta. \tag{4.3}$$

Actually, α can be cut in many directions simultaneously. The factorization property is a simple but strong analytical property, which turns out to be useful in the analysis of gapped ground state phases [36,45–47,49].

Another property we note about $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ is that it does not create a long-range entanglement. For example, it satisfies the following property. If A and B are observables localized in finite regions far away from each other, then α almost preserves the tensor product form of $A \otimes B$, namely, there are operators $\tilde{A}$, $\tilde{B}$ strictly localized in some finite disjoint areas such that $\tilde{A} \otimes \tilde{B}$ approximates $\alpha(A \otimes B)$ in the norm topology. In fact, our α can be regarded as a version of a quantum circuit with finite depth, which is regarded as a quantum circuit which does not create long-range entanglement [3]. From this point of view, we say a state has a short-range entanglement if it is of the form

$$\left(\bigotimes_{\boldsymbol{x}\in\mathbb{Z}^\nu} \rho_{\boldsymbol{x}}\right) \circ \alpha, \tag{4.4}$$

with infinite tensor product state $\bigotimes_{\boldsymbol{x}\in\mathbb{Z}^\nu} \rho_{\boldsymbol{x}}$ and an automorphism $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$. Otherwise, we say it has a long-range entanglement.

In the physics literature, the classification of states with respect to local unitaries is considered [14]. Two states are equivalent if there is a local unitary connecting them. In our framework, these local unitaries can be understood as automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$, and the classification in [14] can be reformulated as follows. For two states ω_1, ω_0 on $\mathcal{A}_{\mathbb{Z}^\nu}$, we write $\omega_1 \sim_{\mathrm{l.}u.} \omega_0$ if there is an automorphism $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ such that $\omega_1 = \omega_0 \circ \alpha$. This gives some equivalence relation. From the fact that automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ do not create long-range entanglement, this is one physically natural criterion of classification of states.

5. THE CLASSIFICATION OF GAPPED GROUND STATE PHASES

The automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ are of fundamental importance in the classification problem of gapped ground state phases. In a word, ground state spaces of two interactions $\Phi_0, \Phi_1 \in \mathcal{P}$ (Definition 3.1) are connected to each other via such automorphisms if they are equivalent in the classification of gapped ground state phases. In this section, we introduce such a theorem, called the automorphic equivalence. The automorphic equivalence started as Hasting's adiabatic lemma [23] in finite-dimensional quantum mechanical system. There have been seminal mathematical improvements and generalizations after that [4,40] in the context of the thermodynamic limit of quantum spin systems. Here we introduce a version from [33], where we require the spectral gap only in the infinite systems (i.e., the setting in Section 3).

The classification problem of gapped ground states in infinite systems can be roughly described as follows.

We say that two interactions $\Phi_0, \Phi_1 \in \mathcal{P}$ are equivalent if there is a path of interactions $\Phi : [0,1] \to \mathcal{P}$ satisfying the following conditions:

(1) $\Phi(0) = \Phi_0$ and $\Phi(1) = \Phi_1$;

(2) $[0,1] \ni s \mapsto \Phi(X;s) \in \mathcal{A}_X$ is continuous and piecewise C^1. The interaction $\Phi(s)$ and its derivative are of finite range, bounded with respect to some norm uniformly in $s \in [0,1]$ (see (ii)–(iv) of Assumption 1.2 in [33]);

(3) For each pure τ_{Φ_0}-ground state φ_0, there is a unique smooth path of states φ_s where each φ_s is a pure $\tau_{\Phi(s)}$-ground state. (Here, smooth means the expectation value of some class of elements in $\mathcal{A}_{\mathbb{Z}^\nu}$ with respect to φ_s is differentiable, and its derivative is not too large compared to some norm; see [33, ASSUMPTION 1.2(VII)].) For each $s \in [0,1]$, the map $\operatorname{ex} \mathcal{G}_{\Phi_0} \ni \varphi_0 \mapsto \varphi_s \in \operatorname{ex} \mathcal{G}_{\Phi_s}$ gives a bijection;

(4) The gap is uniformly bounded from below by some $\gamma > 0$ along the path, i.e., $\sigma(H_{\psi_s,\Phi(s)}) \setminus \{0\} \subset [\gamma, \infty)$ for all $s \in [0,1]$ and a pure τ_{Φ_s}-ground state ψ_s.

We write $\Phi_0 \sim \Phi_1$ if $\Phi_0, \Phi_1 \in \mathcal{P}$ are equivalent in this sense.

The automorphic equivalence in this setting is given as follows.

Theorem 5.1 ([33]). *If $\Phi_0 \sim \Phi_1$, then there is an $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ such that*

$$\mathcal{G}_{\Phi_1} = \mathcal{G}_{\Phi_0} \circ \alpha. \tag{5.1}$$

Proof. We use the notation above for $\Phi_0 \sim \Phi_1$. From Remark 1.4 of [33], there is a path of automorphisms $\alpha_s \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ satisfying $\varphi_s = \varphi_0 \circ \alpha_s$ for each state φ_0, φ_s in (3). This α_s is independent of the choice of φ_0. Because $\mathcal{G}_{\Phi(s)}$ is a convex weak$*$-compact set, it coincides with the weak$*$-closure of the convex hull of extremal points of $\mathcal{G}_{\Phi(s)}$. Hence we see that this α_s maps $\mathcal{G}_{\Phi(0)}$ to $\mathcal{G}_{\Phi(s)}$ bijectively. ■

Hence automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ connect ground state spaces of Φ_0 and Φ_1. For this reason, this class of automorphisms is of fundamental importance. The point here is that it is not only that there is some automorphism connecting the ground state spaces, but also that we know the details of the automorphisms.

Note that for interactions $\Phi_1, \Phi_0 \in \mathcal{P}$ with unique ground states $\omega_{\Phi_1}, \omega_{\Phi_0}$, $\Phi_1 \sim \Phi_0$ implies $\omega_{\Phi_1} \sim_{\mathrm{l.u.}} \omega_{\Phi_0}$ by Theorem 5.1. At the moment of writing, it is not clear to us if the converse is true.

We call an on-site interaction (defined in (3.3)) with a unique gapped ground state a trivial interaction. The unique ground state ω_{Φ_0} of a trivial interaction Φ_0 is of infinite tensor product form. One can easily see that any two trivial interactions are equivalent. The equivalence class $\mathcal{P}_0$ of interactions including these trivial interactions is called a trivial phase. Any interaction Φ in the trivial phase has a unique ground state, and, by Theorem 5.1, it has a short-range entanglement (4.4).

6. SYMMETRY PROTECTED TOPOLOGICAL (SPT) PHASES

The trivial phase $\mathcal{P}_0$ consists of interactions that are connected to trivial interactions, and as a result, its ground state has a short-range entanglement which is basically the same as product states. From this point of view, the trivial phase itself may not be that interesting. However, if we introduce some symmetry to the game, we can extract some interesting mathematical structure out of it. This is so-called symmetry protected topological (SPT) phases, which were introduced by Gu and Wen [12, 13, 21]. Throughout this section, ω_Φ for $\Phi \in \mathcal{P}_0$ indicates the unique ground state of Φ.

In this talk, as a symmetry, we consider an on-site finite group symmetry, which is defined as follows. (A study on the global reflection symmetry in one-dimensional systems can be found in [46].) We fix a finite group G and a (projective) unitary representation U of G on $\mathbb{C}^d$. Then there is a unique automorphism β_g satisfying

$$\beta_g(A) = \left(\bigotimes_{x\in\Lambda} U(g)\right) A \left(\bigotimes_{x\in\Lambda} U(g)^*\right), \quad g \in G,\ A \in \mathcal{A}_\Lambda,\ \Lambda \in \mathfrak{S}_{\mathbb{Z}^\nu}.$$

Clearly, this gives an action of G on $\mathcal{A}_{\mathbb{Z}^\nu}$, i.e., $\beta_g\beta_h = \beta_{gh}$ for $g, h \in G$. We call this action of G, an on-site symmetry given by G and U. We say an interaction Φ is β-invariant if $\beta_g(\Phi(X)) = \Phi(X)$ for all $X \in \mathfrak{S}_{\mathbb{Z}^\nu}$ and $g \in G$. For a ground state φ of a β-invariant interaction Φ, one can check that $\varphi \circ \beta_g$ is also a ground state of Φ. Therefore, if a β-invariant interaction Φ has a unique ground state ω_Φ, the ground state is β-invariant, $\omega_\Phi \circ \beta_g = \omega_\Phi$.

What we are interested in, in this section, is the set of all β-invariant interactions in the trivial phase $\mathcal{P}_0$. We denote the set of all such interactions by $\mathcal{P}_{0,\beta}$. We would like to classify them with respect to the following criterion. Two interactions Φ_0, Φ_1 are β-equivalent if there is a smooth path of interactions in $\mathcal{P}_{0,\beta}$ satisfying the conditions (1)–(4) we saw in Section 5. We write $\Phi_0 \sim_\beta \Phi_1$ in this case. The difference between $\sim$ and $\sim_\beta$ is that we require the symmetry to be preserved along the path. Because of this additional condition, there can be interactions $\Phi_0, \Phi_1 \in \mathcal{P}_{0,\beta}$, which satisfy $\Phi_0 \sim \Phi_1$ (by definition) but not $\Phi_0 \sim_\beta \Phi_1$. In other words, $\mathcal{P}_{0,\beta}$ may split into possibly multiple equivalence classes. The resulting equivalence classes are the symmetry-protected topological (SPT) phases.

For this SPT classification problem, physicists and algebraic topologists have a conjecture [26, 56]. They say that SPT-phases should be understood in terms of the invertible quantum field theory. As a result, for a finite group G, SPT-phases should be classified by the Pontryagin dual of bordism group on the classifying space BG of G. In one and two dimensions, these Pontryagin duals are $H^2(G, \mathrm{U}(1))$, $H^3(G, \mathrm{U}(1))$. In fact, we can derive these group-cohomology-valued invariants out of our general microscopic models of in those dimensions.

Theorem 6.1 ([45, 47]). *There is an $H^2(G, \mathrm{U}(1))$-valued invariant for one-dimensional SPT-phases. There is an $H^3(G, \mathrm{U}(1))$-valued invariant for two-dimensional SPT-phases.*

For the rest of this section, we explain how to find such invariants out of general models. In the analysis of gapped ground state phases, there is a general guiding principle to find an invariant. That is, cut the system into two and look at the edge. This principle is sometimes called the bulk-edge correspondence. In order to derive the invariant in Theorem 6.1, we follow this principle and restrict our group action β to the half of the system.

Namely, we consider the group actions

$$\beta_g^R := \mathrm{id}_{\mathcal{A}_L} \otimes \bigotimes_{x\geq 0} \mathrm{Ad}\big(U(g)\big), \quad \beta_g^U := \mathrm{id}_{\mathcal{A}_{H_D}} \otimes \bigotimes_{(x,y)\in H_U} \mathrm{Ad}\big(U(g)\big), \tag{6.1}$$

in one and two dimensions, respectively. We investigate the effect of these actions on our unique ground state ω_Φ for $\Phi \in \mathcal{P}_{0,\beta}$.

Let us start with one-dimensional systems. Recall that ω_Φ has a short-range entanglement, and is β-invariant. From these facts, we expect that the effect of β^R is not much recognizable on the left infinite chain, far away from the origin. On the other hand, on the right infinite chain, far away from the origin, the differences between β and β^R are not much recognizable. Combining this and the fact that ω_Φ is β-invariant, we conclude that the effect of β^R is not much recognizable on the right infinite chain, far away from the origin. As a result, we expect that the effect of β^R on ω_Φ should be localized around the origin. In other words, ω_Φ and $\omega_\Phi \circ \beta_g^R$ are macroscopically the same. It turns out to be true, mathematically, in the following sense.

Proposition 6.1. *The states ω_Φ and $\omega_\Phi \circ \beta_g^R$ are equivalent.*

This can be seen very easily. Recall from the definition that $\Phi \in \mathcal{P}_0$ means $\Phi \sim \Phi_0$ with some trivial interaction Φ_0. By Theorem 5.1, we have $\omega_\Phi = \omega_{\Phi_0} \circ \alpha$ with some $\alpha \in \mathrm{QAut}(\mathcal{A}_\mathbb{Z})$. Recall that, as a trivial interaction, Φ_0 has a unique ground state of infinite tensor product form. In particular, we can write ω_{Φ_0} as $\omega_{\Phi_0} = \omega_L \otimes \omega_R$ with pure states ω_L, ω_R on the left and right infinite chains $\mathcal{A}_L$, $\mathcal{A}_R$, respectively. Recall also that our α satisfies the factorization property (4.1). Combining these facts, we conclude that

$$\omega_\Phi \simeq (\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R), \tag{6.2}$$

with some automorphisms α_L, α_R on $\mathcal{A}_L$, $\mathcal{A}_R$. From this and the invariance of ω_Φ under β_g, we see that $\omega_L\alpha_L\beta_g^L \otimes \omega_R\alpha_R\beta_g^R \simeq \omega_L\alpha_L \otimes \omega_R\alpha_R$, where β^L, β^R are the restrictions of β to the left and right infinite chains, respectively. This implies $\omega_R\alpha_R\beta_g^R \simeq \omega_R\alpha_R$, hence we get

$$\omega_\Phi\beta_g^R \simeq \omega_L\alpha_L \otimes \omega_R\alpha_R\beta_g^R \simeq \omega_L\alpha_L \otimes \omega_R\alpha_R \simeq \omega_\Phi, \tag{6.3}$$

proving the claim.

Note from Section 3 that Proposition 6.1 means β_g^R is implementable by a unitary u_g in the GNS representation $(\mathcal{H}_{\omega_\Phi}, \pi_{\omega_\Phi})$ of ω_Φ, i.e.,

$$\mathrm{Ad}(u_g) \circ \pi_{\omega_\Phi} = \pi_{\omega_\Phi} \circ \beta_g^R. \tag{6.4}$$

Because β^R is a group action, we have

$$\mathrm{Ad}(u_g u_h) \circ \pi_{\omega_\Phi} = \pi_{\omega_\Phi} \circ \beta_g^R\beta_h^R = \pi_{\omega_\Phi} \circ \beta_{gh}^R = \mathrm{Ad}(u_{gh}) \circ \pi_{\omega_\Phi}, \quad g, h \in G. \tag{6.5}$$

Recall that ω_Φ is a unique ground state of Φ, hence it is pure. As a result, $\pi_{\omega_\Phi}(\mathcal{A}_\mathbb{Z})$ is dense in $\mathcal{B}(\mathcal{H}_{\omega_\Phi})$ with respect to the strong operator topology. From this, (6.5) implies that there is some $\sigma(g,h) \in \mathrm{U}(1)$ such that

$$u_g u_h = \sigma(g,h) u_{gh}, \quad g, h \in G. \tag{6.6}$$

In other words, (u_g) forms a projective representation. As a result, we obtain $H^2(G, \mathrm{U}(1))$-valued index out of it.

Using the automorphic equivalence Theorem 5.1 and the factorization property of the automorphism therein, one can show that it is in fact an invariant of our classification $\sim_\beta$ [45]. The point of the proof is, when $\Phi_0 \sim_\beta \Phi_1$, that the time-dependent interactions giving $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}})$ in Theorem 5.1 can be taken to be β-invariant. Proposition 6.1 itself holds for general β-invariant unique gapped ground state. This is thanks to the theorem by Matsui [31] showing the split property for unique gapped ground states. Projective representations associated to split states have been known since the year 2000 [30] among operator algebraists. What is new here is that the associated cohomology class is an invariant of our classification. In fact, this $H^2(G, \mathrm{U}(1))$-valued index is a complete invariant of pure β-invariant split states with respect to some classification [48]. This index can be used to show Lieb–Schultz–Mattis-type theorems [2,29,30,38] (no-go theorems for the existence of unique gapped ground state under some symmetry), for finite groups symmetries [50,51].

For two dimensions, $\omega_\Phi \circ \beta_g^U$ is not equivalent to ω_Φ in general. However, an analogous argument as in the one-dimensional case lets us expect that the effect of β_g^U should be localized around the x-axis. In fact, it turns out to be true mathematically.

Proposition 6.2. *For any* $0 < \theta < \frac{\pi}{2}$, *there are* $\eta_{g,L} \in \mathrm{Aut}(\mathcal{A}_{C_{\theta,L}})$ *and* $\eta_{g,R} \in \mathrm{Aut}(\mathcal{A}_{C_{\theta,R}})$ *such that*

$$\omega_\Phi \circ \beta_g^U \simeq \omega_\Phi(\eta_{g,L} \otimes \eta_{g,R}).$$

It means macroscopically that the effect of β_g^U on ω_Φ is localized around $C_{\theta,L}$ and $C_{\theta,R}$ for any $0 < \theta < \frac{\pi}{2}$. This $\eta_{g,R}$ is our source of the $H^3(G, \mathrm{U}(1))$-valued index.

Now we fix some $0 < \theta < \frac{\pi}{2}$, and set $\gamma_g^R := \beta_g^{UR} \circ \eta_{g,R}^{-1}$, $\gamma_g^L := \beta_g^{UL} \circ \eta_{g,L}^{-1}$ with $\eta_{g,R}$, $\eta_{g,L}$ for this θ. Here, β_g^{UR}, β_g^{UL} are group actions of G given by

$$\beta_g^{UR} := \mathrm{id}_{(H_U \cap H_R)^c} \otimes \bigotimes_{(x,y) \in H_U \cap H_R} \mathrm{Ad}\big(U(g)\big),$$

$$\beta_g^{UL} := \mathrm{id}_{(H_U \cap H_L)^c} \otimes \bigotimes_{(x,y) \in H_U \cap H_L} \mathrm{Ad}\big(U(g)\big).$$

From Proposition 6.2, we have

$$\omega_\Phi \circ \left(\gamma_g^L \otimes \gamma_g^R\right) \simeq \omega_\Phi, \quad g \in G. \tag{6.7}$$

On the other hand, recall from the definition that $\Phi \in \mathcal{P}_0$ means $\Phi \sim \Phi_0$ with some trivial interaction Φ_0. By Theorem 5.1, we have $\omega_\Phi = \omega_{\Phi_0} \circ \alpha$, with $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^\nu})$ satisfying the factorization property, i.e.,

$$\alpha = \mathrm{Ad}(v) \circ (\alpha_L \otimes \alpha_R) \circ \Theta, \quad \alpha_L \in \mathrm{Aut}\,\mathcal{A}_{H_L}, \ \alpha_R \in \mathrm{Aut}\,\mathcal{A}_{H_R}, \ \Theta \in \mathrm{Aut}\,\mathcal{A}_{C_\theta^c}, \tag{6.8}$$

for our fixed θ. Recall that as a trivial interaction, Φ_0 has a unique ground state ω_{Φ_0} of infinite tensor product form. In particular, we can write ω_{Φ_0} as $\omega_{\Phi_0} = \omega_L \otimes \omega_R$ with pure states ω_L, ω_R on $\mathcal{A}_{H_L}$, $\mathcal{A}_{H_R}$, respectively. Combining these, we conclude that

$$\omega_\Phi \simeq (\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \Theta. \tag{6.9}$$

Repeated use of (6.7) gives

$$\omega_\Phi \circ \left(\gamma_g^L \gamma_h^L \left(\gamma_{gh}^L\right)^{-1} \otimes \gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1}\right) \simeq \omega_\Phi. \tag{6.10}$$

Applying (6.9) to this, we obtain

$$\begin{aligned}(\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \Theta \circ \left(\gamma_g^L \gamma_h^L \left(\gamma_{gh}^L\right)^{-1} \otimes \gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1}\right)\\ \simeq (\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \Theta.\end{aligned} \tag{6.11}$$

Note that

$$\gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1} = \left(\beta_g^{UR} \eta_{g,R}^{-1} \left(\beta_g^{UR}\right)^{-1}\right)\left(\beta_{gh}^{UR} \eta_{h,R}^{-1} \eta_{gh,R} \left(\beta_{gh}^{UR}\right)^{-1}\right) \in \operatorname{Aut}(\mathcal{A}_{C_{\theta,R}}). \tag{6.12}$$

Similarly, we have $\gamma_g^L \gamma_h^L (\gamma_{gh}^L)^{-1} \in \operatorname{Aut}(\mathcal{A}_{C_{\theta,L}})$. Therefore, they commute with $\Theta \in \operatorname{Aut}(\mathcal{A}_{C_\theta^c})$. From this and (6.11), we obtain

$$(\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \left(\gamma_g^L \gamma_h^L \left(\gamma_{gh}^L\right)^{-1} \otimes \gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1}\right) \simeq (\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R),$$

which implies

$$\omega_R \alpha_R \gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1} \simeq \omega_R \alpha_R. \tag{6.13}$$

Recall from Section 3 that this means the automorphism $\gamma_g^R \gamma_h^R (\gamma_{gh}^R)^{-1}$ is implementable by a unitary $u(g,h)$ in the GNS representation $(\mathcal{H}_R, \pi_R)$ of $\omega_R \alpha_R$, i.e.,

$$\operatorname{Ad}\big(u(g,h)\big)\pi_R = \pi_R \gamma_g^R \gamma_h^R \left(\gamma_{gh}^R\right)^{-1}. \tag{6.14}$$

Note also that (6.9) and (6.7) imply

$$(\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \Theta \circ \left(\gamma_g^L \otimes \gamma_g^R\right) \simeq (\omega_L \otimes \omega_R) \circ (\alpha_L \otimes \alpha_R) \circ \Theta. \tag{6.15}$$

Therefore, with $(\mathcal{H}_L, \pi_L)$ a GNS representation of $\omega_L \alpha_L$, there is a unitary W_g on $\mathcal{H}_L \otimes \mathcal{H}_R$ implementing $\Theta \circ (\gamma_g^L \otimes \gamma_g^R) \circ \Theta^{-1}$ in the GNS representation $(\mathcal{H}_L \otimes \mathcal{H}_R, \pi_L \otimes \pi_R)$ of $\omega_L \alpha_L \otimes \omega_R \alpha_R$, i.e.,

$$\operatorname{Ad}(W_g)(\pi_L \otimes \pi_R) = (\pi_L \otimes \pi_R) \circ \Theta \circ \left(\gamma_g^L \otimes \gamma_g^R\right) \circ \Theta^{-1}. \tag{6.16}$$

For these $u(g,h)$ (6.14) and W_g (6.16), we claim that there are $c(g,h,k) \in \mathrm{U}(1)$ such that

$$\begin{aligned}\operatorname{Ad}(W_g)\big(\mathbb{I}_L \otimes u(h,k)\big) \cdot \big(\mathbb{I}_L \otimes u(g,hk)\big)\\ = c(g,h,k)\big(\mathbb{I}_L \otimes u(g,h)u(gh,k)\big), \quad g,h,k \in G.\end{aligned} \tag{6.17}$$

To see this, consider $\pi_L \otimes \pi_R \gamma_g^R \gamma_h^R \gamma_k^R$. On the one hand, with the repeated use of (6.14), we have

$$\begin{aligned}\pi_L \otimes \pi_R \gamma_g^R \gamma_h^R \gamma_k^R &= \operatorname{Ad}\big(\mathbb{I}_L \otimes u(g,h)\big)\big(\pi_L \otimes \pi_R \gamma_{gh}^R \gamma_k^R\big)\\ &= \operatorname{Ad}\big(\mathbb{I}_L \otimes u(g,h)u(gh,k)\big)\big(\pi_L \otimes \pi_R \circ \gamma_{ghk}^R\big).\end{aligned} \tag{6.18}$$

On the other hand, note that both of $\gamma_h^R\gamma_k^R(\gamma_{hk}^R)^{-1}$ and $\gamma_g^R(\gamma_h^R\gamma_k^R(\gamma_{hk}^R)^{-1})(\gamma_g^R)^{-1}$ commute with Θ as before. Hence we have

$$\begin{aligned}&\mathrm{id}_L\otimes\gamma_g^R\big(\gamma_h^R\gamma_k^R\big(\gamma_{hk}^R\big)^{-1}\big)\big(\gamma_g^R\big)^{-1}\\&\quad=\Theta\big(\gamma_g^L\otimes\gamma_g^R\big)\Theta^{-1}\big(\mathrm{id}_L\otimes\gamma_h^R\gamma_k^R\big(\gamma_{hk}^R\big)^{-1}\big)\Theta\big(\gamma_g^L\otimes\gamma_g^R\big)^{-1}\Theta^{-1}.\end{aligned}\tag{6.19}$$

From this and repeated use of (6.14), (6.16), we have

$$\begin{aligned}&\pi_L\otimes\pi_R\gamma_g^R\gamma_h^R\gamma_k^R\\&\quad=(\pi_L\otimes\pi_R)\Theta\big(\gamma_g^L\otimes\gamma_g^R\big)\Theta^{-1}\big(\mathrm{id}_L\otimes\gamma_h^R\gamma_k^R\big(\gamma_{hk}^R\big)^{-1}\big)\Theta\big(\gamma_g^L\otimes\gamma_g^R\big)^{-1}\\&\qquad\times\Theta^{-1}\big(\mathrm{id}_L\otimes\gamma_g^R\gamma_{hk}^R\big)\\&\quad=\mathrm{Ad}\big(W_g\big(\mathbb{I}_L\otimes u(h,k)\big)W_g^*\big(\mathbb{I}_L\otimes u(gh,k)\big)\big)\big(\pi_L\otimes\pi_R\gamma_{ghk}^R\big).\end{aligned}\tag{6.20}$$

Comparing this and (6.18), we have

$$\begin{aligned}&\mathrm{Ad}\big(\mathbb{I}_L\otimes u(g,h)u(gh,k)\big)(\pi_L\otimes\pi_R)\\&\quad=\mathrm{Ad}\big(W_g\big(\mathbb{I}_L\otimes u(h,k)\big)W_g^*\big(\mathbb{I}_L\otimes u(gh,k)\big)\big)(\pi_L\otimes\pi_R).\end{aligned}\tag{6.21}$$

Note that, because $(\mathcal{H}_L\otimes\mathcal{H}_R,\pi_L\otimes\pi_R)$ is a GNS representation of a pure state $\omega_L\alpha_L\otimes\omega_R\alpha_R$, $(\pi_L\otimes\pi_R)(\mathcal{A}_{\mathbb{Z}^2})$ is dense in $\mathcal{B}(\mathcal{H}_L\otimes\mathcal{H}_R)$ with the strong operator topology. As a result, (6.21) implies our claim (6.17).

The situation in (6.14), (6.17) is pretty much similar to that of cocycle actions [15,24]. In fact, following the argument in [24], we can show that $c(g,h,k)$ satisfies the 3-cocycle relation. Hence, out of it, we obtain an $H^3(G,\mathrm{U}(1))$-valued index. Using the automorphic equivalence Theorem 5.1 and the factorization property of the automorphism therein, one can show that it is in fact an invariant of our classification $\sim_\beta$.

A derivation of indices for SPT-phases was initially carried out in tensor network models, matrix product states MPS [52–54] in one dimension, and projected entangled pair states [32]. Our indices coincide with theirs in those models. In other words, thanks to those works, there are many examples. Our approach introduced in this section is operator algebraic. Recently, some quantum information based approaches were reported [25,55].

7. ANYONS IN TOPOLOGICAL PHASES

In this section, we consider the classification $\sim_{\mathrm{l.u.}}$ in two dimensions. Recall that states which are equivalent to an infinite tensor product state with respect to $\sim_{\mathrm{l.u.}}$ are said to have a short-range entanglement, and otherwise they are said to have a long-range entanglement. It is frequently said that in the two-dimensional systems, the existence of an "anyon" means the long-range entanglement of the state [28]. In this section, we formulate this statement in our operator-algebraic setting.

An anyon is a string-like excitation with a braiding structure. How to formulate an anyon mathematically is a nontrivial question of mathematical physics. Our answer, motivated by AQFT [27] and studies of Kitaev models [10,19,34,35] is that it is a superselection sector. It is defined in terms of cones. By a cone we mean a subset of $\mathbb{Z}^2$ of the form

$$\Lambda_{\boldsymbol{a},\theta,\varphi} := \left\{ \boldsymbol{x} \in \mathbb{Z}^2 \mid (\boldsymbol{x} - \boldsymbol{a}) \cdot \boldsymbol{e}_\theta > \cos\varphi \cdot \|\boldsymbol{x} - \boldsymbol{a}\| \right\},$$

with some $\boldsymbol{a} \in \mathbb{R}$, $\theta \in \mathbb{R}$, and $\varphi \in (0, \pi)$. Here we set $\boldsymbol{e}_\theta := (\cos\theta, \sin\theta)$. For a cone $\Lambda := \Lambda_{\boldsymbol{a},\theta,\varphi}$ and $\boldsymbol{b} \in \mathbb{R}^2$, $\varepsilon > 0$, we set $\Lambda_\varepsilon + \boldsymbol{b} := \Lambda_{\boldsymbol{a}+\boldsymbol{b},\theta,\varphi+\varepsilon}$, $|\arg\Lambda| := 2\varphi$, and $\boldsymbol{e}_\Lambda := \boldsymbol{e}_\theta$.

Definition 7.1. Let $(\mathcal{H}, \pi_0)$ be an irreducible representation of $\mathcal{A}_{\mathbb{Z}^2}$. We say that a representation π of $\mathcal{A}_{\mathbb{Z}^2}$ on $\mathcal{H}$ satisfies the superselection criterion for π_0 if

$$\pi|_{\mathcal{A}_{\Lambda^c}} \simeq_{u.e.} \pi_0|_{\mathcal{A}_{\Lambda^c}},$$

for any cone Λ in $\mathbb{Z}^2$. (Here, $\simeq_{u.e.}$ means that the two representations are unitarily equivalent.) Such representations are called superselection sectors for π_0.

Superselection sectors are objects studied extensively in AQFT. In the context of quantum spin systems, P. Naaijkens and his coauthors carried out studies on Kitaev's quantum double model from the point of view of superselection sectors [10, 19, 34, 35], where they drove a braiding structure.

We can see the importance of the sector theory for us from the fact that it is an invariant of $\sim_{\mathrm{l}.u.}$.

Theorem 7.1 ([36]). *Let $(\mathcal{H}, \pi_0)$ be an irreducible representation and let $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$. Suppose that a representation π satisfies the superselection criterion for π_0. Then $\pi \circ \alpha$ satisfies the superselection criterion for $\pi_0 \circ \alpha$.*

Let ω_1, ω_0 be pure states such that $\omega_1 \sim_{\mathrm{l}.u.} \omega_0$ with $\omega_1 = \omega_0 \circ \alpha$, $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$. Then, by Theorem 7.1, α gives a bijection between the set of all superselection sectors of π_{ω_0} and the set of all superselection sectors of π_{ω_1}.

The proof of Theorem 7.1 is a simple argument using the factorization property. For $\varepsilon > 0$, analogous to (4.3), we have a decomposition

$$\alpha = \mathrm{Ad}(v) \circ \Xi \circ (\alpha_\Lambda \otimes \alpha_{\Lambda^c}), \tag{7.1}$$

where α_Λ, α_{Λ^c}, Ξ are automorphisms on $\mathcal{A}_\Lambda$, $\mathcal{A}_{\Lambda^c}$, $\mathcal{A}_{\Lambda_\varepsilon}$, respectively. (We choose $\varepsilon > 0$ small enough so that Λ_ε is still a cone.) Then for a superselection sector π for π_0, we have

$$\begin{aligned} \pi \circ \alpha|_{\mathcal{A}_\Lambda} &\sim_{\mathrm{u}.e.} \pi \circ \Xi \circ \alpha_\Lambda|_{\mathcal{A}_\Lambda} = \pi|_{\mathcal{A}_{\Lambda_\varepsilon}} \circ \Xi \circ \alpha_\Lambda|_{\mathcal{A}_\Lambda} \\ &\sim_{\mathrm{u}.e.} \pi_0|_{\mathcal{A}_{\Lambda_\varepsilon}} \circ \Xi \circ \alpha_\Lambda|_{\mathcal{A}_\Lambda} \sim_{\mathrm{u}.e.} \pi_0 \circ \alpha|_{\mathcal{A}_\Lambda}, \end{aligned} \tag{7.2}$$

proving the claim.

We say that π_0 has a trivial sector theory if any representation satisfying the superselection criterion for π_0 is quasiequivalent to π_0. Otherwise, we say π_0 has a nontrivial sector theory. One can show that for a pure state of infinite tensor product form, its GNS representation has a trivial sector theory [36]. Combing this and Theorem 7.1, we obtain the following.

Corollary 7.1. *If a pure state has a short-range entanglement, then its GNS representation has a trivial sector theory.*

In other words, the existence of nontrivial superselection sectors implies the long-range entanglement. If we regard superselection sectors as anyons, it is a mathematical realization of the folklore saying that the existence of anyons implies long-range entanglement of the state.

The reason why we expect superselection sectors to be related to anyons comes from AQFT. Using the tools from AQFT, in [11] Cha–Naaijkens–Nachtergaele derived a braiding structure in a general setting of semigroup of almost localized endomorphisms in quantum spin systems. It is well known that anyons show up in AQFT surprisingly naturally [5, 8, 9, 16, 20, 27]. More precisely, under some condition called Haag duality, a braided C^*-tensor category can be associated to the irreducible representation with nontrivial sector theory. The Haag duality is the property $\pi_0(\mathcal{A}_{\Lambda^c})' = \pi_0(\mathcal{A}_\Lambda)''$, for all cones Λ in $\mathbb{Z}^2$.

The problem for us about introducing this condition in quantum spin systems is that it does not look to be plausible that this condition is stable under automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$. Recalling that automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$ are the fundamental operations in the classification problem of gapped ground state phases, this situation is not convenient for us. For this reason, we introduce a weaker version of Haag duality.

Definition 7.2 (Approximate Haag duality [49]). Let $(\mathcal{H}, \pi_0)$ be an irreducible representation of $\mathcal{A}_{\mathbb{Z}^2}$. We say that $(\mathcal{H}, \pi_0)$ satisfies the approximate Haag duality if the following conditions hold: For any $\varphi \in (0, 2\pi)$ and $\varepsilon > 0$ with $\varphi + 4\varepsilon < 2\pi$, there is some $R_{\varphi,\varepsilon} > 0$ and decreasing functions $f_{\varphi,\varepsilon,\delta}(t)$, $\delta > 0$ on $\mathbb{R}_{\geq 0}$ with $\lim_{t\to\infty} f_{\varphi,\varepsilon,\delta}(t) = 0$ such that

(i) for any cone Λ with $|\arg \Lambda| = \varphi$, there is a unitary $U_{\Lambda,\varepsilon} \in \mathcal{U}(\mathcal{H})$ satisfying

$$\pi_0(\mathcal{A}_{\Lambda^c})' \subset \mathrm{Ad}(U_{\Lambda,\varepsilon})\big(\pi_0(\mathcal{A}_{(\Lambda - R_{\varphi,\varepsilon}\boldsymbol{e}_\Lambda)_\varepsilon})''\big), \tag{7.3}$$

and

(ii) for any $\delta > 0$ and $t \geq 0$, there is a unitary $\tilde{U}_{\Lambda,\varepsilon,\delta,t} \in \pi_0(\mathcal{A}_{\Lambda_{\varepsilon+\delta} - t\boldsymbol{e}_\Lambda})''$ satisfying

$$\|U_{\Lambda,\varepsilon} - \tilde{U}_{\Lambda,\varepsilon,\delta,t}\| \leq f_{\varphi,\varepsilon,\delta}(t). \tag{7.4}$$

The good point about this weaker version is that we know it is stable under automorphisms in $\mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$.

Proposition 7.1. *Let $(\mathcal{H}, \pi_0)$ be an irreducible representation of $\mathcal{A}_{\mathbb{Z}^2}$ satisfying the approximate Haag duality. Then for any automorphism $\alpha \in \mathrm{QAut}(\mathcal{A}_{\mathbb{Z}^2})$, $(\mathcal{H}, \pi_0 \circ \alpha)$ also satisfies the approximate Haag duality.*

It turns out that even with this weaker version of Haag duality and the setting of gapped ground state phases (which is different from that of AQFT), we can still derive a braided C^*-tensor category (see [41] for the definition) out of superselection sectors where, unlike endomorphisms, the multiplication rule is not a priori given [49]. The proof is a modification of the argument in AQFT and some additional argument using the gap condition Definition 3.1. More precisely, let Φ be a uniformly bounded finite range interaction on $\mathcal{A}_{\mathbb{Z}^2}$ with gapped ground states. Let ω be a pure τ_Φ-ground state with a GNS representation

$(\mathcal{H}, \pi_0, \Omega)$. We assume that π_0 has a nontrivial sector theory, and π_0 satisfies the approximate Haag duality. Fix some $\theta \in \mathbb{R}$ and $\varphi \in (0, \pi)$, and denote by $\mathcal{C}_{(\theta,\varphi)}$ the set of all cones whose angle does not intersects with $[\theta - \varphi, \theta + \varphi]$. We set

$$\mathcal{B}_{(\theta,\varphi)} := \overline{\bigcup_{\Lambda \in \mathcal{C}_{(\theta,\varphi)}} \pi_0(\mathcal{A}_{\Lambda^c})'}. \tag{7.5}$$

Here $\bar{\cdot}$ denotes the norm closure. Using the approximate Haag duality, using the argument in [9], each superselection sector $\rho : \mathcal{A}_{\mathbb{Z}^2} \to \mathcal{B}(\mathcal{H}_\omega)$ for π_0 extends to an endomorphism on $\mathcal{B}_{(\theta,\varphi)}$. We denote the extension by the same symbol ρ. Via these extensions, we can introduce compositions between superselection sectors. With this composition as a tensor, the superselection sectors of π_0 are the objects of our braided C^*-tensor category. Our morphisms are given by the intertwiners. Namely, for objects ρ, σ, the morphisms from ρ to σ are bounded operators R on $\mathcal{H}$ such that $R\rho(A) = \sigma(A)R$, for any $A \in \mathcal{A}_{\mathbb{Z}^2}$. The set of all morphisms from ρ to σ is denoted by (ρ, σ). Note that (ρ, σ) is a Banach space and (ρ, ρ) is a C^*-algebra. Following AQFT, the tensor of morphisms $R_1 \in (\rho_1, \sigma_1)$, $R_2 \in (\rho_2, \sigma_2)$ are defined by

$$R_1 \otimes R_2 := R_1 \rho_1(R_2) \in (\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2). \tag{7.6}$$

In fact, each intertwiner belongs to $\mathcal{B}_{(\theta,\varphi)}$ such that $\rho_1(R_2)$ is well-defined. Using the gap inequality and the nontriviality of the sector theory, we can show for any cone Λ that $\pi_0(\mathcal{A}_\Lambda)''$ is either type II_∞ or type III factor. It means that there are isometries $u_\Lambda, v_\Lambda \in \pi_0(\mathcal{A}_\Lambda)''$ such that $u_\Lambda u_\Lambda^* + v_\Lambda v_\Lambda^* = \mathbb{I}$. Using this, for any superselection sectors ρ, σ, we can define their direct sum $\rho \bigoplus \sigma : \mathcal{A}_{\mathbb{Z}^2} \to \mathcal{B}(\mathcal{H}_0)$ by

$$\left(\rho \bigoplus \sigma\right)(A) := u_\Lambda \rho(A) u_\Lambda^* + v_\Lambda \sigma(A) v_\Lambda^*, \quad A \in \mathcal{A}_{\mathbb{Z}^2}. \tag{7.7}$$

From the same fact, we can also define subobjects. Namely, if $p \in (\rho, \rho)$ is a nonzero projection, we can find some superselection sector σ and an isometry v such that $vv^* = p$ and $\rho(A)v = v\sigma(A)$ for all $A \in \mathcal{A}_{\mathbb{Z}^2}$. Hence we obtain the following theorem.

Theorem 7.2 ([49]). *In the above setting, superselection sectors of π_0 form a braided C^*-tensor category. If two of such states $\omega_{\Phi_1}, \omega_{\Phi_2}$ satisfy $\omega_{\Phi_1} \sim_{\mathrm{l.u.}} \omega_{\Phi_2}$, then corresponding braided C^*-tensor categories are monoidally equivalent.*

FUNDING

This work was partially supported by JSPS KAKENHI Grant Number 16K05171 and 19K03534. It was also supported by JST CREST Grant Number JPMJCR19T2.

REFERENCES

[1] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Valence bond ground states in isotropic quantum antiferromagnets. *Comm. Math. Phys.* **115** (1988), 477–528.

[2] I. Affleck and E. H. Lieb, A proof of part of Haldane's conjecture on spin chains. *Lett. Math. Phys.* **12** (1986), 57–69.

[3] S. Bachmann and M. Lange, Trotter product formulae for $*$-automorphisms of quantum lattice systems. 2021, arXiv:2105.14168.

[4] S. Bachmann, S. Michalakis, B. Nachtergaele, and R. Sims, Automorphic equivalence within gapped phases of quantum lattice systems. *Comm. Math. Phys.* **309** (2012), 835–871.

[5] M. Bischoff, Y. Kawahigashi, R. Longo, and K. H. Rehren, *Tensor categories and endomorphisms of von Neumann algebras (with applications to quantum field theory)*. Springer Briefs Math. Phys. 3, Springer 2015.

[6] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics 1*. Springer, 1986.

[7] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics 2*. Springer, 1996.

[8] D. Buchholz, S. Doplicher, G. Morchio, J. E. Roberts, and F. Strocchi, Asymptotic abelianness and braided tensor C^*-categories. In *Rigorous quantum field theory*, edited by A. B. de Monvel, D. Buchholz, and U. Moschella, Progr. Math. 251, Birkhäuser, Basel, 2007.

[9] D. Buchholz and K. Fredenhagen, Locality and the structure of particle states. *Comm. Math. Phys.* **84** (1982), 1–54.

[10] M. Cha, P. Naaijkens, and B. Nachtergaele, The complete set of infinite volume ground states for Kitaev's abelian quantum double models. *Comm. Math. Phys.* **357** (2018), 125–157.

[11] M. Cha, P. Naaijkens, and B. Nachtergaele, On the stability of charges in infinite quantum spin systems. *Comm. Math. Phys.* **373** (2020), 219–264.

[12] X. Chen, Z.-C. Gu, and X.-G. Wen, Classification of gapped symmetric phases in one-dimensional spin systems. *Phys. Rev. B* **83** (2011), 035107.

[13] X. Chen, Z. C. Gu, Z. X. Liu, and X. G. Wen, Symmetry protected topological orders and the group cohomology of their symmetry group. *Phys. Rev. B* **87** (2013), 155114.

[14] X. Chen, Z. C. Gu, and X. G. Wen, Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order. *Phys. Rev. B* **82** (2010), 155138.

[15] A. Connes, Periodic automorphisms of the hyperfinite factor of type II_1. *Acta Sci. Math.* **39** (1977), 39–66.

[16] S. Doplicher, R. Haag, and J. E. Roberts, Local observables and particle statistics I. *Comm. Math. Phys.* **23** (1971), 199–23.

[17] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated states on quantum spin chains. *Comm. Math. Phys.* **144** (1992), 443–490.

[18] M. Fannes, B. Nachtergaele, and R. F. Werner, Finitely correlated pure states. *J. Funct. Anal.* **120** (1994), 511–534.

[19] L. Fiedler and P. Naaijkens, Haag duality for Kitaev's quantum double model for abelian groups. *Rev. Math. Phys.* **27** (2015), 1550021:1–43.

[20] K. Fredenhagen, K. H. Rehren, and B. Schroer, Superselection sectors with braid group statistics and exchange algebras. *Comm. Math. Phys.* **125** (1989), 201–226.

[21] Z.-C. Gu and X.-G. Wen, Tensor-entanglement-filtering renormalization approach and symmetry-protected topological order. *Phys. Rev. B* **80** (2009), 155131.

[22] M. B. Hastings and T. Koma, Spectral gap and exponential decay of correlations. *Comm. Math. Phys.* **265** (2006), 781–804.

[23] M. B. Hastings and X. G. Wen, Quasi-adiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance. *Phys. Rev. B* **72** (2005), 045141.

[24] V. Jones, Actions of finite groups on the hyperfinite type II_1 factor. *Mem. Amer. Math. Soc.* **28** (1980), no. 237.

[25] A. Kapustin, N. Sopenko, and B. Yang, A classification of phases of bosonic quantum lattice systems in one dimension. 2020, arXiv:2012.15491.

[26] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, Fermionic symmetry protected topological phases and cobordisms. *J. High Energy Phys.* (2015), 1–21.

[27] Y. Kawahigashi, Conformal field theory, Vertex operator algebras and operator algebras. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pp. 2597–2616, World Scientific, 2019.

[28] A. Kitaev, Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303** (2003), 2–30.

[29] E. Lieb, T. Schultz, and D. Mattis, Two soluble models of an antiferromagnetic chain. *Ann. Phys.* **16** (1961), 407–466.

[30] T. Matsui, The split property and the symmetry breaking of the quantum spin chain. *Comm. Math. Phys.* **218** (2001), 393–416.

[31] T. Matsui, Boundedness of entanglement entropy and split property of quantum spin chains. *Rev. Math. Phys.* **25** (2013), 1350017.

[32] A. Molnar, Y. Ge, N. Schuch, and J. I. Cirac, A generalization of the injectivity condition for projected entangled pair states. *J. Math. Phys.* **59** (2018), 021902.

[33] A. Moon and Y. Ogata, Automorphic equivalence within gapped phases in the bulk. *J. Funct. Anal.* **278** (2020), 108422.

[34] P. Naaijkens, Localized endomorphisms in Kitaev's toric code on the plane. *Rev. Math. Phys.* **23** (2011), 347–373.

[35] P. Naaijkens, Haag duality and the distal split property for cones in the toric code. *Lett. Math. Phys.* **101** (2012), 341–354.

[36] P. Naaijkens and Y. Ogata, The split and approximate split property in 2D systems: stability and absence of superselection sectors. 2021, arXiv:2102.07707.

[37] B. Nachtergaele and R. Sims, Lieb–Robinson bounds and the exponential clustering theorem. *Comm. Math. Phys.* **265** (2006), 119–130.

[38] B. Nachtergaele and R. Sims, A multi-dimensional Lieb–Schultz–Mattis theorem. *Comm. Math. Phys.* **276** (2007), 437–472.

[39] B. Nachtergaele and R. Sims, Locality estimates for quantum spin systems. In *New trends in mathematical physics*, edited by V. Sidoravicius, Springer, 2009.

[40] B. Nachtergaele, R. Sims, and A. Young, Quasi-locality bounds for quantum lattice systems. I. Lieb–Robinson bounds, quasi-local maps, and spectral flow automorphisms. *J. Math. Phys.* **60** (2019), 061101.

[41] S. Neshveyev and L. Tuset, *Compact quantum groups and their representation categories*. AMS, 2014.

[42] Y. Ogata, A class of asymmetric gapped Hamiltonians on quantum spin chains and its classification I. *Comm. Math. Phys.* **348** (2016), 847–895.

[43] Y. Ogata, A class of asymmetric gapped Hamiltonians on quantum spin chains and its classification II. *Comm. Math. Phys.* **348** (2016), 897–957.

[44] Y. Ogata, A class of asymmetric gapped Hamiltonians on quantum spin chains and its classification III. *Comm. Math. Phys.* **352** (2017), 1205–1263.

[45] Y. Ogata, A $\mathbb{Z}_2$-index of symmetry protected topological phases with time reversal symmetry for quantum spin chains. *Comm. Math. Phys.* **374** (2020), 705–734.

[46] Y. Ogata, A $\mathbb{Z}_2$-index of symmetry protected topological phases with reflection symmetry for quantum spin chains. *Comm. Math. Phys.* **385** (2021), 1247–1272.

[47] Y. Ogata, A $H^3(G, \mathbb{T})$-valued index of symmetry protected topological phases with on-site finite group symmetry for two-dimensional quantum spin systems. 2021, arXiv:2101.00426.

[48] Y. Ogata, A classification of pure states on quantum spin chains satisfying the split property with on-site finite group symmetries. *Trans. Amer. Math. Soc. Ser. B* **8** (2021), 39–65.

[49] Y. Ogata, A derivation of braided C^*-tensor categories from gapped ground states satisfying the approximate Haag duality. 2021, arXiv:2106.15741.

[50] Y. Ogata, Y. Tachikawa, and H. Tasaki, General Lieb–Schultz–Mattis type theorems for quantum spin chains. *Comm. Math. Phys.* **385** (2021), 79–99.

[51] Y. Ogata and H. Tasaki, Lieb–Schultz–Mattis type theorems for quantum spin chains without continuous symmetry. *Comm. Math. Phys.* **372** (2019), 951–962.

[52] D. Perez-Garcia, M. M. Wolf, M. Sanz, F. Verstraete, and J. I. Cirac, String order and symmetries in quantum spin lattices. *Phys. Rev. Lett.* **100** (2008), 167202.

[53] F. Pollmann, A. Turner, E. Berg, and M. Oshikawa, Entanglement spectrum of a topological phase in one dimension. *Phys. Rev. B* **81** (2010), 064439.

[54] F. Pollmann, A. Turner, E. Berg, and M. Oshikawa, Symmetry protection of topological phases in one-dimensional quantum spin systems. *Phys. Rev. B* **81** (2012), 075125.

[55] Sopenko, An index for two-dimensional SPT states. 2021, arXiv:2101.00801.

[56] K. Yonekura, On the cobordism classification of symmetry protected topological phases. *Comm. Math. Phys.* **368** (2019), 1121–1173.

YOSHIKO OGATA

Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba
Meguro-ku Tokyo 153-8914, Japan, yoshiko@ms.u-tokyo.ac.jp

LIST OF CONTRIBUTORS

Ward, Rachel **7:5140**
Wei, Dongyi **5:3902**
Weiss, Barak **5:3412**
White, Stuart **4:3314**
Wigderson, Avi **2:1392**
Williams, Lauren K. **6:4710**
Willis, George A. **3:1554**
Wittenberg, Olivier **3:2346**
Wood, Melanie Matchett **6:4476**

Xu, Zhouli **4:2768**

Ying, Lexing **7:5154**
Yokoyama, Keita **3:1504**
Young, Robert J. **4:2678**

Zerbes, Sarah Livia **3:1918**
Zhang, Cun-Hui **7:5594**
Zhang, Kaiqing **7:5340**
Zhang, Zhifei **5:3902**
Zheng, Tianyi **4:3340**
Zhou, Xin **4:2696**
Zhu, Chen-Bo **4:3062**
Zhu, Xiaohua **4:2718**
Zhu, Xinwen **3:2012**
Zhuk, Dmitriy **3:1530**
Zograf, Peter **3:2196**
Zorich, Anton **3:2196**

EMS PRESS

https://ems.press

ISBN Set 978-3-98547-058-7

ISBN Volume 5 978-3-98547-063-1