1 Introduction and overview

1.1 The main result and its history

Formally speaking, wave maps are the analogue of harmonic maps where the Minkowski metric is imposed on the independent variables. More precisely, for a smooth $\mathbf{u} : \mathbb{R}^{n+1} \to \mathfrak{M}$ with $(\mathfrak{M}, g)$ Riemannian, define the Lagrangian

$$\mathfrak{L}(\mathbf{u}) := \int_{\mathbb{R}^{n+1}} \left(|\partial_t \mathbf{u}|_g^2 - |\nabla \mathbf{u}|_g^2 \right) dt dx.$$

Then the critical points are defined as $\mathfrak{L}'(\mathbf{u}) = 0$ which means that $\Box \mathbf{u} \perp T_u \mathfrak{M}$ in case $\mathfrak{M}$ is embedded in some Euclidean space. This is called the *extrinsic* formulation, which can also be written as

$$\Box \mathbf{u} + A(\mathbf{u})(\partial_\alpha \mathbf{u}, \partial^\alpha \mathbf{u}) = 0$$

where $A(\mathbf{u})$ is the second fundamental form. In view of this, it is clear that $\gamma \circ \phi$ is a wave map for any geodesic γ in $\mathfrak{M}$ and any free scalar wave ϕ. Moreover, any harmonic map is a stationary wave map. The *intrinsic* formulation is $D^\alpha \partial_\alpha u = 0$, where

$$D_\alpha X^j := \partial_\alpha X^j + \Gamma_{ik}^j \circ \mathbf{u} \, X^i \partial_\alpha u^k$$

is the covariant derivative induced by $\mathbf{u}$ on the pull-back bundle of $T\mathfrak{M}$ under $\mathbf{u}$ (with the summation convention in force). Thus, in local coordinates $\mathbf{u} = (u^1, \ldots, u^d)$ one has

$$\Box u^j + \Gamma_{ik}^j \circ \mathbf{u} \, \partial_\alpha u^i \, \partial^\alpha u^k = 0. \tag{1.1}$$

The central problem for wave maps is to answer the following question:

For which $\mathfrak{M}$ does the Cauchy problem for the wave map $\mathbf{u} : \mathbb{R}^{n+1} \to \mathfrak{M}$ with smooth data $(\mathbf{u}, \dot{\mathbf{u}})|_{t=0} = (\mathbf{u}_0, \mathbf{u}_1)$ have global smooth solutions?

In view of finite propagation speed, one may assume that the data $(\mathbf{u}_0, \mathbf{u}_1)$ are trivial outside of some compact set (i.e., $\mathbf{u}_0$ is constant outside of some compact set, whereas $\mathbf{u}_1$ vanishes outside of that set). Let us briefly describe what is known about this problem.

First, recall that the wave map equation is invariant under the scaling $\mathbf{u} \mapsto \mathbf{u}(\lambda\cdot)$ which is critical relative to $\dot{H}^{\frac{n}{2}}(\mathbb{R}^n)$, whereas the conserved energy

$$\mathcal{E}(\mathbf{u}) = \frac{1}{2}\sum_{\alpha=0}^{n}\int_{\mathbb{R}^n} |\partial_\alpha \mathbf{u}(t,x)|^2\, dx$$

is critical relative to $\dot{H}^1(\mathbb{R}^n)$. In the supercritical case $n \geq 3$ it was observed by Shatah [41] that there are self-similar blow-up solutions of finite energy. In the critical case $n = 2$, it is known that there can be no self-similar blow-up, see [42]. Moreover, Struwe [50] observed that in the equivariant setting, blow-up in this dimension has to result from a strictly slower than self-similar rescaling of a harmonic sphere of finite energy. His arguments were based on the very detailed well-posedness of equivariant wave maps by Christodoulou, Tavildar-Zadeh [4], [5], and Shatah, Tahvildar-Zadeh [44], [45] in the energy class for equivariant wave maps into manifolds that are invariant under the action of $SO(2,\mathbb{R})$. Finally, Rodnianski, Sterbenz [38], as well as the authors together with Daniel Tataru [26] exhibited finite energy wave maps from $\mathbb{R}^{2+1} \to S^2$ that blow up in finite time by suitable rescaling of harmonic maps.

Let us now briefly recall some well-posedness results. The nonlinearity in (1.1) displays a *null-form structure*, which was the essential feature in the subcritical theory of Klainerman–Machedon [16]–[18], and Klainerman–Selberg [20], [21]. These authors proved strong local well-posedness for data in $H^s(\mathbb{R}^n)$ when $s > \frac{n}{2}$. The important critical theory $s = \frac{n}{2}$ was begun by Tataru [64], [63]. These seminal papers proved global well-posedness for smooth data satisfying a smallness condition in $\dot{B}^{\frac{n}{2}}_{2,1}(\mathbb{R}^n) \times \dot{B}^{\frac{n}{2}-1}_{2,1}(\mathbb{R}^n)$. In a breakthrough work, Tao [59], [58] was able to prove well-posedness for data with small $\dot{H}^{\frac{n}{2}} \times \dot{H}^{\frac{n}{2}}$ norm and the sphere as target. For this purpose, he introduced the important *microlocal gauge* in order to remove some "bad" interaction terms from the nonlinearity. Later results by Klainerman, Rodnianski [19], Nahmod, Stephanov, Uhlenbeck [36], Tataru [61], [60], and Krieger [23], [24], [25] considered other cases of targets by using similar methods as in Tao's work.

Recently, Sterbenz and Tataru [47], [48] have given the following very satisfactory answer[1] to the above question: *If the energy of the initial data is smaller than the energy of any nontrivial harmonic map $\mathbb{R}^n \to \mathfrak{M}$, then one has global existence and regularity.*

[1] The conclusions of our work were reached before the appearance of [47], [53].

Notice in particular that if there are no harmonic maps other than constants, then one has global existence for all energies. A particular case of this are the hyperbolic spaces $\mathbb{H}^n$ for which Tao [57]– [53] has achieved the same result (with some a priori global norm control).

The purpose of this book is to apply the method of concentration compactness as in Bahouri, Gérard [1] and Kenig, Merle [14], [15] to the large data wave map problem with the hyperbolic plane $\mathbb{H}^2$ as target. We emphasize that this gives more than global existence and regularity as already in the semilinear case considered by the aforementioned authors. The fact that in the critical case the large data problem should be decided by the geometry of the target is a conjecture going back to Sergiu Klainerman.

Let us now describe our result in more detail. Let $\mathbb{H}^2$ be the upper half-plane model of the hyperbolic plane equipped with the metric $ds^2 = \frac{d\mathbf{x}^2+d\mathbf{y}^2}{\mathbf{y}^2}$. Let $\mathbf{u} : \mathbb{R}^2 \to \mathbb{H}^2$ be a smooth map. Expanding the derivatives $\{\partial_\alpha \mathbf{u}\}_{\alpha=0,1,2}$ (with $\partial_0 := \partial_t$) in the orthonormal frame $\{\mathbf{e}_1, \mathbf{e}_2\} = \{\mathbf{y}\partial_\mathbf{x}, \mathbf{y}\partial_\mathbf{y}\}$ gives rise to smooth coordinate functions $\phi_\alpha^1, \phi_\alpha^2$. In what follows, $\|\partial_\alpha \mathbf{u}\|_X$ will mean $(\sum_{j=1}^2 \|\phi_\alpha^j\|_X^2)^{\frac{1}{2}}$ for any norm $\|\cdot\|_X$ on scalar functions. For example, the energy of $\mathbf{u}$ is

$$E(\mathbf{u}) := \sum_{\alpha=0}^{2} \|\partial_\alpha \mathbf{u}\|_2^2.$$

Next, suppose $\pi : \mathbb{H}^2 \to M$ is a covering map with M some hyperbolic Riemann surface with the metric that renders π a local isometry. In other words, $M = \mathbb{H}^2/\Gamma$ for some discrete subgroup $\Gamma \subset PSL(2,\mathbb{R})$ which operates totally discontinuously on $\mathbb{H}^2$. Now suppose $\mathbf{u} : \mathbb{R}^2 \to M$ is a smooth map which is constant outside of some compact set, say. It lifts to a smooth map $\tilde{\mathbf{u}} : \mathbb{R}^2 \to \mathbb{H}^2$ uniquely, up to composition with an element of Γ. We now define $\|\partial_\alpha \mathbf{u}\|_X := \|\partial_\alpha \tilde{\mathbf{u}}\|_X$. In particular, the energy $E(\mathbf{u}) := E(\tilde{\mathbf{u}})$. Note that due to the fact that Γ is a group of isometries of $\mathbb{H}^2$, these definitions are unambiguous. Our main result is as follows.

Theorem 1.1. *There exists a function $K : (0,\infty) \to (0,\infty)$ with the following property: Let M be a hyperbolic Riemann surface. Suppose $(\mathbf{u}_0, \mathbf{u}_1) : \mathbb{R}^2 \to M \times TM$ are smooth and $\mathbf{u}_0 = const$, $\mathbf{u}_1 = 0$ outside of some compact set. Then the wave map evolution $\mathbf{u}$ of these data as a map $\mathbb{R}^{1+2} \to M$ exists globally as a smooth function and, moreover, for any $\frac{1}{p} + \frac{1}{2q} \leq \frac{1}{4}$ with $2 \leq q < \infty$,*

$\gamma = 1 - \frac{1}{p} - \frac{2}{q}$,

$$\sum_{\alpha=0}^{2} \|(-\Delta)^{-\frac{\gamma}{2}} \partial_\alpha \mathbf{u}\|_{L_t^p L_x^q} \leq C_q \, K(E). \tag{1.2}$$

Moreover, in the case when $M \hookrightarrow \mathbb{R}^N$ is a compact Riemann surface, one has scattering:

$$\max_{\alpha=0,1,2} \left\| \partial_\alpha \mathbf{u}(t) - \partial_\alpha S(t)(f,g) \right\|_{L_x^2} \to 0 \quad \text{as } t \to \pm\infty$$

where $S(t)(f,g) = \cos(t|\nabla|) f + \frac{\sin(t|\nabla|)}{|\nabla|} g$ and suitable $(f,g) \in (\dot{H}^1 \times L^2)(\mathbb{R}^2; \mathbb{R}^N)$. Alternatively, if M is non-compact, then lifting $\mathbf{u}$ to a map $\mathbb{R}^{1+2} \to \mathbb{H}^2$ with derivative components ϕ_α^j as defined above, one has

$$\max_{\alpha=0,1,2} \left\| \phi_\alpha^j(t) - \partial_\alpha S(t)(f^j, g^j) \right\|_{L_x^2} \to 0 \quad \text{as } t \to \pm\infty$$

where $(f^j, g^j) \in (\dot{H}^1 \times L^2)(\mathbb{R}^2; \mathbb{R})$.

We emphasize that (1.2) can be strengthened considerably in terms of the type of norm applied to the Coulomb gauged derivative components of the wave map:

$$\sum_{\alpha=0}^{2} \|\psi_\alpha\|_S^2 \leq C \, K(E)^2 \tag{1.3}$$

The meaning ψ_α as well as of the S norm will be explained below. We now turn to describing this result and our methods in more detail. For more background on wave maps see [13], [61], and [42].

1.2 Wave maps to $\mathbb{H}^2$

The manifold $\mathbb{H}^2$ is the upper half-plane equipped with the metric $ds^2 = \frac{d\mathbf{x}^2 + d\mathbf{y}^2}{\mathbf{y}^2}$. Expanding the derivatives $\{\partial_\alpha \mathbf{u}\}_{\alpha=0,1,2}$ (with $\partial_0 := \partial_t$) of a smooth map $\mathbf{u} : \mathbb{R}^{1+2} \to \mathbb{H}^2$ in the orthonormal frame $\{\mathbf{e}_1, \mathbf{e}_2\} = \{\mathbf{y}\partial_\mathbf{x}, \mathbf{y}\partial_\mathbf{y}\}$ yields

$$\partial_\alpha \mathbf{u} = (\partial_\alpha \mathbf{x}, \partial_\alpha \mathbf{y}) = \sum_{j=1}^{2} \phi_\alpha^j \, \mathbf{e}_j,$$

whence

$$\mathbf{y} = e^{\sum_{j=1,2} \Delta^{-1}\partial_j \phi_j^2}, \quad \mathbf{x} = \sum_{j=1,2} \Delta^{-1}\partial_j(\phi_j^1 \mathbf{y}), \tag{1.4}$$

provided we assume the normalization $\lim_{|x|\to\infty} |\ln \mathbf{y}| = \lim_{|x|\to\infty} |\mathbf{x}| = 0$. Energy conservation takes the form

$$\int_{\mathbb{R}^2} \sum_{\alpha=0}^{2}\sum_{j=1}^{2} \left|\phi_\alpha^j(t,x)\right|^2 dx = \int_{\mathbb{R}^2} \sum_{\alpha=0}^{2}\sum_{j=1}^{2} \left|\phi_\alpha^j(0,x)\right|^2 dx \tag{1.5}$$

where $x = (x_1, x_2)$ and $\partial_0 = \partial_t$. If $\mathbf{u}(t,x)$ is a smooth wave map, then the functions $\{\phi_\alpha^j\}$ for $0 \le \alpha \le 2$ and $j = 1,2$ satisfy the div-curl system

$$\partial_\beta \phi_\alpha^1 - \partial_\alpha \phi_\beta^1 = \phi_\alpha^1 \phi_\beta^2 - \phi_\beta^1 \phi_\alpha^2 \tag{1.6}$$

$$\partial_\beta \phi_\alpha^2 - \partial_\alpha \phi_\beta^2 = 0 \tag{1.7}$$

$$\partial_\alpha \phi^{1\alpha} = -\phi_\alpha^1 \phi^{2\alpha} \tag{1.8}$$

$$\partial_\alpha \phi^{2\alpha} = \phi_\alpha^1 \phi^{1\alpha} \tag{1.9}$$

for all $\alpha, \beta = 0, 1, 2$. As usual, repeated indices are being summed over, and lowering or raising is done via the Minkowski metric. Clearly, (1.6) and (1.7) are integrability conditions which are an expression of the curvature of $\mathbb{H}^2$. On the other hand, (1.8) and (1.9) are the actual wave map system. Since the choice of frame was arbitrary, one still has gauge freedom for the system (1.6)–(1.9). We shall exclusively rely on the Coulomb gauge which is given in terms of complex notation by the functions

$$\psi_\alpha := \psi_\alpha^1 + i\psi_\alpha^2 = (\phi_\alpha^1 + i\phi_\alpha^2) e^{-i\Delta^{-1}\sum_{j=1}^2 \partial_j \phi_j^1}. \tag{1.10}$$

If ϕ_j^1 are Schwartz functions , then $\sum_{j=1}^2 \partial_j \phi_j^1$ has mean zero whence

$$\big(\Delta^{-1}\sum_{j=1}^{2} \partial_j \phi_j^1\big)(z) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \log|z-\zeta| \sum_{j=1}^{2} \partial_j \phi_j^1(\zeta)\, d\zeta \wedge d\bar{\zeta} \tag{1.11}$$

is well-defined and moreover decays like $|z|^{-1}$ (but in general no faster). The gauged components $\{\psi_\alpha\}_{\alpha=0,1,2}$ satisfy the new div-curl system

$$\partial_\alpha \psi_\beta - \partial_\beta \psi_\alpha = i\psi_\beta \Delta^{-1} \sum_{j=1,2} \partial_j(\psi_\alpha^1 \psi_j^2 - \psi_\alpha^2 \psi_j^1) - i\psi_\alpha \Delta^{-1}\partial_j(\psi_\beta^1 \psi_j^2 - \psi_\beta^2 \psi_j^1) \tag{1.12}$$

$$\partial_\nu \psi^\nu = i\psi^\nu \Delta^{-1} \sum_{j=1}^{2} \partial_j (\psi_\nu^1 \psi_j^2 - \psi_\nu^2 \psi_j^1). \tag{1.13}$$

In particular, one obtains the following system of wave equations for the ψ_α:

$$\begin{aligned} \Box \psi_\alpha = {} & i\,\partial^\beta \big[\psi_\alpha \Delta^{-1} \sum_{j=1,2} \partial_j (\psi_\beta^1 \psi_j^2 - \psi_\beta^2 \psi_j^1)\big] \\ & - i\,\partial^\beta \big[\psi_\beta \Delta^{-1} \partial_j (\psi_\alpha^1 \psi_j^2 - \psi_\alpha^2 \psi_j^1)\big] \\ & + i\,\partial_\alpha \big[\psi^\beta \Delta^{-1} \sum_{j=1,2} \partial_j (\psi_\beta^1 \psi_j^2 - \psi_\beta^2 \psi_j^1)\big] \end{aligned} \tag{1.14}$$

Throughout this book we shall only consider *admissible* wave maps $\mathbf{u}$. These are characterized as smooth wave maps $\mathbf{u} : I \times \mathbb{R}^2 \to \mathbb{H}^2$ on some time interval I so that the derivative components ϕ_α^j are Schwartz functions on fixed time slices.

By the method of *Hodge decompositions* from[2] [23]– [25] one exhibits the null-structure present in (1.12)–(1.14). Hodge decomposition here refers to writing

$$\psi_\beta = -R_\beta \sum_{k=1}^{2} R_k \psi_k + \chi_\beta \tag{1.15}$$

where $R_\beta := \partial_\beta |\nabla|^{-1}$ are the usual Riesz transform. Inserting the hyperbolic terms $R_\beta \sum_{k=1}^2 R_k \psi_k$ into the right-hand sides of (1.12)–(1.14) leads to trilinear nonlinearities with a null structure. As is well-known, such null structures are amenable to better estimates since they annihilate "self-interactions", or more precisely, interactions of waves which propagate along the same characteristics, cf. [18]– [17], as well as [20], [21], [11]. Furthermore, inserting at least one "elliptic term" χ_β from (1.15) leads to a higher order nonlinearity, in fact quintic or higher which are easier to estimate (essentially by means of Strichartz norms). To see this, note that

$$\begin{aligned} \sum_{j=1}^{2} \partial_j \chi_j &= 0 \\ \partial_j \chi_\beta - \partial_\beta \chi_j &= \partial_j \psi_\beta - \partial_\beta \psi_j, \end{aligned}$$

[2] In these papers this decomposition is also referred to as "dynamic decomposition".

whence

$$\chi_\beta = i \sum_{j,k=1}^{2} \partial_j \Delta^{-1}\big[\psi_\beta \Delta^{-1}\partial_k(\psi_j^1\psi_k^2 - \psi_k^1\psi_j^2) - \psi_j \Delta^{-1}\partial_k(\psi_\beta^1\psi_k^2 - \psi_k^1\psi_\beta^2)\big]. \tag{1.16}$$

Since we are only going to obtain a priori bounds on ϕ_α^j, it will suffice to assume throughout that the ϕ_α^j are Schwartz functions, whence the same holds for ψ_α. In what follows, we shall never actually *solve the system* (1.12)–(1.14). To go further, the wave-equation (1.14) by *itself is meaningless without assuming the* ψ_α *to satisfy the compatibility relations* (1.12) *and* (1.13). In fact, it is not even clear that (1.12) and (1.13) will hold for all $t \in (-T, T)$ if they hold at time $t = 0$ and (1.14) holds for all $t \in (-T, T)$. Nonetheless, assuming that the ψ_α are defined in terms of the derivative components ϕ_α of a 'sufficiently nice' wave map, it is clear that all three of (1.12)–(1.14) will be satisfied. This being said, we will only use the system (1.14) to derive *a priori estimates* for ψ_α, which will then be shown to lead to suitable bounds on the components ϕ_α^j of derivatives of a wave map $\mathbf{u}$. This is done by means of Tao's device of frequency envelope, see [59] or [23]. This refers to a sequence $\{c_k\}_{k\in\mathbb{Z}}$ of positive reals such that

$$c_k\, 2^{-\sigma|k-\ell|} \le c_\ell \le c_k\, 2^{\sigma|k-\ell|} \tag{1.17}$$

where $\sigma > 0$ is a small number. The most relevant example is given by

$$c_k := \Big(\sum_{\ell\in\mathbb{Z}} 2^{-\sigma|k-\ell|}\, \|P_\ell \psi(0)\|_2^2\Big)^{\frac{1}{2}}$$

which controls the initial data. While it is of course clear that (1.6)–(1.9) imply the system (1.12)–(1.14), the reverse implication is not such a simple matter since it involves solving an elliptic system with large solutions. On the other hand, transferring estimates on the ψ_α in $H^s(\mathbb{R}^2)$ spaces to similar bounds on the derivative components ϕ_α^j does not require this full implication. Indeed, assume the bound $\|\psi\|_{L_t^\infty((-T_0,T_1);H^{\delta_1}(\mathbb{R}^2))} < \infty$ for some small $\delta_1 > 0$ (we will obtain such bounds via frequency envelopes with $0 < \delta_1 < \sigma$). For any fixed time $t \in (-T_0, T_1)$ one now has with P_k being the usual Littlewood–Paley projections to frequency 2^k,

$$\begin{aligned}
\|P_\ell \phi_\alpha\|_{H^{\delta_2}} &= \big\|P_\ell[e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j\phi_j^1}\psi_\alpha]\big\|_{H^{\delta_2}} \\
&\le \big\|P_\ell[P_{<\ell-10}(e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j\phi_j^1})P_{[\ell-10,\ell+10]}\psi_\alpha]\big\|_{H^{\delta_2}} \\
&\quad + \big\|P_\ell[P_{[\ell-10,\ell+10]}(e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j\phi_j^1})P_{<\ell+15}\psi_\alpha]\big\|_{H^{\delta_2}}
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{k>\ell+10} \big\| P_\ell [P_k (e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j \phi_j^1}) P_{k+O(1)}\psi_\alpha] \big\|_{H^{\delta_2}} \\
&\lesssim \| P_{[\ell-10,\ell+10]}\psi_\alpha\|_{H^{\delta_2}} \\
&\quad + \big\| P_{[\ell-10,\ell+10]}(e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j \phi_j^1}) \big\|_{H^{\delta_2}} \| P_{<\ell+15}\psi_\alpha\|_\infty \\
&\quad + \sum_{k>\ell+10} \big\| P_k (e^{i\sum_{j=1}^2 \Delta^{-1}\partial_j \phi_j^1}) \big\|_{H^{\delta_2}} \| P_{k+O(1)}\psi_\alpha\|_\infty
\end{aligned}
$$

Next, one has the bounds

$$
\begin{aligned}
\big\| \nabla_x\, e^{i\Delta^{-1}\sum_{j=1}^2 \partial_j \phi_j^1} \big\|_{L_t^\infty L_x^2} &\lesssim \|\phi_j^1\|_{L_t^\infty L_x^2}, \\
\| P_{<\ell+15}\psi_\alpha\|_{L_x^\infty} &\lesssim 2^{(1-\delta_1)\ell} \|\psi_\alpha\|_{H^{\delta_1}}
\end{aligned}
$$

where the first one is admissible due to *energy conservation for the derived wave map*, see (1.5). In conclusion,

$$
\| P_\ell \phi_\alpha\|_{H^{\delta_2}} \lesssim \| P_{\ell+O(1)}\psi_\alpha\|_{H^{\delta_2}} + 2^{(\delta_2-\delta_1)\ell} \|\phi\|_{L_x^2} \|\psi\|_{H^{\delta_1}}.
$$

Summing over $\ell \geq 0$ yields

$$
\|\phi\|_{L_t^\infty((-T_0,T_1);H^{\delta_2}(\mathbb{R}^2))} < \infty. \tag{1.18}
$$

By the subcritical existence theory of Klainerman and Machedon, see [18]– [16] as well as [20], [21], the solution can now be extended smoothly beyond this time interval. More precisely, the device of frequency envelopes allows one to place the Schwartz data in $H^s(\mathbb{R}^2)$ for all $s > 0$ initially, and as it turns out, also for all times provided $s > 0$ is sufficiently small. The latter claim is of course the entire objective of this book. We should also remark that we bring (1.14) into play only because it fits into the framework of the spaces from [59] and [63]. This will allow us to obtain the crucial energy estimate for solutions of (1.14), whereas it is not clear how to do this directly for the system (1.12), (1.13). As already noted in [23], the price one pays for passing to (1.14) lies with the *initial conditions*, or more precisely, the time derivative $\partial_t \psi_\alpha(0,\cdot)$. While $\psi_\alpha(0,\cdot)$ only involves one derivative of the wave map $\mathbf{u}$, this time derivative involves two. This will force us to essentially "randomize" the initial time.

1.3 The small data theory

In this section we give a very brief introduction to the spaces which are needed to control the ψ system (1.12), (1.13), and (1.14). A systematic development will

be carried out in Chapter 2 below, largely following [58] (we do need to go beyond both [58] and [23] in some instances such as by adding the sharp Strichartz spaces with the Klainerman–Tataru gain for small scales, and by eventually modifying $\|\cdot\|_{S[k]}$ to the stronger $|||\cdot|||_{S[k]}$ which allows for a high-high gain in the $S \times S \to L^2_{tx}$ estimate). First note that it is not possible to bound the trilinear nonlinearities in this system in Strichartz spaces due to slow dispersion in dimension two. Moreover, it is not possible to adapt the $X^{s,b}$-space of the subcritical theory to the scaling invariant case as this runs into logarithmic divergences. For this reason, Tataru [63] devised a class of spaces which resolve these logarithmic divergences. His idea was to allow characteristic frames of reference. More precisely, fix $\omega \in S^1$ and define

$$\theta^{\pm}_{\omega} := (1, \pm\omega)/\sqrt{2}, \quad t_\omega := (t,x)\cdot\theta^{+}_{\omega}, \quad x_\omega := (t,x) - t_\omega\theta^{+}_{\omega},$$

which are the coordinates defined by a generator on the light-cone. Now suppose that ψ_i are free waves such that ψ_1 is Fourier supported on $1 \le |\xi| \le 2$, and both ψ_2 and ψ_3 are Fourier supported on $|\xi| \sim 2^k$ where k is large and negative. Finally, we also assume that the three waves are in "generic position", i.e., that their Fourier supports make an angle of about size one. Clearly, $2^{-k}\psi_1\psi_2\psi_3$ is then a representative model for the nonlinearities arising in (1.14). With

$$\psi_3(t,x) = \int_{\mathbb{R}^2} e^{i[t|\xi|+x\cdot\xi]} f(\xi)\, d\xi$$

we perform the *plane-wave* decomposition $\psi_3(t,x) = \int \phi_\omega(\sqrt{2}t_\omega)\, d\omega$ where

$$\phi_\omega(s) := \int e^{irs} f(r\omega)\, r dr.$$

By inspection,

$$\int \|\phi_\omega\|_{L^2_{t_\omega} L^\infty_{x_\omega}}\, d\omega \lesssim 2^{\frac{k}{2}} \|\psi_3\|_{L^\infty_t L^2_x}. \tag{1.19}$$

Hence,

$$\begin{aligned} 2^{-k}\int \|\phi_\omega\, \psi_1\psi_2\|_{L^1_{t_\omega} L^2_{x_\omega}}\, d\omega &\lesssim 2^{-k}\int \|\phi_\omega\|_{L^2_{t_\omega} L^\infty_{x_\omega}}\, d\omega \|\psi_1\psi_2\|_{L^2_{t_\omega} L^2_{x_\omega}} \\ &\lesssim \|\psi_3\|_{L^\infty_t L^2_x}\, \|\psi_1\|_{L^\infty_t L^2_x} \|\psi_2\|_{L^\infty_t L^2_x} \end{aligned}$$

which is an example[3] of a *trilinear estimate* which will be studied systematically in Chapter 5. Here we used both (1.19) and the standard bilinear L^2_{tx} bilinear

[3] Note that one does not obtain a gain in this case. This fact will be of utmost importance in this book, forcing us to use a "twisted" wave equation resulting from these high-low-low interactions in the linearized trilinear expressions.

L^2-bound for waves with angular separation:

$$\|\psi_1\psi_2\|_{L^2_{t_\omega}L^2_{x_\omega}} = \|\psi_1\psi_2\|_{L^2_t L^2_x} \lesssim 2^{\frac{k}{2}}\|\psi_2\|_{L^\infty_t L^2_x}\|\psi_1\|_{L^\infty_t L^2_x}$$

This suggests introducing an *atomic space* with atoms ψ_ω of Fourier support $|\xi| \sim 1$ and satisfying

$$\|\psi_\omega\|_{L^1_{t_\omega}L^2_{x_\omega}} \leq 1$$

as part of the space $N[0]$ which holds the nonlinearity (the zero here refers to the Littlewood–Paley projection P_0. Below, we refer to this space as NF). In addition, the space defined by (1.19) is also an atomic space and should be incorporated in the space $S[k]$ holding the solution at frequency 2^k (we refer to this below as the PW space). By duality to $L^1_{t_\omega}L^2_{x_\omega}$ in $N[0]$, we then expect to see $L^\infty_{t_\omega}L^2_{x_\omega}$ as part of $S[0]$. The simple observation here (originating in [63]) is that one can indeed bound the energy along a characteristic frame (t_ω, x_ω) of a free wave as long as its Fourier support makes a positive angle with the direction ω. Indeed, recall the local energy conservation identity $\partial_t e - \operatorname{div}(\partial_t\psi\nabla\psi) = 0$ for a free wave where

$$e = \frac{1}{2}(|\partial_t\psi|^2 + |\nabla\psi|^2)$$

is the energy density, over a region of the form $\{-T \leq t \leq T\} \cap \{t_\omega > a\}$. From the divergence theorem one obtains that

$$\int_{t_\omega = a} \chi_{[-T\leq t\leq T]}|\omega^\perp\nabla\psi|^2\, d\mathfrak{L}^2 \lesssim \|\psi\|^2_{L^\infty_t L^2_x}$$

where $\mathfrak{L}^2$ is the planar Lebesgue measure on $\{t_\omega = a\}$. Sending $T \to \infty$ and letting ρ denote the distance between ω and the direction of the Fourier support of $\psi|_{t=0}$, one concludes that

$$\|\psi\|_{L^\infty_{t_\omega}L^2_{x_\omega}} \lesssim \rho^{-1}\|\psi\|_{L^\infty_t L^2_x}.$$

Hence, we should include a piece

$$\sup_{\omega\not\in 2\kappa} d(\omega,\kappa)\|\psi\|_{L^\infty_\omega L^2_{x_\omega}}$$

in the norm $S[0]$ holding $P_0\psi$ provided ψ is a wave packet oriented along the cone of dimensions $1\times 2^k\times 2^{2k}$, projecting onto an angular sector in the ξ-plane associated with the cap $\kappa \subset S^1$, where κ is of size 2^k (this is called NF* below).

Recall that we have made a genericity assumption which guaranteed that the Fourier supports were well separated in the angle. In order to relax this condition,

it is essential to invoke the usual device of *null-forms* which cancel out parallel interactions. One of the discoveries of [23] is a genuinely trilinear null-form expansion, see (5.44) and (5.45), which exploit the relative position of all three waves simultaneously. It seems impossible to reduce the trilinear nonlinearities of (1.14) exclusively to the easier bilinear ones.

It is shown in [63] (and then also in [58] which develops much of the functional framework that we use, as well as [23]) that in low dimensions (especially $n = 2$ but these spaces are also needed for $n = 3$), these null-frame spaces are strong enough – in conjunction with more traditional scaling invariant $X^{s,b}$ spaces – to bound the trilinear nonlinearities, as well as weak enough to allow for an energy estimate to hold. This then leads modulo passing to an appropriate gauge to the small energy theory.

The norm $\|\cdot\|_S$ in (1.3) is of the form $\|\psi\|_S := \big(\sum_{k\in\mathbb{Z}} \|P_k\psi\|^2_{S[k]}\big)^{\frac{1}{2}}$ where $S[k]$ is built from $L^\infty_t L^2_x$, critical $X^{s,b}$, $L^4_t L^\infty_x$ Strichartz norms, as well as the null-frame spaces which we just described.

1.4 The Bahouri–Gérard concentration compactness method

We now come to the core of the argument, namely the Bahouri–Gérard type decomposition and the associated perturbative argument. We remark that independently of and simultaneously with Bahouri, Gérard, Merle and Vega [30] introduced a similar concentration compactness method into the study of nonlinear evolution equations (they considered the L^2-critical nonlinear Schrödinger equation).

In [12] P. Gérard considered defocusing semilinear wave equations in $\mathbb{R}^{3+1}$ of the form $\Box u + f(u) = 0$ with data given by a sequence (ϕ_n, ψ_n) of energy data going weakly to zero. Denote the resulting solutions to the nonlinear problem by u_n, and the free waves with the same data by v_n. Gérard proved that provided $f(u)$ is *subcritical* relative to energy then

$$\|u_n - v_n\|_{L^\infty(I;\mathcal{E})} \to 0, \qquad \text{as } n \to \infty$$

where $\mathcal{E}$ is the energy space. In contrast, for this to hold for the energy critical problem he found via the concentrated compactness method of P. L. Lions that it is necessary and sufficient that $\|v_n\|_{L^\infty(I;L^6(\mathbb{R}^3))} \to 0$. In other words, the critical problem experiences a *loss of compactness*.

The origin of this loss of compactness, as well as the meaning of the L^6 condition were later made completely explicit by Bahouri–Gérard [1]. Their result

reads as follows: *Let* $\{(\phi_n, \psi_n)\}_{n=1}^{\infty} \subset \dot{H}^1 \times L^2(\mathbb{R}^3)$ *be a bounded sequence, and define* v_n *to be a free wave with these initial data. Then there exists a subsequence* $\{v_n'\}$ *of* $\{v_n\}$*, a finite energy free wave* v*, as well as free waves* $V^{(j)}$ *and* $(\varepsilon^{(j)}, x^{(j)}) \in (\mathbb{R}^+, \mathbb{R}^3)^{\mathbb{Z}^+}$ *for every* $j \geq 1$ *with the property that for all* $\ell \geq 1$,

$$v_n'(t,x) = v(t,x) + \sum_{j=1}^{\ell} \frac{1}{\sqrt{\varepsilon_n^{(j)}}} V^{(j)}\left(\frac{t - t_n^{(j)}}{\varepsilon_n^{(j)}}, \frac{x - x_n^{(j)}}{\varepsilon_n^{(j)}}\right) + w_n^{(\ell)}(t,x) \quad (1.20)$$

where

$$\limsup_{n\to\infty} \|w_n^{(\ell)}\|_{L_t^5(\mathbb{R}, L_x^{10}(\mathbb{R}^3))} \to 0 \qquad \text{as } \ell \to \infty,$$

and for any $j \neq k$,

$$\frac{\varepsilon_n^{(j)}}{\varepsilon_n^{(k)}} + \frac{\varepsilon_n^{(k)}}{\varepsilon_n^{(j)}} + \frac{|x_n^{(j)} - x_n^{(k)}| + |t_n^{(j)} - t_n^{(k)}|}{\varepsilon_n^{(j)}} \to \infty, \qquad \text{as } n \to \infty.$$

Furthermore, the free energy E_0 *satisfies the following orthogonality property:*

$$E_0(v_n') = E_0(v) + \sum_{j=1}^{\ell} E_0(V^{(j)}) + E_0(w_n^{(\ell)}) + o(1), \qquad \text{as } n \to \infty.$$

Note that this result characterized the loss of compactness in terms of the appearance of concentration profiles $V^{(j)}$. Moreover, [1] contains an analogue of this result for so-called Shatah–Struwe solutions of the semi-linear problem $\Box u + |u|^4 u = 0$ which then leads to another proof of the main result in [12]. One of the main applications of their work was to show the existence of a function $A : [0,\infty) \to [0,\infty)$ so that every Shatah–Struwe solution satisfies the bound

$$\|u\|_{L_t^5(\mathbb{R}; L_x^{10}(\mathbb{R}^3))} \leq A\big(E(u)\big) \quad (1.21)$$

where $E(u)$ is the energy associated with the semi-linear equation. This is proved by contradiction; indeed, assuming (1.21) fails, one then obtains sequences of bounded energy solutions with uncontrollable Strichartz norm which is then shown to contradict the fact the nonlinear solutions themselves converge weakly to another solution. The decomposition (1.20) compensates for the aforementioned loss of compactness by reducing it precisely to the effect of the *symmetries*, i.e., dilation and scaling. This is completely analogous to the elliptic (in fact, variational) origins of the method of concentration compactness, see Lions [28] and Struwe [49]. See [1] for more details and other applications.

The importance of [1] in the context of wave maps is made clear by the argument of Kenig, Merle [14], [15]. This method, which will be described in more detail later in this section, represents a general method for attacking global well-posedness problems for energy critical equations such as the wave map problem. Returning to the Bahouri–Gérard decomposition, we note that any attempt at implementing this technique for wave maps encounters numerous serious difficulties. These are of course all rooted in the difficult nonlinear nature of the system (1.6)–(1.9). Perhaps the most salient feature of our decomposition, performed in detail in Section 9.2, as compared to [1] is that the free wave equation no longer captures the correct asymptotic behavior for large times; rather, the atomic components $V^{(j)}$ are defined as solutions of a covariant (or "twisted") wave equation of the form

$$\Box + 2iA_\alpha \partial^\alpha \tag{1.22}$$

where the magnetic potential A_α arises from linearizing the wave map equation in the Coulomb gauge. More precisely, the magnetic term here captures the high-low-low interactions in the trilinear nonlinearities of the wave map system where there is no a priori smallness gain. We shall then obtain the concentration profiles via an inductive procedure over increasing frequency scales; in particular, in (1.22) the Coulomb potential A_α (this is a slight misnomer, but the "Coulomb" here and in all other instances where we use this phrase is a reference to the gauge) is defined in terms of lower-frequency approximations which are already controlled, see the next subsection for more details.

In keeping with the Kenig–Merle method, the Bahouri–Gérard decomposition is used to show the following: assume that a uniform bound of the form

$$\|\psi\|_S \leq C(E)$$

for some function $C(E)$ fails for some finite energy levels E. In particular, the set

$$\mathfrak{A} := \{E \in \mathbf{R}_+ | \sup_{\|\psi\|_{L^2_x} \leq E} \|\psi\|_S = \infty\} \neq \emptyset$$

where we loosely denote the energy by $\|\psi\|_{L^2_x} = (\sum_{\alpha=0}^2 \|\psi_\alpha\|^2_{L^2_x})^{\frac{1}{2}}$, and we can then define a number, denoted throughout the rest of this book by E_{crit}, as follows:

$$E_{crit} = \inf_{E \in \mathfrak{A}} E \tag{1.23}$$

Then there must exist a weak wave map $\mathbf{u}_{\text{critical}} : (-T_0, T_1) \to S$ to a compact Riemann surface uniformized by $\mathbb{H}^2$, which enjoys certain compactness properties. In the final part of the argument we then need to rule out the existence of such an object, arriving at an eventual contradiction at the end of the book.

Starting this grand contradiction argument here, we now assume as above that $\mathfrak{A} \neq \emptyset$; this implies that there is a sequence of Schwartz class (on fixed time slices) wave maps $\mathbf{u}^n : (-T_0^n, T_1^n) \times \mathbb{R}^2 \to \mathbb{H}^2$ with the properties that

- $\|\psi^n\|_{L_x^2} \to E_{crit}$,
- $\lim_{n\to\infty} \|\psi^n\|_{S((-T_0^n,T_1^n)\times\mathbb{R}^2)} = \infty$.

Thus all these wave maps have $t = 0$ in their domain of definition. We shall call such a sequence of wave maps *essentially singular*. Roughly speaking, we shall proceed along the following steps. First, recall that the Bahouri–Gérard theorem is a genuine phase-space result in the sense that it identifies the main asymptotic carriers of energy *which are not pure radiation*, which would then sit in $w_n^{(\ell)}$. This refers to the free waves $V^{(j)}$ above, which are "localized" in frequency (namely at scale $(\varepsilon_n^{(j)})^{-1}$) as well as in physical spaces (namely around the space-time points $(t_n^{(j)}, x_n^{(j)})$). The procedure of filtering out the scales $\varepsilon_n^{(j)}$ is due to Metivier–Schochet, see [33].

(1) *Bahouri–Gérard I: Filtering out frequency blocks.*
If we apply the frequency localization procedure of Metivier–Schochet to the derivative components $\phi_\alpha^n = (\frac{\partial_\alpha \mathbf{x}^n}{\mathbf{y}^n}, \frac{\partial_\alpha \mathbf{y}^n}{\mathbf{y}^n})$ of an essentially singular sequence at time $t = 0$, we run into the problem that the resulting frequency components are not necessarily related to an actual map from $\mathbb{R}^2 \to \mathbb{H}^2$. We introduce a procedure to obtain a frequency decomposition which is "geometric", i.e., the frequency localized pieces are themselves derivative components of maps from $\mathbb{R}^2 \to \mathbb{H}^2$. More specifically, in Section 9.2, we start with decompositions

$$\phi_\alpha^n = \sum_{a=1}^{A} \tilde{\phi}_\alpha^{na} + w_\alpha^{nA}, \ \alpha = 0, 1, 2$$

where the $\tilde{\phi}_\alpha^{na}$ are "frequency atoms" obtained from the first stage of the standard Bahouri–Gérard process, see [1]. Here it may be assumed that the frequency scales in the cases $\alpha = 0, 1, 2$ are identical. Since the $\tilde{\phi}_\alpha^{na}$ do not necessarily form the derivative components of admissible maps into $\mathbb{H}^2$, one replaces them by components ϕ_α^{na} which are derivative components of admissible maps, subject to the same frequency scales.

(2) *Refining the considerations on frequency localization; frequency localized approximative maps.* In order to deal with the non-atomic (in the frequency sense) derivative components, which may still have large energy, we need to be able to truncate the derivative components arbitrarily in frequency while still retaining the geometric interpretation. Here we shall use arguments just as in the first step to allow us to "build up" the components ψ_α^n from low fre-

quency ones. In the end, we of course need to show that for some subsequence of the ψ_α^n, the frequency support is essentially atomic. If this were to fail, we deduce an a priori bound on $\|\psi_\alpha^n\|_{S((-T_0^n,T_1^n)\times\mathbb{R}^2)}$. Specifically, we show in Section 9.3 that judicious choice of an interval J, depending on the position of the Fourier support of the frequency atoms ϕ_α^{na} allows us to truncate the components ϕ_α^n to $P_J\phi_\alpha^n$ while retaining their "geometric significance", i.e., the components $P_J\phi_\alpha^{na}$, $\alpha = 0, 1, 2$ are also derivative components of a map up to arbitrarily small errors.

(3) *Assuming the presence of a lowest energy non-atomic type component, establish an a priori estimate for its nonlinear evolution.* More precisely, in Section 9.4, we replace ϕ_α^n by components $\Phi_\alpha^{nA_0^{(0)}}$, which arise by truncating the frequency support of ϕ_α^n to sufficiently low frequencies such that all frequency atoms with energy above a certain threshold are eliminated. In order to obtain a priori bounds on the evolution of the associated Coulomb components $\Psi_\alpha^{nA_0^{(0)}} = \Phi_\alpha^{nA_0^{(0)}} e^{-i\sum_{k=1,2}\Delta^{-1}\partial_k\Phi_k^{nA_0^{(0)}}}$, we use the previous step to approximate the $\Phi_\alpha^{nA_0^{(0)}}$ by frequency truncated $P_{J_j}\Phi_\alpha^{nA_0^{(0)}}$ for judiciously chosen increasing intervals J_j, whose number only depends on the energy E_{crit}. A finite induction procedure then leads to a priori bounds on the $\Psi_\alpha^{nA_0^{(0)}}$, provided n is chosen large enough (only depending on E_{crit}). Here we already encounter the difficulty that the low frequency components appear to interact strongly with the high-frequency components in the nonlinearity, a stark contrast to the defocussing nonlinear critical wave equation. In particular, in order to "bootstrap" the bounds on the differences of the Coulomb potentials associated with the $P_{J_j}\Phi_\alpha^{nA_0^{(0)}}$, we have to invoke energy estimates for covariant wave equations of the form $\Box u + 2i\partial^\nu u A_\nu = 0$.

(4) *Bahouri–Gérard II, applied to the first atomic frequency component.* In Section 9.6, assuming that we have constructed the first "low frequency approximation" $\Phi_\alpha^{nA_0^{(0)}}$ in the previous step, we need to filter out the concentration profiles (analogous to the $V^{(j)}$ at the beginning of this subsection) corresponding to the frequency atoms above the minimum energy threshold and at lowest possible frequency. This is where we have to deviate from Bahouri–Gérard: instead of the free wave operator, we need to use the covariant wave operator $\Box_{A^n} = \Box + 2iA_\nu^n\partial^\nu$ to model the asymptotics as $t \to \pm\infty$, where A_ν^n is the Coulomb potential associated with the low frequency approximation $\Phi_\alpha^{nA_0^{(0)}}$. Thus we obtain the concentration profiles

as weak limits of the data under the covariant wave evolution. Again a lot of effort needs to be expended on showing that the components we obtain are actually the Coulomb derivative components of Schwartz maps from $\mathbb{R}^2 \to \mathbb{H}^2$, up to arbitrarily small errors in energy. Once we have this, we can then use the result from the stability section in order to construct the time evolution of these pieces and obtain their a priori dispersive behavior.

(5) *Bahouri–Gérard II; completion*. Here we repeat Steps 3 and 4 for the ensuing frequency pieces, to complete the estimate for the ψ_α^n. The conclusion is that upon choosing n large enough, we arrive at a contradiction, unless there is precisely one frequency component and precisely one atomic physical component forming that frequency component. These are the data that then gives rise to the weak wave map with the desired compactness properties.

1.5 The Kenig–Merle agument

In [14], [15], Kenig and Merle developed an approach to the global wellposendess for defocusing energy critical semilinear Schrödinger and wave equations; moreover, their argument yields a blow-up/global existence dichotomy in the focusing case as well, provided the energy of the wave lies beneath a certain threshold. See [7] for an application of these ideas to wave maps.

Let us give a brief overview of their argument. Consider

$$\Box u + u^5 = 0$$

in $\mathbb{R}^{1+3}$ with data in $\dot{H}^1 \times L^2$. It is standard that this equation is well-posed for small data provided we place the solution in the energy space intersected with suitable Strichartz spaces. Moreover, if I is the maximal interval of existence, then necessarily $\|u\|_{L_t^8(I;L_x^8(\mathbb{R}^3))} = \infty$ and the energy $E(u)$ is conserved.

Now suppose E_{crit} is the maximal energy with the property that all solutions in the above sense with $E(u) < E_{crit}$ exist globally and satisfy $\|u\|_{L_t^8(\mathbb{R};L_x^8(\mathbb{R}^3))} < \infty$. Then by means of the Bahouri–Gérard decomposition, as well as the perturbation theory for this equation one concludes that a critical solution u_C exists on some interval I^* and that $\|u_C\|_{L_t^8(I^*;L_x^8(\mathbb{R}^3))} = \infty$. Moreover, by similar arguments one obtains the crucial property that the set

$$K := \Big\{ \Big(\lambda^{\frac{1}{2}}(t) u\big(\lambda(t)\big(x - y(t)\big), t\Big), \lambda^{\frac{3}{2}}(t) \partial_t u\Big(\lambda(t)\big(x - y(t)\big), t\Big)\Big) \ : \ t \in I \Big\}$$

is precompact in $\dot{H}^1 \times L^2(\mathbb{R}^3)$ for a suitable path $\lambda(t), y(t)$. To see this, one applies the Bahouri–Gérard decomposition to a sequence u_n of solutions with energy $E(u_n) \to E_{crit}$ from above. The logic here is that due to the minimality assumption on E_{crit} *only a single limiting profile can arise in* (1.20) up to errors that go to zero in energy as $n \to \infty$. Indeed, if this were not the case then due to fact that the profiles diverge from each other in physical space as $n \to \infty$ one can then apply the perturbation theory to conclude that each of the individual nonlinear evolutions of the limiting profiles (which exist due to the fact that their energies are strictly below E_{crit}) can be superimposed to form a global nonlinear evolution, contradicting the choice of the sequence u_n. The fact that $\ell = 1$ allows one to rescale and re-translate the unique limiting profile to a fixed position in phase space (meaning spatial position and spatial frequency) which then gives the desired nonlinear evolution u_C. The compactness follows by the same logic: assuming that it does not hold, one then obtains a sequence $u_C(\cdot, t_n)$ evaluated at times $t_n \in I^*$ converging to an endpoint of I^* such that for $n \neq n'$, the rescaled and translated versions of $u_C(\cdot, t_n)$ and $u_C(\cdot, t_{n'})$ remain at a minimal positive distance from each other in the energy norm. Again one applies Bahouri–Gérard and finds that $\ell = 1$ by the choice of E_{crit} and perturbation theory. This gives the desired contradiction. The compactness property is of course crucial; indeed, for illustrative purposes suppose that u_C is of the form

$$u_C(t,x) = \lambda(t)^{\frac{1}{2}} U\Big(\lambda(t)\big(x - x(t)\big)\Big)$$

where $\lambda(t) \to \infty$ as $t \to 1$, say. Then u_C blows up at time $t = 1$ (in the sense that the energy concentrates at the tip of a cone) and

$$\lambda(t)^{-\frac{1}{2}} u_C\big(\lambda(t)^{-1}x + x(t)\big) = U(x)$$

is compact for $0 \leq t < 1$. Returning to the Kenig–Merle argument, the logic is now to show that u_C acts in some sense like a blow-up solution, at least if I^* is finite in one direction.

The second half of the Kenig–Merle approach then consists of a rigidity argument which shows that a u_C with the stated properties cannot exist. This is done mainly by means of the conservation laws, such as the Morawetz and energy identities. More precisely, the case where I^* is finite at one end is reduced to the self-similar blow-up scenario. This, however, is excluded by reducing to the stationary case and an elliptic analysis which proves that the solution would have to vanish. If I^* is infinite, one basically faces the possibility of stationary solutions which are again shown not to exist.

For the case of wave maps, we follow the same strategy. More precisely, our adaptation of the Bahouri–Gérard decomposition to wave maps into $\mathbb{H}^2$ leads to a critical wave map with the desired compactness properties. In the course of our proof, it will be convenient to project the wave map onto a compact Riemann surface $\mathcal{S}$ (so that we can avail ourselves of the *extrinsic formulation* of the wave map equation). However, it will be important to work simultaneously with this object as well as the lifted one which takes its values in $\mathbb{H}^2$ (since it is for the latter that we have a meaningful well-posedness theory for maps with energy data).

The difference from [14] lies mainly with the rigidity part. In fact, in our context the conservation laws are by themselves not sufficient to yield a contradiction. This is natural, since the geometry of the target will need to play a crucial role. As indicated above, the two scenarios that are lead to a contradiction are the self-similar blow-up supported inside of a light-cone and the stationary weak wave map, which is of course a weakly harmonic map (which cannot exist since the target $\mathcal{S}$ is compact with negative curvature). The former is handled as follows: in self-similar coordinates, one obtains a harmonic map defined on the disk with the hyperbolic metric and with finite energy (the stationarity is derived as in [14]). Moreover, there is the added twist that one controls the behavior of this map at the boundary in the trace sense (in fact, one shows that this trace is constant). Therefore, one can apply the boundary regularity version of Helein's theorem which was obtained by Qing [37]. Lemaire's theorem [27] then yields the constancy of the harmonic map, whence the contradiction (for a version of this argument under the a priori assumption of regularity all the way to the boundary see Shatah–Struwe [42]).

1.6 An overview of the book

The book is essentially divided into two parts: The *modified Bahouri–Gérard method* is carried out in its entirety starting with Chapter 2, and ending with Chapter 9. Indeed, all that precedes Chapter 9 leads to this section, which constitutes the core of this book. The *Kenig–Merle method adapted to Wave Maps* is then performed in the much shorter Chapter 10. We commence by describing in detail the contents of Chapter 2 to Chapter 9.

1.6.1 Preparations for the Bahouri–Gérard process

As explained above, we describe admissible wave maps $u : R^{2+1} \to \mathbb{H}^2$ mostly in terms of the associated Coulomb derivative components ψ_α. Our goals then are to

(1) *Develop a suitable functional framework*, in particular a space-time norm $\|\psi\|_{S(\mathbb{R}^{2+1})}$, together with time-localized versions $\|\psi\|_{S([I\times\mathbb{R}^2)}$ for closed time intervals I, which have the property that

$$\limsup_{I\subset\tilde{I}} \|\psi\|_{S(I\times\mathbb{R}^2)} < \infty$$

for some open interval $\tilde{I}$ implies that the underlying wave map u can be extended smoothly and admissibly beyond any endpoint of $\tilde{I}$, provided such exists.

(2) *Establish an a priori bound* of the form

$$\|\psi\|_{S(I\times\mathbb{R}^2)} \leq C(E)$$

for some function $C : \mathbb{R}_+ \to \mathbb{R}_+$ of the energy E. This latter step will be accomplished by the Bahouri–Gérard procedure, arguing by contradiction.

We first describe (1) above in more detail: in Chapter 2, we introduce the norms $\|\cdot\|_{S[k]}$, $\|\cdot\|_{N[k]}$, $k \in \mathbb{Z}$, which are used to control the frequency localized components of ψ and the nonlinear source terms, respectively. The norm $\|\cdot\|_S$ is then obtained by square summation over all frequency blocks. The basic paradigm for establishing estimates on ψ then is to formulate a wave equation

$$\Box\psi = F$$

or more accurately typically in frequency localized form

$$\Box P_0\psi = P_0 F,$$

and to establish bounds for $\|P_0 F\|_{N[0]}$ which may then be fed into an energy inequality, see Section 2.3, which establishes the link between the S and N-spaces. In order to be able to estimate the nonlinear source terms F, we need to manipulate the right-hand side of (1.14), making extensive use of (1.15). The precise description of the actual nonlinear source terms that we will use for F is actually rather involved, and given in Chapter 3. In order to estimate the collection of trilinear as well as higher order terms, we carefully develop the necessary estimates in Chapters 4, 5, as well as 6. We note that the estimates in [23], while similar,

are not quite strong enough for our purposes, since we need to gain in the largest frequency in case of high-high cascades. This requires us to subtly modify the spaces by comparison to loc. cit. Moreover, the fact that we manage here to build in sharp Strichartz estimates allows us to replace several arguments in [23] by more natural ones, and we opted to make our present account as self-contained as possible.

With the null-form estimates from Chapters 4, 5, 6 in hand, we establish the role of $\|\cdot\|_S$ as a "regularity controlling" device in the sense of (1) above in Chapter 7, see Proposition 7.2. The proof of this reveals a somewhat unfortunate feature of our present setup, namely the fact that working at the level of the differentiated wave map system produces sometimes too many time derivatives, which forces us to use somewhat delicate "randomization" of times arguments. In particular, in the proof of all a priori estimates, we need to distinguish between a "small time" case (typically called Case 1) and a "long time" Case 2, by reference to a fixed frequency scale. In the short time case, one works exclusively in terms of the div-curl system, while in the long-time case, the wave equations start to be essential.

Chapter 7 furthermore explains the well-posedness theory at the level of the ψ_α, see the most crucial Proposition 7.11. We do not prove this proposition in Chapter 7, as it follows as a byproduct of the core perturbative Proposition 9.12 in Chapter 9. Proposition 7.11 and the technically difficult but fundamental Lemma 7.10 allow us to define the "Coulomb wave maps propagation" for a tuple ψ_α, $\alpha = 0, 1, 2$ which are only L^2 functions at time $t = 0$, provided the latter are the L^2-limits of the Coulomb components of admissible maps. Indeed, this concept of propagation is independent of the approximating sequence chosen and satisfies the necessary continuity properties.

We also formulate the concept of a "wave map at infinity" at the level of the Coulomb components, see Proposition 7.15 and the following Corollary 7.16. Again the proofs of these results will follow as a byproduct of the fundamental Proposition 9.12 and Proposition 9.30 in the core Chapter 9.

In Chapter 8, we develop some auxiliary technical tools from harmonic analysis which will allow us to implement the first stage of the Bahouri–Gérard process, namely crystallizing frequency atoms from an "essentially singular" sequence of admissible wave maps. These tools are derived from the embedding $\dot{B}^1_{2,\infty}(\mathbb{R}^2) \to \mathrm{BMO}$ as well as weighted (relative to A_p) Coifman–Rochberg–Weiss commutator bounds.

As mentioned before, Chapter 9 represents the core of this book. In Section 9.2, starting with an *essentially singular* sequence $\mathbf{u}^n$ of admissible wave maps with deteriorating bounds, i.e., $\|\psi^n_\alpha\|_S \to \infty$ as $n \to \infty$ but with the

crucial criticality condition $\lim_{n\to\infty} E(\mathbf{u}^n) = E_{crit}$, we show that the derivative components ϕ_α^n may be decomposed as a sum

$$\phi_\alpha^n = \sum_{a=1}^{A} \phi_\alpha^{na} + w_\alpha^{nA}$$

where the ϕ_α^{na} are derivative components of admissible wave maps which have frequency supports "drifting apart" as $n \to \infty$, while the error w_α^{nA} satisfies

$$\limsup_{n\to\infty} \|w_\alpha^{nA}\|_{\dot{B}^0_{2,\infty}} < \delta,$$

provided $A \geq A_0(\delta)$ is large enough.

In Section 9.3, we then select a number of "principal" frequency atoms ϕ^{na}, $a = 1, 2, \ldots, A_0$, as well as a (potentially very large) collection of "small atoms" ϕ^{na}, $a = A_0 + 1, \ldots, A$. We order these atoms by the frequency scale around which they are supported starting with those of the lowest frequency. The idea now is as follows: under the assumption that there are at least two frequency atoms, or else in case of only one frequency atom that it has energy $< E_{crit}$, we want to obtain a contradiction to the essential criticality of the underlying sequence $\mathbf{u}^n$. To achieve this, we define in Section 9.3 sequences of approximating wave maps, which are essentially obtained by carefully truncating the initial data sequence ϕ^{na} in frequency space.

In Section 9.4, we establish an a priori bound for the lowest frequency approximating map which comprises all the minimum frequency small atoms as well as the component of the small Besov error of smallest frequency, see Proposition 9.9. The proof of this follows again by truncating the data suitably in frequency space, and applying an inductive procedure to a sequence of approximating wave maps. This hinges crucially on the core perturbative result Proposition 9.12, which plays a fundamental role in this book. The main technical difficulty encountered in the proof of the latter comes from the issue of *divisibility*: let us be given a schematically written expression

$$\partial^\nu \epsilon A_\nu$$

which is *linear* in the perturbation (so that we cannot perform a bootstrap argument based solely on the smallness on ϵ itself), while A_ν denotes some null-form depending on a priori controlled components ψ. "Divisibility" means the property that upon suitably truncating time into finitely many intervals I_j *whose number only depends on* $\|\psi\|_S$, one may bound the expression by

$$\|\partial^\nu \epsilon A_\nu\|_{N(I_j\times\mathbb{R}^2)} \ll \|\epsilon\|_S.$$

In other words, by shrinking the time interval, we ensure that we can iterate the term away. While this would be straightforward provided we had an estimate for $\|A_\nu\|_{L^1_t L^\infty_x}$ (which is possible in space dimensions $n \geq 4$), in our setting, the spaces are much too weak and complicated. Our way out of this impasse is to build those terms for which we have no obvious divisibility into the linear operator, and thereby form a new operator

$$\Box_A \epsilon := \Box \epsilon + 2i\,\partial^\nu \epsilon A_\nu$$

with a magnetic potential term. Fortunately, it turns out that if A_ν is supported at much lower frequencies than ϵ (which is precisely the case where divisibility fails), one can establish an approximate energy conservation result, which in particular gives a priori control over a certain constituent of $\|\cdot\|_S$. With this in hand, one can complete the bootstrap argument, and obtain full control over $\|\epsilon\|_S$.

Having established control over the lowest-frequency "essentially non-atomic" approximating wave map in Section 9.4, we face the task of "adding the first large atomic component", ϕ^{n1}. It is here that we have to depart crucially from the original method of Bahouri–Gérard: instead of studying the free wave evolution of the data, we extract concentration cores by applying the "twisted" covariant evolution associated with

$$\Box_{A^n} u = 0,$$

which is essentially defined as above. The key property that makes everything work is an almost exact energy conservation property associated with its wave flow. This is a rather delicate point, and uses the Hamiltonian structure of the covariant wave flow.

It then requires a fair amount of work to show that the profile decomposition at time $t = 0$ in terms of covariant free waves is "geometric", in the sense that the concentration profiles can indeed by approximated by the Coulomb components of admissible maps, up to a constant phase shift, see Proposition 9.24.

Finally, in Proposition 9.30 we show that we may evolve the data including the first large frequency atom, provided all concentration cores have energy strictly less than E_{crit}.

As most of the work has been done at this point, adding on the remaining frequency atoms in Section 9.9 does not provide any new difficulties, and can be done by the methods of the preceding sections.

In conjunction with the results of Chapter 7, we can then infer that given an essentially singular sequence of wave maps $\mathbf{u}^n$, we may select a subsequence of them whose Coulomb components ψ^n_α, up to re-scalings and translations, converge to a limiting object $\Psi^\infty_\alpha(t, x)$, which is well-defined on some interval $I \times \mathbb{R}^2$

where I is either a finite time interval or (semi)-infinite, and the limit of the Coulomb components of admissible maps there. Moreover, most crucially for the sequel, $\Psi_\alpha^\infty(t,x)$ satisfies a remarkable *compactness property*, see Proposition 9.36. This sets the stage for the method of Kenig–Merle, which we adapt to the context of wave maps in Chapter 10.