Mathematisches Forschungsinstitut Oberwolfach

Report No. 41/2025

DOI: 10.4171/OWR/2025/41

New Mathematical Directions in Coding Theory

Organized by
Noga Ron-Zewi, Haifa
Mary Wootters, Stanford

7 September – 12 September 2025

Abstract. Coding theory is concerned with the design and analysis of error-correcting codes, a method for protecting data from noise or corruption. In addition to their wide practical applicability, error-correcting codes are also supported by a rich theory, with connections to diverse disciplines in mathematics, science, and engineering. This workshop focused on exciting mathematical methods in the design of error-correcting codes, including high-dimensional expanders, convex optimization, and structured random ensembles. These methods have led to recent breakthroughs in coding theory. The goal of this workshop was to bring together researchers from multiple communities to exchange ideas around these and other mathematical techniques, hopefully leading to further advances.

Mathematics Subject Classification (2020): 94Bxx.

Introduction by the Organizers

The workshop *New Mathematical Directions in Coding Theory*, organized by Mary Wootters (Stanford) and Noga Ron-Zewi (Haifa) gathered more than 50 participants from a broad range of institutions worldwide. The event brought together researchers from diverse backgrounds in mathematics, computer science, and electrical engineering, all united by their interest in the modern theory of error-correcting codes. The workshop provided an engaging setting that fostered lively discussions, exchange of ideas, and exploration of emerging mathematical approaches in coding theory.

Coding theory is concerned with the design and analysis of error-correcting codes, which enable the reliable transmission and storage of information in the presence of noise. Beyond their practical importance, such codes exhibit deep mathematical structures and connections to algebra, geometry, probability theory, and graph theory. The workshop focused on emerging mathematical directions that have recently advanced the field, including approaches based on high-dimensional expanders, combinatorics, graph theory, and structured random ensembles. By emphasizing these modern mathematical tools, the workshop aimed to bridge perspectives across different subfields and to stimulate new collaborations.

The scientific program featured a mix of survey and research talks, highlighting recent advances and open problems. The sessions covered topics such as list decoding and list recovery of algebraic and random codes, connections between coding theory and complexity, graph-based and geometric constructions, as well as emerging applications in quantum computing and DNA storage. Open problem sessions and informal discussions complemented the formal program, creating opportunities for participants to engage in deeper exchanges and to initiate new research directions. The week also included a traditional Oberwolfach hike and social gatherings, which contributed to a collegial and stimulating atmosphere.

The remainder of this report presents summaries of the talks delivered during the workshop. They highlight not only recent advances across a wide range of topics in coding theory but also open questions and emerging lines of research that point toward exciting future developments.

Acknowledgement: The MFO and the workshop organizers would like to thank the National Science Foundation for supporting the participation of junior researchers in the workshop by the grant DMS-2230648, “US Junior Oberwolfach Fellows”.

Workshop: New Mathematical Directions in Coding Theory

Table of Contents

Abstracts

Recent Advances in the Combinatorial List Decodability of Algebraic Codes (Survey Talk)

Zeyu Guo

List decoding, first introduced by Elias and Wozencraft in the late 1950s, is a central notion in coding theory. In the combinatorial setting, one asks how many codewords can lie in a Hamming ball of a given radius: a code is said to be (ρ, L) list decodable if every ball of radius ρn contains at most L codewords. This captures the tradeoff between rate, error tolerance, and list size, without regard to algorithmic efficiency. A useful strengthening is the average-radius notion, which instead requires that the average distance from the center to any $L+1$ codewords exceeds ρn.

For a Reed–Solomon (RS) code of rate R and relative distance $1-R$, the classical Johnson bound guarantees list decodability up to radius $1-\sqrt{R}$. On the other hand, the list-decoding capacity theorem shows that a random code of rate R can tolerate up to $1-R-\varepsilon$ fraction errors with list size $O(1/\varepsilon)$. Whether random RS codes can approach this limit remained an important open problem until recently.

1. Generalized Singleton Bound

The recent line of progress begins with Shangguan and Tamo [1]. They proved a *generalized Singleton bound*, later refined by Roth [2]. This bound states that if a code of rate R is (average-radius) (ρ, L) list decodable, then

$$\rho \leq \tfrac{L}{L+1}(1-R).$$

This generalizes the classical Singleton bound ($L=1$) and provides a sharp limitation on list decoding. Moreover, setting $L = \Theta(1/\varepsilon)$ recovers the list-decoding capacity $1-R-\varepsilon$. Shangguan and Tamo [1] conjectured that random RS codes attain this bound.

2. Resolution by Brakensiek–Gopi–Makam

Subsequent results [3, 4, 5] improved the known bounds on the combinatorial list decodability of random RS codes, culminating in the work of Brakensiek, Gopi, and Makam [5], which confirmed Shangguan and Tamo's conjecture. They proved that with high probability random RS codes meet the generalized Singleton bound for every fixed L over large fields. In doing so, they introduced the framework of *higher-order MDS codes* and showed the equivalence of several formulations (MDS(L), GZP(L), LD-MDS(L)). A central ingredient was the *GM–MDS theorem*, proved independently by Lovett [6] and Yildiz and Hassibi [7]. Thus, generic RS codes are optimally list decodable in the average-radius sense, matching the best possible guarantees.

3. Field Size Improvements

A key limitation of the results in [1, 3, 5] was the requirement of exponentially large alphabets. Guo–Zhang [8] reduced the necessary field size to polynomial in n, at the cost of slightly relaxing the decoding radius to $\frac{L}{L+1}(1-R-\varepsilon)$. They showed that random RS codes over fields of size $\Theta_{L,\varepsilon}(n^2)$ are still nearly optimally list decodable with high probability. Later, Alrabiah, Guruswami, and Li [9] improved the analysis of [8] and proved that random RS codes over fields of linear size $\Theta_{L,\varepsilon}(n)$ already achieve the same list decodability. The same paper also proved that random linear codes over fields of size $\Theta_{L,\varepsilon}(1)$ achieve the same guarantee.

4. Extensions to Other Codes

The ideas developed for random Reed–Solomon codes extend to other algebraic families. For algebraic geometry (AG) codes, Brakensiek, Dhar, Gopi, and Zhang [10] showed that random AG codes over small fields can also achieve the generalized Singleton bound, exploiting curves with many rational points. In the rank metric, Gabidulin codes play the analogous role of RS codes. Recent work by Guo, Xing, Yuan, and Zhang [11, 12] established that random Gabidulin codes are list decodable up to the generalized Singleton radius, with polynomial or linear extension degree depending on whether one allows a small slack.

Explicit constructions are also advancing. Folded RS codes and multiplicity codes, which are known to achieve list-decoding capacity [13, 14], have seen steady improvements in their combinatorial list size bounds [15, 16, 17]. In particular, Srivastava [16] and Chen and Zhang [17] independently showed that folded RS and multiplicity codes achieve the list-decoding radius $\frac{L}{L+1}(1-R-\varepsilon)$ with list size $(L-1)^2+1$ and L, respectively. The latter result essentially matches the generalized Singleton bound.

5. Summary and Open Problems

The combinatorial list decodability of Reed–Solomon codes and other algebraic codes is now much better understood. The remaining challenges include:

- Extending the results on list decodability of algebraic codes to the more general setting of list recoverability, with better trade-offs between parameters.
- Constructing better explicit higher-order MDS codes. Current constructions [1, 2, 18] require doubly exponential field sizes in n or Lk, where n is the block length, L is the list size, and k is the dimension of the code.

In addition, it would be interesting to streamline the current proofs of GM–MDS-type theorems [6, 7, 19, 11], and to find further applications of such theorems.

References

[1] C. Shangguan and I. Tamo, *Combinatorial list-decoding of Reed-Solomon codes beyond the Johnson radius*, Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (2020), 538–551.

[2] R. M. Roth, *Higher-order MDS codes*, IEEE Transactions on Information Theory **68**.12 (2022), 7798–7816.

[3] Z. Guo, R. Li, C. Shangguan, I. Tamo, and M. Wootters, *Improved list-decodability and list-recoverability of Reed–Solomon codes via tree packings*, SIAM Journal on Computing **53**.2 (2024), 389–430.

[4] A. Ferber, M. Kwan, and L. Sauermann, *List-decodability with large radius for Reed-Solomon codes*, IEEE Transactions on Information Theory, **68(6)** (2022), 3823–3828.

[5] J. Brakensiek, S. Gopi, and V. Makam, *Generic Reed–Solomon codes achieve list-decoding capacity*, SIAM Journal on Computing (2024), STOC23-118–STOC23-154.

[6] S. Lovett, *MDS matrices over small fields: A proof of the GM-MDS conjecture*, 2018 IEEE 59th Annual Symposium on Foundations of Computer Science (2018), 194–199.

[7] H. Yildiz and B. Hassibi, *Optimum linear codes with support-constrained generator matrices over small fields*, IEEE Transactions on Information Theory **65(12)** (2019), 7868–7875.

[8] Z. Guo and Z. Zhang, *Randomly punctured Reed-Solomon codes achieve the list decoding capacity over polynomial-size alphabets*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (2023), 164–176.

[9] O. Alrabiah, V. Guruswami, and R. Li, *Randomly punctured Reed–Solomon codes achieve list-decoding capacity over linear-sized fields*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 1458–1469.

[10] J. Brakensiek, M. Dhar, S. Gopi, and Z. Zhang, *AG codes achieve list decoding capacity over constant-sized fields*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 740–751.

[11] Z. Guo, C. Xing, C. Yuan, and Z. Zhang, *Random Gabidulin codes achieve list decoding capacity in the rank metric*, IEEE 65th Annual Symposium on Foundations of Computer Science (2024), 1846–1873.

[12] Z. Guo, C. Xing, C. Yuan, and Z. Zhang, *Gabidulin Codes Achieve List Decoding Capacity with an Order-Optimal Column-To-Row Ratio*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025), Leibniz International Proceedings in Informatics (LIPIcs) **353** (2025), 43:1–43:20.

[13] V. Guruswami and A. Rudra, *Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy*, IEEE Transactions on information theory **54(1)** (2008), 135–150.

[14] S. Kopparty, *List-decoding multiplicity codes*, Theory of Computing **11(1)** (2015), 149–182.

[15] I. Tamo, *Tighter list-size bounds for list-decoding and recovery of folded Reed-Solomon and multiplicity codes*, IEEE Transactions on Information Theory **70(12)** (2024), 8659–8668.

[16] S. Srivastava, *Improved list size for folded Reed-Solomon codes*, Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (2025), 2040–2050.

[17] Y. Chen and Z. Zhang, *Explicit folded Reed-Solomon and multiplicity codes achieve relaxed generalized Singleton bounds*, Proceedings of the 57th Annual ACM Symposium on Theory of Computing (2025), 1–12.

[18] J. Brakensiek, M. Dhar, and S. Gopi, *Improved field size bounds for higher order MDS codes*, 2023 IEEE International Symposium on Information Theory (ISIT) (2023), 1243–1248.

[19] J. Brakensiek, M. Dhar, and S. Gopi, *Generalized GM-MDS: Polynomial codes are higher order MDS*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 728–739.

Fast List-Decoding and List-Recovery of Univariate Multiplicity and FRS Codes

Prahladh Harsha

(joint work with Rohan Goyal, Mrinal Kumar, and Ashutosh Shankar)

We show that the known list-decoding and list-recovery algorithms for univariate multiplicity and folded Reed–Solomon (FRS) codes can be made to run in $\tilde{O}(n)$ time.

Univariate multiplicity codes and FRS codes are natural variants of Reed–Solomon codes that were discovered and studied for their applications to list decoding. It is known that for every $\epsilon > 0$, and rate $r \in (0,1)$, there exist explicit families of these codes that have rate r and can be list decoded from a $(1-r-\epsilon)$ fraction of errors with constant list size in polynomial time [1, 2]. In this talk based on [3], we present randomized algorithms that perform the above list-decoding tasks in $\tilde{O}(n)$, where n is the block-length of the code. We also prove similar nearly-linear time for the list-recovery algorithms.

Our algorithms have two main components. The first component builds upon the lattice-based approach in [4], where a $\tilde{O}(n)$ time list-decoding algorithm for Reed–Solomon codes approaching the Johnson radius was designed. As part of the second component, we design $\tilde{O}(n)$ time algorithms for two natural algebraic problems: given a $(m+2)$-variate polynomial $Q(x, y_0, \ldots, y_m) = \tilde{Q}(x) + \sum_{i=0}^{m} Q_i(x) \cdot y_i$ the first algorithm solves order-m linear differential equations of the form $Q\left(x, f(x), \frac{df}{dx}, \ldots, \frac{d^m f}{dx^m}\right) \equiv 0$ while the second solves functional equations of the form $Q\left(x, f(x), f(\gamma x), \ldots, f(\gamma^m x)\right) \equiv 0$, where m is an arbitrary constant and γ is a field element of sufficiently high order.

References

[1] V. Guruswami and C. Wang, *Linear-Algebraic List Decoding for Variants of Reed–Solomon Codes*, IEEE Transactions on Information Theory **59(6)** (2013), 3257–3268.

[2] S. Kopparty, N. Ron-Zewi, S. Saraf, and M. Wootters, *Improved list decoding of folded Reed–Solomon and multiplicity codes*, SIAM Journal on Computing **52(3)** (2023), 794–840.

[3] Rohan Goyal, Prahladh Harsha, Mrinal Kumar, and Ashutosh Shankar, *Fast list-decoding of univariate multiplicity and folded Reed–Solomon codes*, Proc. 65th IEEE Symp. on Foundations of Comp. Science (FOCS 2024), 328–343.

[4] M. Alekhnovich, *Linear diophantine equations over polynomials and soft decoding of Reed–Solomon codes*, IEEE Transactions on Information Theory **51(7)** (2005), 2257–2265.

Decoding Graph-Based Codes (Survey Talk)

Shashank Srivastava
(joint work with Madhur Tulsiani)

Introduction and Background

We consider codes $C \subseteq \Sigma^n$ that encode k symbols into n symbols so that their blocklength is n, and the rate is $R = k/n$. The minimum Hamming distance $\delta(C)$, normalized by the blocklength n, between distinct codewords is a measure of the code's error tolerance. The Singleton bound says that for a rate R code, its distance δ can be at most $1 - R$. The well known Reed-Solomon codes achieve this bound, and classical unique decoding algorithms such as Berlekamp-Welch can correct up to $(1 - R)/2$ errors in near-linear time.

List decoding generalizes unique decoding: the decoder outputs all codewords within a certain distance of the received word, which is crucial when the error fraction is large (approaching $1 - R$). For many algebraic codes (e.g., folded Reed-Solomon codes, univariate multiplicity codes), efficient list decoding algorithms exist up to the optimal radius $1 - R - \varepsilon$ [1, 2]. A recent line of work [3] has shown that the list size can be $O(1/\varepsilon)$, and this is optimal up to constant factors.

Graph-Based Codes and Expansion

Most progress in list decoding has focused on algebraic codes. In contrast, **graph-based codes**, in particular those constructed using spectral expander graphs, offer several advantages, such as efficient, parallelizable decoders, beating the Zyablov bound, c^3-LTCs, etc. The following are three well-studied instances:

- **Tanner codes**: Codewords sit on edges, subject to local codes on vertex neighborhoods. First explicit construction of LDPC codes [4, 5].
- **Alon-Edmonds-Luby (AEL) codes**: Use spectral expanders to amplify distance while causing near-optimal drop in rate. Approach Singleton bound with alphabet size independent of blocklength, and linear time encoding and unique decoding [6, 7].
- **Ta-Shma's codes**: Explicit binary codes approaching Gilbert-Varshamov bound [8].

The main message of this talk is that expansion properties not only help with code's structural guarantees, but also enable new algorithmic techniques for decoding. These algorithms have common themes, and make more direct use of expansion than erstwhile unique decoders for these algorithms.

List Decoding Algorithms for Graph-Based Codes

Typical unique decoders work by reducing the global decoding problem to many, many small decoding problems that are each much smaller. Typically, this smaller problem is simply to decode locally in the neighborhood of a vertex, and with sparse graphs, this is a constant sized problem. Thus, efficiency is not a concern for these local instances of the decoding problem. In the unique decoding regime, most of these local decoding problems succeed in finding a component of the true global solution, and the remaining error can be made arbitrarily small based on the quality of expansion.

The key difficulty with list decoding is that while this reduction to many different small decoding problems still works, the potential presence of two different solutions means that the components of both of these global solutions will be represented in the outputs of these smaller problems. In other words, one is left with a small *list* of solutions for each local problem, such that any true global solution is represented in most of these lists, but we do not know which item of the list corresponds to a particular codeword.

Our Solution: Regularity Lemmas

Nevertheless, we show how to stitch these lists into a consistent global solution. These combinatorial tools can be seen as forms of low-rank projection of a graph, that allow for efficient estimation of cut sizes of arbitrary graphs.

A couple of issues arise. Regularity lemmas are typically used for dense graphs, but we can use the fact that our sparse graphs are dense subgraphs of expanders to extract a dense model [9]. Secondly, regularity lemmas are combinatorial tools, but we would like to design algorithms using them. We show that a semidefinite programming based approximation algorithm of Alon and Naor [10] for the cut-norm problem can be used to efficiently build regularity lemma decompositions, even for dense subgraphs of sparse expanders.

The algorithmic step above can be implemented in near-linear time, and the regularity decompositon can be used to efficiently combine local lists into global candidates. This also gives a combinatorial proof for the list size being small beyond the Johnson bound. For AEL codes, where list size is known to be near-optimal due to [11], we show the following,

Main Result: A near-linear time algorithm to list decode AEL codes up to the optimal radius $1 - R - \varepsilon$, with list size and runtime depending only polynomially on $1/\varepsilon$ (not on blocklength).

The SoS Approach. Sum-of-Squares (SoS) hierarchy is another algorithmic technique that is applicable for decoding several of the expander-based codes. These algorithms are typically used for constraint satisfaction problems, but can be adapted to work for codes. However, SoS algorithms are typically much slower than regularity lemma-based algorithms, which also often provide additional combinatorial insights.

Implications for other codes.

- **Tanner Codes**: Can be list decoded up to their designed distance, matching tensor codes which are their dense graph analogs. Regularity lemmas are the only known way to control list size for such radii.
- **AEL Codes**: Achieve list decoding and list recovery up to capacity in near-linear time using regularity lemmas.

- **Ta-Shma's Codes**: Can be list decoded up to a radius approaching 1/2 in near-linear time for binary codes in near-linear time, and even better up to Johnson bound, but with slower Sum-of-Squares algorithms.

Future Directions

Graph-based codes are the only known way to get linear-time unique decoders. Therefore, these AEL based codes become natural candidates to achieve linear time list decoding up to capacity, a long-standing open problem.

From a technique point of view, the stitching problem for local lists is relevant in other areas such as heavy hitters, compressed sensing, and group testing. The regularity lemma approach may offer new tools in these domains.

References

[1] V. Guruswami and A. Rudra, *Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy*, IEEE Transactions on information theory **54(1)** (2008), 135–150.

[2] S. Kopparty, *List-decoding multiplicity codes*, Theory of Computing **11(1)** (2015), 149–182.

[3] Y. Chen and Z. Zhang, *Explicit folded Reed-Solomon and multiplicity codes achieve relaxed generalized Singleton bounds*, Proceedings of the 57th Annual ACM Symposium on Theory of Computing (2025), 1–12.

[4] R. Tanner, *A recursive approach to low complexity codes*, IEEE Transactions on Information Theory **27(5)** (1981), 533–547.

[5] M. Sipser and D. Spielman, *Expander codes*, IEEE Transactions on Information Theory **42(6)** (1996), 1710–1722.

[6] N. Alon, J. Edmonds, and M. Luby, *Linear time erasure codes with nearly optimal recovery*, Proceedings of IEEE 36th Annual Foundations of Computer Science (1995), 512–519.

[7] V. Guruswami and P. Indyk, *Linear-time encodable/decodable codes with near-optimal rate*, IEEE Transactions on Information Theory **51(10)** (2005), 3393–3400.

[8] A. Ta-Shma, *Explicit, almost optimal, epsilon-balanced codes*, Proceedings of the 49th ACM Symposium on Theory of Computing (2017).

[9] L. Trevisan, M. Tulsiani, and S. Vadhan, *Boosting, regularity and efficiently simulating every high-entropy distribution*, Proceedings of the 24th IEEE Conference on Computational Complexity (2009).

[10] N. Alon and A. Naor, *Approximating the cut-norm via grothendieck's inequality*, Proceedings of the 36th ACM Symposium on Theory of Computing (2004), 72–80.

[11] F. G. Jeronimo, T. Mittal, S. Srivastava, and M. Tulsiani, *Explicit codes approaching generalized Singleton bound using expanders*, Proceedings of the 57th ACM Symposium on Theory of Computing (2025).

Expander Graphs and Optimally List-Decodable Codes

Madhur Tulsiani

(joint work with Fernando G. Jeronimo, Tushant Mittal, Shashank Srivastava)

We discuss a new family of explicit codes that are list decodable to capacity and achieve an optimal list size of $O(1/\epsilon)$. In contrast to existing explicit constructions of codes achieving list decoding capacity, our arguments do not rely on algebraic structure but utilize simple combinatorial properties of expander graphs.

Our construction is based on a celebrated distance amplification procedure due to Alon, Edmonds, and Luby [1], which transforms any high-rate code into one with near-optimal rate-distance tradeoff. We generalize it to show that the same procedure can be used to transform any high-rate code into one that achieves list decoding capacity. Our proof can be interpreted as a "local-to-global" phenomenon for (a slight strengthening of) the generalized Singleton bound.

As a corollary of our result, we also obtain the first explicit construction of LDPC codes achieving list decoding capacity, and in fact arbitrarily close to the generalized Singleton bound.

Open Problems.

- Can expander graphs be used to obtain similar local-to-global results for other code properties?
- Is there an analogue of this list decodability result for quantum LDPC codes?

References

[1] N. Alon, J. Edmonds, and M. Luby. *Linear time erasure codes with nearly optimal recovery*, Proceedings of IEEE 36th Annual Foundations of Computer Science (1995), 512–519.

Polynomials over Grids: Local Testing and Decoding (Survey Talk)

Madhu Sudan

(joint work with Prashanth Amireddy, Mitali Bafna, Amik Raj Behara, Manaswi Paraashar, and Srikanth Srinivasan)

It has long been known that evaluations of low-degree multivariate polynomials over product sets leads to a class of functions with combinatorially nice error-correcting properties. However the algorithmic ability to correct these polynomials were only traditionally known when the evaluation domain was a vector space. About ten years back Kim and Kopparty [1] highlighted this question and gave the first global decoding algorithms correcting polynomials over product sets upto half the minimum distance of the underlying code. This result inspired our sequence of works [2, 3, 4, 5, 6, 7] that explored local testing and correcting of polynomials over product sets (of the form $S_1 \times \cdots S_n$) with grids ($S_i = S \forall i$) and hypercube ($S_i = \{0, 1\}$) being special cases.

In this talk we surveyed the results from our joint works that reveal the following: (1) Unlike the vector space setting local decoding can not be done with $O_{d,s}(1)$ queries where d is the degree of the polynomial and $s = \max_i\{S_i\}$. This result even holds for $d = 1$ and the domain being the hypercube. (2) In contrast, local testing can be performed with $O_{d,s}(1)$ queries over grids, but not over general product sets [2, 3]. (3) The class of degree d polynomials in n variables can be locally corrected with $(\log n)^{O_d(1)}$ queries over the hypercube and even list-decoded upto its minimum distance [4, 5]. (4) The local correcting results can be extended to the setting of general grids, but now only correcting some $\Omega_{d,s}(1)$ fraction of errors [7]. A concept that is highlighted by these works is the role of junta-degree of functions.

References

[1] John Y. Kim and Swastik Kopparty, *Decoding reed-muller codes over product sets*, Theory Comput. **13(1)** (2017), 1–38.

[2] Mitali Bafna, Srikanth Srinivasan, and Madhu Sudan, *Local decoding and testing of polynomials over grids*, Random Struct. Algorithms **57(3)** (2020), 658–694.

[3] Prashanth Amireddy, Srikanth Srinivasan, and Madhu Sudan, *Low-degree testing over grids*, Approximation, Randomization, and Combinatorial Optimization, APPROX/RANDOM (2023), 41:1–41:22.

[4] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan, *Local correction of linear functions over the boolean cube*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC (2024), 764–775.

[5] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan, *Low degree local correction over the boolean cube*, Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA (2025), 5504–5511.

[6] Prashanth Amireddy, Amik Raj Behera, Srikanth Srinivasan, and Madhu Sudan, *A near-optimal polynomial distance lemma over boolean slices*, Proceedings of the 52nd International Colloquium on Automata, Languages, and Programming, ICALP (2025), 11:1–11:17.

[7] Prashanth Amireddy, Amik Raj Behera, Srikanth Srinivasan, and Madhu Sudan, *Eigenvalue bounds for symmetric markov chains on multislices with applications*, Approximation, Randomization, and Combinatorial Optimization, APPROX/RANDOM (2025), 34:1–34:12.

Incidence Theorems for Multivariate Polynomials via Cayley Color Graphs

Chong Shangguan

(joint work with Yulin Yang and Tao Zhang)

We prove several bounds on the number of incidences between two sets of multivariate polynomials of bounded degree over finite fields. From these results, we deduce bounds on incidences between points and multivariate polynomials, extending and strengthening a recent bound of Tamo for points and univariate polynomials. Our bounds are asymptotically tight for a wide range of parameters.

To prove these results, we establish a novel connection between the incidence problem and a naturally defined Cayley color graph, in which the weight of colored edges faithfully reflects the number of incidences. This motivates us to prove an expander mixing lemma for general abelian Cayley color graphs, which generalizes the classic mixing lemma of Alon and Chung, and controls the total weight of colored edges crossing two vertex subsets via eigenvalues.

The talk is based on a joint work [1] with Yulin Yang (Shandong University) and Tao Zhang (Xidian University).

References

[1] Chong Shangguan, Yulin Yang, and Tao Zhang, *Incidence theorems for multivariate polynomials over finite fields*, **arXiv:2509.17563** (2025).

High Rate Multivariate Polynomial Evaluation Codes

Swastik Kopparty
(joint work with Mrinal Kumar, John Kim)

The classical Reed-Muller codes over a finite field $\mathbb{F}_q$ are based on evaluations of m-variate polynomials of degree at most d over a product set U^m, for some $d < |U|$. Because of their good distance properties, as well as the ubiquity and expressive power of polynomials, these codes have played an influential role in coding theory and complexity theory. This is especially so in the setting of U being $\mathbb{F}_q$ where they possess deep locality properties. However, these Reed-Muller codes have a significant limitation in terms of the rate achievable — the rate cannot be more than $\frac{1}{m!} = \exp(-m \log m)$.

In this work, we give the first constructions of multivariate polynomial evaluation codes which overcome the rate limitation – concretely, we give explicit evaluation domains $S \subseteq \mathbb{F}_q^m$ on which evaluating m-variate polynomials of degree at most d gives a good code. For $m = O(1)$, these new codes have relative distance $\Omega(1)$ and rate $1 - \epsilon$ for any $\epsilon > 0$. In fact, we give two quite different constructions, and for both we develop efficient decoding algorithms for these codes that can decode from half the minimum distance.

The first of these codes is based on evaluating multivariate polynomials on simplex-like sets. The distance of this code is proved via a generalized Schwartz-Zippel lemma on the probability of non-zeroness when evaluating polynomials on sparser subsets of U^m – the final bound only depends on the "shape" of the set, and recovers the Schwartz-Zippel bound for the case of the full U^m, while still being $\Omega(1)$ for much sparser simplex-like subsets of U^m.

The second of these codes is more algebraic and, surprisingly (to us), has some strong locality properties. It is based on evaluating multivariate polynomials at the intersection points of hyperplanes in general position. It turns out that these evaluation points have many large subsets of collinear points. These subsets form the basis of a simple local characterization, and using some deeper algebraic tools generalizing ideas from Polischuk-Spielman [3], Raz-Safra [1] and Ben-Sasson-Sudan [2], we show that this gives a local test for these codes. Interestingly, the set of evaluation points for these locally testable multivariate polynomial evaluation codes can be as small as $O(d^m)$, and need not occupy a constant or even noticeable fraction of the full space $\mathbb{F}_q^m$.

References

[1] Ran Raz and Shmuel Safra, *A Sub-Constant Error-Probability Low-Degree Test, and a Sub-Constant Error-Probability PCP Characterization of NP*. Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing (1997) 475–484.

[2] Eli Ben-Sasson and Madhu Sudan, *Robust Locally Testable Codes and Products of Codes*. Random Structures & Algorithms **28(4)** (2006), 387–402.

[3] Alexander Polishchuk and Daniel A. Spielman, *Nearly-linear size holographic proofs*, Symposium on the Theory of Computing (1994).

A $2^{\Omega(k^{1/4})}$ Lower-Bound for Linear 3-LCC

Tal Yankovitz

A q-locally correctable code (LCC) $C : 0, 1^k \to 0, 1^n$ is a code in which it is possible to correct every bit of a (not too) corrupted codeword by making at most q queries to the word. The cases in which q is constant are of special interest, and so are the cases that C is linear. In a breakthrough result Kothari and Manohar [1] showed that for linear 3-LCC $n = 2^{\Omega(k^{1/8})}$. In this work [2] we prove that $n = 2^{\Omega(k^{1/4})}$. As Reed-Muller codes yield 3-LCC with $n = 2^{O(k^{1/2})}$, this brings us closer to closing the gap. Moreover, in the special case of design-LCC (into which Reed-Muller fall) the bound we get is $n = 2^{\Omega(k^{1/3})}$.

References

[1] P. K. Kothari and P. Manohar, *An Exponential Lower Bound for Linear 3-Query Locally Correctable Codes*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 776–787.

[2] T. Yankovitz, *A Stronger Bound for Linear 3-LCC*, 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS) (2024), 1786–1801.

Recent Advances and Applications of Algebraic Geometry Codes (Survey Talk)

Gretchen L. Matthews

In this talk, we begin with a review of algebraic geometry codes. Algebraic geometry codes may be defined using curves over finite fields. Given a projective curve $\mathcal{X}$ over a finite field $\mathbb{F}_q$, let G and $D = P_1 + \cdots + P_n$ be divisors on $\mathcal{X}$ such that each P_i is an $\mathbb{F}_q$-rational point on $\mathcal{X}$ and no P_i appears in the support of G. The associated algebraic geometry code is

$$C(D, G) = \{(f(P_1), f(P_2), \ldots, f(P_n)) : f \in \mathcal{L}(G)\} \subseteq \mathbb{F}_q^n$$

where

$$\mathcal{L}(G) := \{f \in \mathbb{F}(X) : (f) \geq -G\} \cup \{0\}.$$

Next, we consider several families of algebraic geometry codes. The most common example of an algebraic geometry code arises from taking $\mathcal{X}$ to be the projective line and $G = \alpha P_\infty$ to be a positive multiple of the point at infinity, in which case the code is a Reed-Solomon code. Other standard examples include codes from larger genus curves such as norm-trace curves given by $\mathcal{X}_{q,r} : Tr(y) = N(x)$ over $\mathbb{F}_{q^r}$; a quotient of norm-trace curve $\mathcal{X}_{q,r,u} : Tr(y) = x^u$ over $\mathbb{F}_{q^r}$ where $u \mid \frac{q^r-1}{q-1}$; and curves from Kummer extensions $\mathcal{X} : L(y) = x^u$ over $\mathbb{F}_{q^r}$ where $L(y) := \sum_{i=0}^{d} a_i y^{q^i}$ with $a_0, a_d \neq 0$ has d distinct roots in $\mathbb{F}_{q^r}$, $r \geq 2$, and $u \mid \frac{q^r-1}{q-1}$. In each of these cases, explicit bases for the Riemann-Roch spaces of divisor which are linear combination of certain points on a line and possibly the point at infinity [14, 16]. Hence, multipoint codes, meaning those in which the support of the divisor G consists of more than one point, may be defined explicitly.

We focus on two objects related to linear codes and applications. Recall that the hull of a code C over $\mathbb{F}_q$ is

$$\mathrm{hull}(C) := C \cap C^{\perp}.$$

A code is said to be linearly complementary dual (LCD) if and only if its hull is trivial. LCD codes, introduced by Massey [1], have been shown to simultaneously protect data from side channel and fault injection attacks [2]. Hulls appear naturally in a number of applications such as the classification of other mathematical objects [3] and cryptography [4]. The same is true for

$$Aut(C) := \{\sigma \in S_n : \sigma(c) \in C \forall c \in C\},$$

the permutation automorphism group of a linear code C of length n. Two linear codes C and C' are said to be monomially equivalent if and only if there exists a monomial matrix M such that $cM \in C'$ for all codewords $c \in C$.

It has been known since 2006 that non-special divisors of small degree exist [5] yet explicit constructions appeared elusive. The desire to define small degree non-special divisors is related to LCD codes. We give some constructions of non-special divisors of small degree which are necessarily multipoint divisors [6]. In other work, motivated by entanglement-assisted quantum codes [7, 8], we provide a procedure that takes as input a linear code C and an integer $l \leq \dim \mathrm{hull}(C)$ and then produces a monomially equivalent linear code C' such that $\dim \mathrm{hull}(C') = l$. As a consequence, we recover some results of [9].

We survey some results on automorphisms of curves and how they relate to code automorphisms. If $\mathcal{X}$ is a curve of genus $g \geq 2$ with divisors D and G having disjoint support, $\deg D \geq (1+g)\deg G$, and $\deg G \geq 2g+1$, then the set of automorphisms of $\mathcal{X}$ that fix the divisors D and G gives rise to some automorphisms of the associated algebraic geometry code, meaning

$$Aut_{D,G}(\mathcal{X}) = Aut\left(C(D,G)\right).$$

We share some recent results demonstrating how code automorphisms provide permutation decoding of burst errors for Hermitian one-point codes.

We end by noting that the algebraic geometry code construction is quite flexible and can be modified to yield locally recoverable codes that are designed to recover erasures using small sets of surviving coordinates; see [10] for instance. Codes constructed using fiber products of curves give rise to multiple recovery sets for each coordinate [11, 12].

Many additional applications of algebraic geometry codes may be found in the literature.

References

[1] J. L. Massey, *Linear codes with complementary duals*, Discrete Math. **106–107** (1992), 337–342.

[2] C. Carlet and S. Guilley, *Complementary dual codes for counter-measures to side-channel attacks, Coding Theory and Applications*, Springer (2015), 97–105.

[3] E. F. Assmus and J. D. Key, *Affine and projective planes*, Discret. Math. **83(2-3)** (1990),161–187.

[4] M. Bardet, A. Otmani, and M. Saeed-Taha, *Permutation code equivalence is not harder than graph isomorphism when hulls are trivial*, Proceedings of IEEE International Symposium on Information Theory (ISIT, 2019), 2464–2468.

[5] S. Ballet and D. Le Brigand, *On the existence of non-special divisors of degree g and $g-1$ in algebraic function fields over $\mathbb{F}_q$*, J. Number Theory **116** (2006), 293–310.

[6] E. Camps-Moreno, H. H. López, and G. L. Matthews, *Explicit Non-special Divisors of Small Degree, Algebraic Geometric Hulls, and LCD Codes from Kummer Extensions*, SIAM Journal on Applied Algebra and Geometry **8(2)**(2024), 394–413.

[7] G. Luo, M. F. Ezerman, N, Grassl, and S. Ling, *How Much Entanglement Does a Quantum Code Need?*, **arXiv:2207.05647** (2022).

[8] K. Guenda, S. Jitman, and T. A. Gulliver, *Constructions of good entanglement- assisted quantum error correcting codes*, Des. Codes and Cryptogr., **86(1)** (2018), 121–136.

[9] C. Carlet, S. Mesnager, C. Tang, Y. Qi and R. Pellikaan, *Linear Codes Over $\mathbb{F}_q$ Are Equivalent to LCD Codes for $q > 3$*, IEEE Trans. Inform. Theory, **64(4)** (2018), 3010–3017.

[10] A. Barg, I. Tamo, and S. Vlăduţ, *Locally recoverable codes on algebraic curves*, IEEE Trans. Inf. Theory **63(8)** (2017), 4928–4939.

[11] K. Haymaker, B. Malmskog, and G. Matthews, *Algebraic hierarchical locally recoverable codes with nested affine subspace recovery*, Des. Codes Cryptogr. **93** (2025), 111–132.

[12] K. Haymaker, B. Malmskog, and G. L. Matthews, *Locally recoverable codes with availability from fiber products of curves*, Adv. Math. Commun. **12(2)** (2018), 317–336.

[13] S. E. Anderson, E. Camps-Moreno, H. H. López, G. L. Matthews, D. Ruano and I. Soprunov, *Relative Hulls and Quantum Codes*, IEEE Transactions on Information Theory **70(5)** (2024), 3190-3201.

[14] H. Maharaj, G. L. Matthews, and G. Pirsic. *Riemann-Roch spaces of the Hermitian function field with applications to algebraic geometry codes and low-discrepancy sequences*, J. Pure Appl. Algebra **195(3)** (2005), 261–280.

[15] S. Mesnager, C. Tang, and Y. Qi, *Complementary dual algebraic geometry codes*, IEEE Trans. Inform. Theory, **64(4)** (2018), 2390–2397.

[16] J. Peachey, *Bases and applications of Riemann-Roch spaces of function fields with many rational places*, Doctoral dissertation, Clemson University (2011).

[17] N. Sendrier, *Finding the permutation between equivalent linear codes: the support splitting algorithm*, IEEE Trans. Inform. Theory **46(4)** (2000), 1193–1203.

Trace Codes Arising from Algebraic Curves

Gil Cohen

(joint work with Dean Doron, Noam Goldgraber, and Tomer Manket)

In this talk we present our work which put forth a candidate for a family of explicit constructions over constant-size fields that meet—or perhaps even surpass—the Gilbert–Varshamov (GV) bound, with the primary goal of achieving this for binary codes. Our main technical contribution is to take the first steps toward their analysis and to understand their limitations. The results we prove are general, extend beyond our original motivation, and may find further applications.

Our idea is simple: we aim to leverage the underlying algebraic structure of Algebraic-Geometric codes—which enables them to beat the Gilbert-Varshamov bound—in the alphabet-reduction step as well. The hope is that, by exploiting this structure, the reduction will not substantially compromise the excellent distance–rate tradeoff of the original AG code. In its most basic form, the alphabet-reduction method we propose applies the field trace to each coordinate of every codeword. While this is the variant we focus on in this work, we view it as a special case within a broader family of constructions whose common feature is the aforementioned strategy of leveraging structure for alphabet reduction.

Specifically, we prove a Hasse–Weil–type estimate that bounds the number of rational points on a curve. Unlike the classical situation, our result applies in greater generality: we work with extensions of general curves rather than only the projective line. The standard bound in this setting, Grothendieck's trace formula, is insufficient for bounding the minimum distance of trace AG codes. Our bound strengthens this bound and provides satisfactory lower bounds on the distance.

Using Gröbner Basis to Search for Dual-Containing Codes over Rings

Hedongliang Liu

(joint work with Cornelia Ott and Felix Ulmer)

Gröbner basis is a powerful tool to solve systems of polynomial equations. It can be seen as a generalization of Gaussian elimination for linear systems. The output of Gröbner basis gives a simpler description of the solution space of the system, from which we can easily extract all solutions. Dual-containing codes are important in the construction of quantum error-correcting codes using the Cadelbank-Shor-Stean (CSS) [1, 2] construction. This work focuses on finding dual-containing skew polycyclic codes using Gröbner basis methods.

In this work [3], we looked for dual-containing polycyclic codes over rings via skew polynomials [4]. Our method involves setting up a system of polynomial equations on the coeffiients of a generator polynomial and a parity check polynomial based on the definition of (θ, δ) polycyclic codes and dual-containing condition, then solving this system using Gröbner basis in Magma.

As a result, we have computed all dual-containing (θ, δ)-codes of length up to 13 over the base ring $\mathbb{F}_2[v]$ and up to 10 over $\mathbb{F}_2[u]$. Some of the codes found have unique weight enumerators that cannot be found using ordinary polynomials.

For future research, it would be insightful to compare the (θ, δ)-codes with other existing codes over rings or bounds on the cardinality or the distance of codes over rings, such as Singleton-like bounds, sphere-packing bounds and Gilbert-Varshamov (GV) bounds.

References

[1] A. R. Calderbank and P. W. Shor, *Good quantum error-correcting codes exist*, Physical Review A **54** (1996), 1098–1105.

[2] A. M. Steane, *Simple quantum error-correcting codes*, Physical Review A **54** (1996), 4741–4751.

[3] H. Liu, C. Ott, and F. Ulmer. *A Gröbner approach to dual-containing cyclic left module (θ, δ)-codes over finite commutative Frobenius rings*, Advances in Mathematics of Communications **19(6)** (2025), 1723–1741.

[4] Ø. Ore, *Theory of non-commutative polynomials*, Annals of Mathematics **34(3)** (1933) 480–508.

Optimal Additive and Linear *b*-Symbol Codes for Large Distances

Sascha Kurz

For a finite *alphabet* $\mathcal{A}$ a *code* C of *length* n and *minimum distance* d is a subset of $\mathcal{A}^n$ such that any two elements differ in at least d positions. E.g. $C = \{cabbdb, bcabbd, abcdbb, cdadcc, acdcdc, dacccd, dcbdaa, bdcada, cbdaad\}$ is a code with length 6 and minimum distance $d = 5$ over the alphabet $\mathcal{A} = \{a, b, c, d\}$. Given parameters n, d, and $\#\mathcal{A}$, the aim is to maximize the code size $\#C$. In our example size 9 is indeed maximal [1]. Alternatively, one can minimize n given d, $\#\mathcal{A}$, and $\#C$. For alphabet $\mathcal{A} = \mathbb{F}_q$ we say that C is *linear* if it is linearly closed. The parameters of a linear code are related by the so-called *Griesmer bound* [2, 3]

$$n \ge \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil =: g_q(k, d), \tag{1}$$

where $k = \log_q \#C$. Interestingly enough, this bound can always be attained with equality if the minimum distance d is sufficiently large [3]. If C is only additively closed we call it *additive*. Each additive code is linear over some subfield so that we set $\mathcal{A} = \mathbb{F}_{q^h}$ and assume that C is linear over $\mathbb{F}_q$, so that $\#C = q^k$.

Let $G \in \mathbb{F}_q^{k\times n}$ be a *generator matrix* of a linear code C, i.e. a matrix whose rows form a basis of C. The columns of G span 1-dimensional subspaces which form a multiset of n points in the projective geometry $\operatorname{PG}(k-1, q)$ such that each hyperplanes contains at most $n - d$ points [4]. Similarly, an additive code over $\mathcal{A} = \mathbb{F}_{q^h}$ is given as the $\mathbb{F}_q$-row span of a full-rank matrix $G \in \mathbb{F}_{q^h}^{k\times n}$. Choosing an $\mathbb{F}_q$-basis $\mathcal{B}$ of $\mathbb{F}_{q^h}$ we can rewrite G as $\widetilde{G} \in \mathbb{F}_q^{kh\times n}$. Writing $\mathbb{F}_4 \simeq \mathbb{F}_2[\omega]/\big(\omega^2 + \omega + 1\big)$, we can start with the generator matrix of a linear code, interprete it as the generator matrix of an additive code and use the basis $\mathcal{B}$ to obtain the example

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & \omega & \omega^2 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & \omega & \omega & \omega & \omega \\ 1 & 0 & 1 & \omega & \omega^2 \\ \omega & 0 & \omega & \omega^2 & 1 \end{pmatrix} \to \begin{pmatrix} 00 & 10 & 10 & 10 & 10 \\ 00 & 01 & 01 & 01 & 01 \\ 10 & 00 & 10 & 01 & 11 \\ 01 & 00 & 01 & 11 & 10 \end{pmatrix}.$$

The blocks of h subsequent columns span subspaces of dimension at most h, which are elements in $\operatorname{PG}(k-1, q)$ of geometric dimension at most $h-1$. Indeed, additive codes over $\mathcal{A} = \mathbb{F}_{q^h}$ with length n, minimum distance d, and size q^k are in one-to-one correspondence to multisets of n subspaces of geometric dimension at most $h-1$ in $\operatorname{PG}(k-1, q)$ such that each hyperplane contains at most $n-d$ of those subspaces [5]. Replacing those subspaces by their contained points we obtain a multiset of points that corresponds to a linear code over $\mathbb{F}_q$ with length $n \cdot (q^h - 1)/(q - 1)$ and minimum distance $q^{h-1} \cdot d$ (assuming some non-degeneracy) [6]. Applying the Griesmer bound (1) to the obtained linear code yields, see [5, Theorem 12],

$$n \ge \left\lceil \frac{g_q\big(r, d \cdot q^{h-1}\big) \cdot (q - 1)}{q^h - 1} \right\rceil = \left\lceil \frac{(q-1) \cdot \sum\limits_{i=0}^{r-1} \left\lceil d \cdot q^{h-1-i} \right\rceil}{q^h - 1} \right\rceil. \tag{2}$$

One main conclusion of our work is that this bound can always be attained with equality if the minimum distance d is sufficiently large. Reverting the above chain of changes between coding theory and geometry boils down to the problem of partitioning a multiset of points into subspaces of geometric dimension $h-1$, which is rather hard in general, so that we consider a special subclass.

Each minimum distance $d \in \mathbb{N}$ can be uniquely written as $d = \sigma q^{k-1} - \sum_{i=1}^{k-1} \varepsilon_i \cdot q^{i-1}$, where $\sigma \in \mathbb{N}$ and $\varepsilon_i \in \{0, 1, \ldots, q-1\}$ for all $1 \le i \le k-1$. With this, the Griesmer bound from (1) is attained with equality, i.e. $n = g_q(k, d)$, iff

$$n = \sigma \cdot \frac{q^k - 1}{q-1} - \sum_{i=1}^{k-1} \varepsilon_i \cdot \frac{q^i - 1}{q-1}. \tag{3}$$

In order to describe a special variant of the Solomon–Stiffler construction, see [3], we assume a chain $S_1 \le S_2 \le \cdots \le S_k$ of subspaces S_i with algebraic dimension i in $\operatorname{PG}(k-1, q)$. For each subspace T we denote its *characteristic function* by χ_T, i.e. $\chi_T(P) = 1$ if point P is contained in T and $\chi_T(P) = 0$ otherwise. With this, $\mathcal{M} = \sigma\chi_{S_k} - \sum_{i=1}^{k-1} \varepsilon_i \chi_{S_i}$ is a multiset of points whose corresponding linear code attains the Griesmer bound (1) if $\sigma \ge \sum_{i=1}^{k-1} \varepsilon_i$ and $0 \le \varepsilon_i \le q-1$, i.e. if d is sufficiently large. More generally, we say that a multiset of points $\mathcal{M}$ in $\operatorname{PG}(k-1, q)$ has type $\sigma[k] - \sum_{i=1}^{k-1} \varepsilon_i[i]$, where $\sigma \in \mathbb{N}$ and $\varepsilon_i \in \mathbb{Z}$ for all $1 \le i \le k-1$, if $\mathcal{M} = \sigma\chi_{S_k} - \sum_{i=1}^{k-1} \varepsilon_i \chi_{S_i}$. Note that σ needs to be sufficiently large. We say that a multiset of points is *h-partitionable* if it can be written as the sum of characteristic functions of subspaces of algebraic dimension h.

Theorem. ([6]) Let q be a prime power, $k > h \ge 1$, $g := \gcd(k, h)$, and $\varepsilon_1, \ldots, \varepsilon_{k-1} \in \mathbb{Z}$ such that q^{h-i} divides ε_i for all $1 \le i < h$ and $\sum_{i=1}^{k-1} \varepsilon_i \cdot \frac{q^i-1}{q-1} \equiv 0 \pmod{\frac{q^g-1}{q-1}}$. Then, there exists a $\sigma \in \mathbb{N}$ such that there exists an h-partionable multiset of points in $\operatorname{PG}(k-1, q)$ with type $\left(\sigma + t \cdot \frac{q^h-1}{q^g-1}\right)[k] - \sum\limits_{i=1}^{k-1} \varepsilon_i[i]$ for all $t \in \mathbb{N}$.

We remark that the stated conditions are also necessary for the assumed chain $S_1 \le \cdots \le S_k$ and that the corresponding proof is constructive. As an example we mention the existence of a 2-partionable multiset of points in $\operatorname{PG}(7, 2)$ with type $t[8] - [7] - [5] - [3]$ for each $t \ge 3$. Those $85t - 55$ lines in $\operatorname{PG}(7, 2)$ have the property that each hyperplane contains at most $21t - 13$ lines, i.e. the corresponding additive code C_t over $\mathbb{F}_4$ has length $n_t = 85t - 55$, minimum distance $d_t = 64t - 42$, and cardinality $\#C_t = 2^8$. We remark that $\mathbb{F}_4$-linear codes with the same minimum distance and cardinality require lengths at least $85t - 53$, i.e. additive codes outperform linear codes for the stated parameters of d, $\#\mathcal{A}$, and $\#C$.

Corollary. For given q, k, h, and sufficiently large d the Griesmer bound for additive codes (2) can always be attained. Moreover, additive codes outperform linear codes if h does not divide k or if the the difference of (2) and (1) is positive.

Open problem. Find constructions for additive codes outperforming linear codes for relatively small minimum distances, see e.g. [6, 7, 9, 10], or decrease the necessary σ in the Solomon–Stiffler type construction. Find improved upper bounds.

In storage applications the reading device is sometimes insufficient to isolate adjacent symbols, which makes it necessary to adjust the standard coding-theoretic error model. Cassuto and Blaum studied a model where pairs of adjacent symbols are read in every step and introduced the so-called symbol-pair metric for codes [8]. This notion was generalized to the b-symbol metric where b-tuples of adjacent symbols are read at every step [11]. For linear codes each b subsequent columns of a generator matrix span a subspace with dimension at most b, so that the Griesmer bound for additive codes (2) applies [12]. Again, the Griesmer bound can always be attained if the minimum distance is sufficiently large [13].

Open problem Find Griesmer type bound bounds and attaining Solomon–Stiffler type constructions for other metric spaces.

References

[1] G. T. Bogdanova, A. E. Brouwer, S. N. Kapralov, and P. R. J. Östergård. *Error-correcting codes over an alphabet of four elements*, Des. Codes Cryptogr. **23** (2001), 333–342.

[2] J. H. Griesmer. *A bound for error-correcting codes*, IBM J. Res. Dev. **4(5)** (1960), 532–542.

[3] G. Solomon and J. J. Stiffler. *Algebraically punctured cyclic codes*, Inf. Control. **8(2)** (1965), 170–179.

[4] S. M. Dodunekov and J. Simonis. *Codes and projective multisets*, Electron. J. Combin. **6** (1998), 1–23.

[5] S. Ball, M. Lavrauw, and T. Popatia. *Griesmer type bounds for additive codes over finite fields, integral and fractional MDS codes*, Des. Codes Cryptogr. **93(1)** (2025), 175–196.

[6] S. Kurz. *Additive codes attaining the Griesmer bound*, **arXiv: 2412.14615** (2024), 100 pages.

[7] S. Ball and T. Popatia. *Additive codes from linear codes*, **arXiv:2506.03805** (2025), 9 pages.

[8] Y. Cassuto and M. Blaum. *Codes for symbol-pair read channels*, IEEE Trans. Inf. Theory **57(12)** (2011), 8011–8020.

[9] C. Guan, R. Li, Y. Liu, and Z. Ma. *Some quaternary additive codes outperform linear counterparts*, IEEE Trans. Inf. Theory **69(11)** (2023), 7122–7131.

[10] C. Guan, J. Lv, G. Luo, and Z. Ma. *Combinatorial constructions of optimal quaternary additive codes*, IEEE Trans. Inf. Theory **70(11)** (2024), 7690–7700.

[11] E. Yaakobi, J. Bruck, and P. H. Siegel. *Constructions and decoding of cyclic codes over b-symbol read channels*, IEEE Trans. Inf. Theory **62(4)** (2016), 1541–1551.

[12] G. Luo, M. F. Ezerman, C. Güneri, S. Ling, and F. Özbudak. *Griesmer bound and constructions of linear codes in b-symbol metric*, IEEE Trans. Inf. Theory **70(11)** (2024), 7840–7847.

[13] S. Kurz, *Linear codes for b-symbol read channels attaining the Griesmer bound*, **arXiv:2507.07728** (2025), 27 pages.

Local Properties and Code Reductions (Survey Talk)

Jonathan Mosheiff

Several well-studied notions in coding theory—most notably *list-decodability* and *list-recoverability*—are said to be *local*. A property of codes is called *local* if its (non)satisfaction can be determined by examining only a bounded number of codewords. Equivalently, if a code violates the property, the violation can already be witnessed by a small subset of its codewords.

For example, being (ρ, L)-list-decodable is an $(L+1)$-local property: a code $C \subseteq \mathbb{F}_q^n$ fails to be (ρ, L)-list-decodable precisely when there exist $L+1$ codewords contained in a single radius-ρ ball. Thus, to convince Bob that a code C is not (ρ, L)-list-decodable, Alice need only present such a collection of $L+1$ codewords as a witness.

The framework of local properties was introduced in [1] and developed further in subsequent works [2, 3, 4, 5, 6]. Collectively, these papers show that local properties exhibit sharp threshold phenomena in natural random code ensembles, and that the thresholds coincide across several such ensembles. The main findings can be summarized as follows:

(1) **Threshold behavior in elementary ensembles.** Elementary ensembles such as random linear codes (RLCs) and plain random codes (PRCs) exhibit a sharp threshold for every local property $\mathcal{P}$ of q-ary codes. Specifically, there exists a threshold rate $R^*_{\mathcal{P}}$ such that a random q-ary RLC of rate $R^*_{\mathcal{P}} - \varepsilon$ satisfies $\mathcal{P}$ with high probability, while one of rate $R^*_{\mathcal{P}} + \varepsilon$ fails it with high probability.

(2) **Characterization of the threshold rate.** For a local property $\mathcal{P}$, the threshold rate $R^*_{\mathcal{P}}$ admits an elementary probabilistic characterization in terms of expectations over random small subsets of codewords.

(3) **Equivalence between ensembles.** Using this characterization, one can show that different code ensembles exhibit the same threshold behavior. In particular, [6] proves that *random Reed–Solomon* codes have the same thresholds for all local properties as random linear codes. Consequently, random Reed–Solomon codes share the same list-decodability and list-recoverability thresholds as RLCs of comparable rate.

As discussed above, the framework of local properties implies that any results concerning the list-decodability or list-recoverability of random linear codes (RLCs) can be transferred to other code ensembles, such as random Reed–Solomon codes. This phenomenon—known as a *reduction between code ensembles*—further reinforces the importance of understanding the list-decodability and list-recoverability behavior of RLCs, which is already a central topic in the theory of random codes.

In this talk, we survey the current state of knowledge on the list-decodability and list-recoverability of random linear codes. We then review the aforementioned sequence of works on *local properties*, highlighting their key techniques, threshold results, and implications for structured ensembles.

We conclude with several open problems:

(1) **Beyond coordinate-wise constraints.** Existing results about local properties require a violating tuple of codewords to violate a specified set of *coordinate-wise* constraints. Consequently, the current framework does not capture constraints involving interactions *across* coordinates. Extending it in this direction would make it possible to address properties such as list-decodability and list-recoverability in insertion–deletion channels.

(2) **Beyond the first level of the logical hierarchy.** Local properties, as discussed here, can be viewed as corresponding to coNP or to Π_1 logical sentences. Extending the framework to higher levels of the arithmetic hierarchy, such as Π_2 and Σ_2, may make it possible to reason about properties like covering, and to reduce the covering behavior of one code ensemble to that of another.

(3) **List-decodability and list-recoverability above capacity.** Consider list-decodability and list-recoverability above capacity. For instance, in a q-ary code of rate $R = 1 - h_q(\rho) + \epsilon$, a ball of radius ρ around a uniformly random center contains $q^{\epsilon n}$ codewords in expectation. Is it true that, for a random linear code, with high probability every such ball contains at most $q^{\epsilon n} \cdot (1 + o(1))$ codewords? This question lies beyond the reach of the current local-property framework, since the size of a violating list is exponential in n. It would be interesting to extend the present framework to handle this setting.

References

[1] J. Mosheiff, N. Resch, N. Ron-Zewi, S. Silas, and M. Wootters, *LDPC codes achieve list decoding capacity*, Proc. 61st IEEE Annual Symposium on Foundations of Computer Science (FOCS) (2020), 458–469.

[2] V. Guruswami, R. Li, J. Mosheiff, N. Resch, S. Silas, and M. Wootters, *Bounds for list-decoding and list-recovery of random linear codes*, IEEE Trans. Inf. Theory **68** (2022), 923–939.

[3] V. Guruswami, J. Mosheiff, N. Resch, S. Silas, and M. Wootters, *Threshold rates for properties of random codes*, IEEE Trans. Inf. Theory **68** (2022), 905–922.

[4] V. Guruswami and J. Mosheiff, *Punctured low-bias codes behave like random linear codes*, Proc. 63rd IEEE Annual Symposium on Foundations of Computer Science (FOCS) (2022), 36–45.

[5] J. Mosheiff, N. Resch, K. Shang, and C. Yuan, *Randomness-efficient constructions of capacity-achieving list-decodable codes*, **arXiv:2402.11533** (2024).

[6] M. Levi, J. Mosheiff, and N. Shagrithaya, *Random Reed–Solomon codes and random linear codes are locally equivalent*, **arXiv:2406.02238** (2024).

List-Decoding and List-Recovery of Random Codes Ensembles

Ray Li

(joint work with Omar Alrabiah, Venkatesan Guruswami, and Nikhil Shagrithaya)

List-decodable and list-recoverable codes are important in theory and practice. In practice, their relaxed decoding guarantee allows them to tolerate more error, giving them applications such as group testing and compressed sensing. In theory, their fundamental mathematical definitions give them "extraneous" applications that have no obvious need for error correction, such as in pseudorandomness, computational complexity, and cryptography.

A fundamental question is: what kinds of codes are optimally list-decodable and list-recoverable? Uniformly random codes, where each codeword is sampled independently at random, achieve the best-known parameters in both list-decoding and list-recovery. I will discuss recent progress [1, 2] on whether much more structured random ensembles of codes — such as random linear codes and randomly punctured Reed-Solomon codes — enjoy list-decoding and list-recovery guarantees comparable to those of uniformly random codes.

References

[1] Omar Alrabiah, Venkatesan Guruswami, Ray Li, *Randomly punctured Reed–Solomon codes achieve list-decoding capacity over linear-sized fields*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC (2024), 1458–1469.

[2] Ray Li, Nikhil Shagrithaya, *Near-Optimal List-Recovery of Linear Code Families*, Approximation, Randomization, and Combinatorial Optimization, APPROX/RANDOM **353** (2025), 53:1–53:14.

List-Recovery of Random Linear Codes over Small Fields

Dean Doron

(joint work with Jonathan Mosheiff, Nicolas Resch and João Ribeiro)

We study list-recoverability of random linear codes over small fields, both from errors and from erasures. We consider codes of rate ε-close to capacity, and aim to bound the dependence of the output list size L on ε, the input list size ℓ, and the alphabet size q. Prior to our work, the best upper bound was $L = q^{O(\ell/\varepsilon)}$ by Zyablov and Pinsker [1].

Previous work has identified cases in which *linear* codes provably perform worse than non-linear codes with respect to list-recovery. While there exist non-linear codes that achieve $L = O(\ell/\varepsilon)$, we know that $L \geq \ell^{\Omega(1/\varepsilon)}$ is necessary for list recovery from erasures over fields of small characteristic, and for list recovery from errors over large alphabets.

We show that in other relevant regimes there is no significant price to pay for linearity, in the sense that we get the correct dependence on the gap-to-capacity ε and go beyond the Zyablov–Pinsker bound for the first time. Specifically, when q is constant and ε approaches zero,

- For list-recovery from erasures over *prime fields*, we show that $L \leq C_1/\varepsilon$. By prior work, such a result cannot be obtained for low-characteristic fields.
- For list-recovery from errors over *arbitrary fields*, we prove that $L \leq C_2/\varepsilon$.

Above, C_1 and C_2 depend on the decoding radius, input list size, and field size. We provide concrete bounds on the constants above, and the upper bounds on L improve upon the Zyablov–Pinsker bound whenever $q \leq 2^{(1/\varepsilon)^c}$ for some small universal constant $c > 0$.

References

[1] V. V. Zyablov, and M. S. Pinsker, *List concatenated decoding*, Problemy Peredachi Informatsii **17.4** (1981), 29–33.

List-Decodable/-Recoverable Codes in the Zero-Rate Regime

Nicolas Resch
(joint work with Chen Yuan and Yihan Zhang)

A classical result of Plotkin [1] establishes that over binary alphabets, positive rate codes cannot have minimum distance greater than 1/2, and thus can uniquely correct at most a 1/4 fraction of errors. It is additionally known that if one insists on constructing a code correcting a $1/4 + \varepsilon$ fraction of errors (for small $\varepsilon > 0$), then this code can have size at most $O(1/\varepsilon)$, and that this is tight.

If one moves to list-decoding binary codes with list-size L – that is, the decoder may output up to L guesses for the transmitted message, as long as one of the guesses is correct – Blinovsky [2] computed a similar threshold: the answer is the (somewhat inscrutable) value of

$$p_L = \frac{1}{2} - \frac{\binom{2k}{k}}{2^{2k+1}} \tag{1}$$

where L is $2k-1$ or $2k$. Blinovsky's argument additionally shows that if one asks for a $(p_L + \varepsilon, L)$-list-decodable code $\mathcal{C} \subseteq \{0,1\}^n$, then it will have size $O_{\varepsilon,L}(1)$: independent of n, but with a (massive) dependence on ε. Later, Alon, Bukh and Polyanskiy [3] showed that for *odd* L, such codes have size $O_L(1/\varepsilon)$ (as with Plotkin's bound), but already with $L = 2$ such codes of size $\Omega(1/\varepsilon^{3/2})$ exist.

In this talk, we will generalize all of these results to the list-decoding/-recovery setting for general (but constant) alphabet sizes. For list-decoding over alphabets of size q, we prove that the threshold is

$$p_{q,L} := \mathop{\mathbb{E}}_{X_1,\dots,X_{L+1}\sim\mathrm{Unif}([q])} [\mathsf{pl}(X_1,\dots,X_{L+1})] \ ,$$

where the notation $X_1,\dots,X_{L+1} \sim \mathrm{Unif}([q])$ means that we take $L+1$ independent and uniform samples from the alphabet $[q]$, and the function $\mathsf{pl}(x_1,\dots,x_{L+1})$

returns the number of $i \in [L+1]$ such that x_i equals a most popular value. That is,

$$\begin{aligned} \mathsf{pl} : [q]^{L+1} &\to \mathbb{R} , \\ (x_1, \ldots, x_{L+1}) &\mapsto \max\{|S| : S \subseteq [L+1] \text{ and } \forall i, j \in S,\ x_i = x_j\} . \end{aligned}$$

A routine computation shows this definition indeed recovers (1) when $q = 2$. A crucial tool in the proof is the concept of Schur convexity, which in certain cases allows one to show that the optimizing value for a function on a space of distributions is the uniform distribution.

Next, we generalize the argument of Alon, Bukh and Polyanskiy [3] to this setting of non-binary alphabets, showing that codes $\mathcal{C} \subseteq [q]^n$ that are $(p_{q,L} + \varepsilon, L)$-list-decodable necessarily have size $O_{q,L}(1/\varepsilon)$. Notably, for non-binary alphabets we find that the parity of the list-size L no longer plays a role.

Lastly, following a classical argument independently discovered by Elias and Bassalygo [4], we show how such bounds on codes in the "zero-rate regime" allow one to provide bounds on codes in the (standard) positive rate regime.

References

[1] M. Plotkin, *Binary codes with specified minimum distance*, IRE Transactions on Information Theory **6** (1960), 445–450.

[2] V. Blinovsky, *Bounds for codes in the case of list decoding of finite volume*, Problems of Information Transmission **22** (1986), 7–19.

[3] N. Alon, B. Bukh, and Y. Polyanskiy. *List-decodable zero-rate codes*, IEEE Transactions on Information Theory **65** (2018), 1657-1667.

[4] L. A. Bassalygo. *New upper bounds for error-correcting codes*, Problemy Peredachi Informatsii **1** (1965), 41–44.

[5] N. Resch, C. Yuan, and Y. Zhang. *Zero-rate thresholds and new capacity bounds for list-decoding and list-recovery* IEEE Transactions on Information Theory **70** (2024), 6211–6238.

[6] N. Resch, C. Yuan, and Y. Zhang. *Tight Bounds on List-Decodable and List-Recoverable Zero-Rate Codes* Proceedings of the 16th Innovations in Theoretical Computer Science Conference **325** (2025), 82:1–82:21.

Survey on Reed–Solomon Codes Against Insertions and Deletions (Survey Talk)

Roni Con

The performance of Reed–Solomon (RS) codes against synchronization errors, primarily insertions and deletions (insdel), has been studied extensively in recent years. In this talk, I will survey several results demonstrating that RS codes can indeed correct insdel errors.

First, I will present a sufficient condition on the evaluation points of an RS code that guarantees maximal error-correction capability (attaining the so called "half-Singleton" bound). Second, I will show that, over sufficiently large finite fields, there exist evaluation points for which the corresponding RS code meets this bound. Third, I will describe the current state-of-the-art explicit construction, which unfortunately requires extremely large field sizes.

I will also show a novel connection between RS codes that can correct against an adversary who first applies a *permutation* to the codeword and then deletes symbols, and a new cryptographic notion called *anonymous secret sharing.* In particular, I will show how such codes naturally give rise to a variant of Shamir's secret-sharing scheme in which the secret can be reconstructed anonymously (i.e., without revealing the identities of the participants). Moreover, for any unauthorized set of participants (set of participants that cannot learn anything about the secret), their identities remain hidden even if an adversary observes their shares.

Finally, I will discuss several open questions: (i) Can we reduce the field size required by the explicit constructions, or alternatively prove lower bounds on the field size needed for RS codes to attain maximal correction capability? (ii) Is there an efficient decoding algorithm for correcting deletions in RS codes? (iii) Are there additional cryptographic or coding-theoretic applications of such codes?

This survey is based on the works [1, 2, 3, 4, 5].

References

[1] Roni Con, Amir Shpilka, and Itzhak Tamo, *Reed solomon codes against adversarial insertions and deletions*, IEEE Transactions on Information Theory **69(5)** (2023), 2991–3000.

[2] Roni Con, Amir Shpilka, and Itzhak Tamo, *Optimal two-dimensional reed–solomon codes correcting insertions and deletions*, IEEE Transactions on Information Theory **70(7)** (2024), 5012–5016.

[3] Roni Con, Zeyu Guo, Ray Li, and Zihan Zhang, *Random reed-solomon codes achieve the half-singleton bound for insertions and deletions over linear-sized alphabets*, Proceedings of 52nd International Colloquium on Automata, Languages, and Programming, ICALP (2025), 60–1.

[4] Peter Beelen, Roni Con, Anina Gruica, Maria Montanucci, and Eitan Yaakobi, *Reed-solomon codes against insertions and deletions: Full-length and rate-*1/2 *codes*, **arXiv:2501.11371** (2025).

[5] Roni Con, *Anonymous shamir's secret sharing via reed-solomon codes against permutations, insertions, and deletions*, **arXiv:2412.17003** (2024).

Decoding Insertions/Deletions via List Recovery

Anisha Banerjee

(joint work with Roni Con, Antonia Wachter-Zeh, and Eitan Yaakobi)

We examine the challenge of efficiently decoding of codes from insertions and deletions. Most existing constructions with efficient decoders utilize synchronization strings, which transform the problem of decoding insertions and deletions into that of decoding substitutions and erasures. Our approach simplifies the problem of decoding of insertions and deletions to a list recovery problem. More specifically, any $(\rho, 2\rho n + 1, L)$-list-recoverable code is a (ρ, L)-list decodable insdel code. For instance, we apply this method to Reed-Solomon (RS) codes, for which efficient list-recovery algorithms up to the Johnson bound exist. In the adversarial insdel model, this decoding approach enables efficient (list) decoding from t insdel errors, provided that $t \cdot k = O(n)$. This leads to the first efficient insdel decoder applicable to $[n, k]$ RS codes when $k > 2$. We also investigate random insdel models like the

Davey-MacKay channel and find that for certain values of ρ, a $(\rho, n^{1/2+0.001}, L)$-list-recoverable code of length n can be efficiently list decode the channel output such that with high probability, the transmitted codeword is in the output list. In the realm of RS codes, this results in an improved rate-error tradeoff for these channels compared to the adversarial error setting. Finally, we adapt the Koetter-Vardy algorithm, a well-known soft-decision list decoding method for RS codes, to address the insertions and deletions caused by the Davey-MacKay channel.

References

[1] B. Haeupler and A. Shahrasbi, *Synchronization Strings: Codes for Insertions and Deletions Approaching the Singleton Bound*, Journal of the ACM **68(5)** (2021), 36:1–36:39.

[2] J. Brakensiek, V. Guruswami, and S. Zbarsky, *Efficient Low-Redundancy Codes for Correcting Multiple Deletions*, IEEE Transactions on Information Theory **64(5)** (2018), 3403–3410.

[3] R. Con, A. Shpilka, and I. Tamo, *Reed Solomon Codes Against Adversarial Insertions and Deletions*, IEEE Transactions on Information Theory **69(5)** (2023), 2991–3000.

[4] P. Beelen, R. Con, A. Gruica, M. Montanucci, and E. Yaakobi, *Reed-Solomon Codes Against Insertions and Deletions: Full-Length and Rate-1/2 Codes*, **arXiv:2501.11371** (2025).

[5] R. Con, Z. Guo, R. Li, and Z. Zhang, *Random Reed-Solomon Codes Achieve the Half-Singleton Bound for Insertions and Deletions over Linear-Sized Alphabets*, **arXiv:2407.07299** (2024).

[6] M. C. Davey and D. J. C. MacKay, *Reliable Communication over Channels with Insertions, Deletions, and Substitutions*, IEEE Transactions on Information Theory **47(2)** (2001), 687–698.

[7] R. Koetter and A. Vardy, *Algebraic soft-decision decoding of Reed-Solomon codes*, IEEE Transactions on Information Theory **49(11)** (2003), 2809–2825.

[8] V. Guruswami and M. Sudan, *Improved decoding of Reed-Solomon and algebraic-geometry codes*, IEEE Transactions on Information Theory **45(6)** (1999), 1757–1767.

Hardness Amplification via Coding Theory: New Results and New Open Problems

Ronen Shaltiel

(joint work with Marshall Ball and Jad Silbak)

Hardness amplification is the complexity theoretic task of transforming a given function f that is hard on the worst-case for small circuits, into a function g that is hard on average (on a uniformly chosen input) agsiant (slightly smaller) circuits. By the seminal work of Sudan, Trevisan and Vadhan [1], hardness amplification can be achieved using coding theoretic methods, and specifically, using error-correcting codes that are locally list-decodable.

Recently, motivated by both complexity theoretic and coding theoretic applications, a new twist on the hardness amplification problem was suggested by Shaltiel and Silbak [2]. In this twist, one requires that g is not only hard on average on a uniformly chosen input X, but also on every samplable distribution X with sufficient min-entropy, where the min-entropy threshold k is a parameter. As one wants a stronger conclusion, a stronger assumption is needed on the hard function f, and it is now required that f is hard on the worst-case even against nondeterministic circuits.

It turns out that this problem is also closely related to several coding theoretic problems (with the common theme being that now the code needs to be set up against a huge number of erasures, as well as few errors). In a sequence of works, Shaltiel and Silbak [2], Ball, Shaltiel and Silbak [3] and Shaltiel [4] make progress on this new hardness amplification problem. This progress gives improved codes for computationally bounded channels, as well as improved extractors for samplable distributions.

Some of this progress is achieved by noting that for these applications one does not need to solve the general case of the coding theoretic problem, but rather a (somewhat unnatural) relaxation of the coding theoretic problem which suffices for the application.

References

[1] Madhu Sudan, Luca Trevisan, and Salil Vadhan, *Pseudorandom generators without the XOR lemma*, Proceedings of the 31st annual ACM symposium on Theory of computing (1999), 537–546.

[2] Ronen Shaltiel and Jad Silbak, *Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 2028–2038.

[3] Marshall Ball, Ronen Shaltiel, and Jad Silbak, *Extractors for Samplable Distributions with Low Min-Entropy*, Proceedings of the 57th Annual ACM Symposium on Theory of Computing (2025), 596–603.

[4] Ronen Shaltiel, *Extractors for Samplable Distribution with Polynomially Small Min-Entropy* (2025).

Seeded Linear Bit-Fixing Extractors and Applications in Erasure Coding

Mahdi Cheraghchi

(joint work with Yeyuan Chen and Nikhil Shagrithaya)

A function $\mathrm{Ext}\colon [N] \times [D] \to [M]$, where $[N] := \{1, \dots, N\}$, is a (k, ε)-extractor if for every distribution X on $[N]$ with min-entropy

$$H_\infty(X) := \min_{x \in [N]} \{-\log_2 \Pr[X = x]\} \geq k,$$

the distribution $(Z, \mathrm{Ext}(X, Z))$ is ε-close in statistical distance to uniform on $[D] \times [M]$ when Z is chosen uniformly from $[D]$. The probabilistic method shows the existence of extractors with seed space $D = O(\frac{\log(N/K)}{\varepsilon^2})$ and output space $M = \Omega(\varepsilon^2 K)$. Achieving explicit constructions with parameters of this quality has been a central open problem for decades.

In this work, we investigate the case of *linear seeded extractors*, i.e. functions $\mathrm{Ext}\colon \mathbb{F}_2^n \times [D] \to \mathbb{F}_2^m$ such that for each fixed seed $z \in [D]$, the map $x \mapsto \mathrm{Ext}(x, z)$ is linear. Of special interest are (oblivious) *bit-fixing sources*, distributions on $\{0, 1\}^n$ in which a subset of coordinates are fixed adversarially while the remaining coordinates are uniformly random. It is known that random functions act as seedless extractors for such sources, and nearly optimal explicit seedless constructions are available in the nonlinear setting. However, linear seedless extractors for bit-fixing sources cannot achieve strong guarantees, rendering the seeded linear case a natural and important object of study.

There is a direct correspondence between linear extractors and error-correcting codes. A linear map $\mathrm{Ext}\colon \mathbb{F}_2^n \to \mathbb{F}_2^m$ is represented by an $m \times n$ generator matrix G, which in turn defines a binary linear code $C \subseteq \mathbb{F}_2^n$. The ability of Ext to extract randomness from a bit-fixing source is equivalent to the ability of the code C to correct erasures on the set of frozen coordinates. Thus a seeded linear bit-fixing extractor corresponds to an ensemble of linear erasure codes, one for each seed. In particular, if the extractor extracts nearly all entropy, then almost all codes in the ensemble are capacity-achieving for the binary erasure channel, as observed in [1].

The parameters obtainable via the probabilistic method are as follows. For every $X \in \mathbb{F}_2^n$ with $H_\infty(X) \geq \delta n$, there exists a seeded linear bit-fixing extractor outputting $(\delta - \eta)n$ bits with error ε and seed parameter $D = O(1/(\varepsilon\eta))$. This is independent of n, unlike what any general-purpose extractor can achieve. Our main result provides an explicit and strongly explicit seeded linear bit-fixing extractor whose parameters match those guaranteed by the probabilistic method, up to a polynomial factor.

Theorem (Informal). *For every $\delta, \eta, \varepsilon > 0$ and $n \in \mathbb{N}$, there exists a strongly explicit family of linear maps*

$$\mathrm{Ext} : \mathbb{F}_2^n \times [D] \to \mathbb{F}_2^{(\delta-\eta)n}, \qquad D = \left(\frac{1}{\varepsilon\eta}\right)^{O(1)},$$

that extracts from all bit-fixing sources on $\mathbb{F}_2^n$ of min-entropy at least δn with error at most ε.

Beyond the binary erasure channel (as in [1]), this framework extends to more structured erasure models. For example, erasures can be arranged in a matrix structure, where codewords are binary matrices and an adversary is able to erase any δ fraction of the rows and δ' fraction of the columns. Equivalently, codewords can be thought of as adjacency matrices of bipartite graphs where the adversary is allowed to pick a subset of the vertices on each part (up to the designated erasure budget) and erase all edges adjacent to the picked vertices. In this setting, we obtain near-MDS codes achieving rate $R = (1-\delta)(1-\delta') - o(1)$, which is optimal.

We also consider the non-bipartite counterpart where codewords correspond to non-biparite graphs and edges incident to up to δ fraction of the vertices (adversarially chosen) can be erased. Equivalently, this is when codewords are binary square matrices that are symmetric and with zeri diagonals. In this model, we improve the best known explicit constructions by achieving rate $(1-\sqrt{\delta})^4 - o(1)$, improving on earlier results [2, 3].

Furthermore, using our explicit erasure code families, we derive strongly explicit codes over constant sized alphabets whose rate and distance trade-offs are arbitrarily close to the Singleton bound. Thereby, this recovers the celebrated framework of Alon, Edmonds, and Luby [4] but with additional strong explicitness guarantees. That is, each entry of an adjacency matrix for code can be computed in polynomial time in the number bits describing the row and column indices.

The full manuscript for this result is available at [5].

References

[1] Mahdi Cheraghchi, *Capacity achieving codes from randomness conductors*, Proceedings of the Annual IEEE International Symposium on Information Theory (ISIT) (2009), 2639–2643.

[2] Swastik Kopparty, Aditya Potukuchi, and Harry Sha, *Error-correcting graph codes*, Proceedings of the Annual Conference on Innovations in Theoretical Computer Science (ITCS) (2025), 67:1–67:20.

[3] Lev Yohananov, Yuval Efron, and Eitan Yaakobi, *Double and triple node-erasure-correcting codes over complete graphs*, IEEE Transactions on Information Theory, **66(7)** (2020), 4089–4103.

[4] Noga Alon, Jeff Edmonds, and Michael Luby, *Linear time erasure codes with nearly optimal recovery*, Proceedings of the Annual IEEE Symposium on Foundations of Computer Science (FOCS) (1995), 512–519.

[5] Yeyuan Chen, Mahdi Cheraghchi, and Nikhil Shagrithaya, *Optimal erasure codes and codes on graphs*, **arXiv:2504.03090** (2025).

Explicit Lossless Vertex Expanders

Rachel Zhang

(joint work with Jun-Ting Hsieh, Ting-Chun Lin, Alexander Lubotzky, Sidhanth Mohanty, Ryan O'Donnell, and Assaf Reiner)

We give the first explicit construction of lossless vertex expanders. These are d-regular graphs where every small set S of vertices has $(1-\varepsilon)d|S|$ distinct neighbors. Previously, the strongest known explicit vertex expanders were those given by Ramanujan graphs, whose spectral properties imply that every small set S of vertices has $0.5d|S|$ distinct neighbors [1]. Unfortunately, we also know that $0.5d|S|$ is the best one can do with spectral tools [2]. In our work [3, 4], we identify a new connection between strong vertex expansion and high dimensional expanders. Using a construction based on Ramanujan cubical complexes [5], we construct the first explicit lossless vertex expanders.

References

[1] Nabil Kahale, *Eigenvalues and expansion of regular graphs*, Journal of the ACM (JACM) **42(5)** (1995), 1091–1106.

[2] Amitay Kamber and Tali Kaufman, *Combinatorics via closed orbits: number theoretic Ramanujan graphs are not unique neighbor expanders*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (2022), 426–435.

[3] Jun-Ting Hsieh, Ting-Chun Lin, Sidhanth Mohanty, Ryan O'Donnell, and Rachel Yun Zhang, *Explicit Two-Sided Vertex Expanders Beyond the Spectral Barrier*, Proceedings of the 57th Annual ACM Symposium on Theory of Computing (2025).

[4] Jun-Ting Hsieh, Alexander Lubotzky, Sidhanth Mohanty, Assaf Reiner, and Rachel Yun Zhang, *Explicit Lossless Vertex Expanders*, Proceedings of the 52nd annual 66th IEEE Symposium on Foundations of Computer Science (FOCS) (2025).

[5] Bruce W Jordan and Ron Livné, *The Ramanujan property for regular cubical complexes*, Duke Math. J. **104(1)** (2000), 85–103.

Asymptotically Good Quantum LDPC Codes (Survey Talk)

Gilles Zémor

In this talk we give a short, partial account of the ideas leading up to the construction of asymptotically good quantum LDPC codes, and of the present state of knowledge.

A quantum CSS code of length n is defined by two classical binary codes of length n, C_X and C_Z, with the property that their dual codes $C_X^\perp$ and $C_Z^\perp$ are orthogonal to each other. The quantum code dimension is given by $k = n - \dim C_X^\perp - \dim C_Z^\perp$. The quantum code minimum distance d is defined as the minimum of two quantities (distances), d_X and d_Z, that are defined respectively as the smallest weight of a vector of C_X (C_Z) not in $C_Z^\perp$ (not in $C_X^\perp$). We are interested in quantum LDPC codes, meaning that we wish the two classical codes C_X and C_Z to be LDPC, i.e., defined by two parity-check matrices of bounded row and column weights.

The theory of quantum LDPC codes starts with the Kitaev toric code [1]. The Kitaev code, in its original instantiation, has parameters of the form $[[n = 2m^2, k = 2, d = m]]$, though they can also take the form $[[m^2, 2, m]]$. The toric code has many desirable features, but one wishes to improve both its dimension and its minimum distance, ideally obtaining a behaviour that is linear in the codelength n.

Improving the minimum distance and obtaining a growth that surpasses $\sqrt{n}$ is arguably the most challenging of the two tasks. In 2002, Freedman et al. [2] obtained a construction of codes of dimension 2 with a distance that scales as $\sqrt{n}\log^{1/4} n$. This stayed the state of the art for almost twenty years, until this record was extended to $\sqrt{n}\log n$ in [3], with follow-up work appearing in [4], borrowing some ideas from the theory of high-dimensional expansion.

In the meantime, the introduction of *Hypergraph Product* codes [5] allowed one to transform an asymptotically good classical LDPC code into a quantum code with a linear dimension and a minimum distance that scales as $\sqrt{n}$, i.e. no worse than the toric code.

On the minimum distance front, we witnessed a significant breakthrough with the advent of *Fiber Bundle codes* [6]: these codes have non-linear dimension but were shown to achieve a minimum distance of $n^{3/5}$. The central idea consists of introducing a *twist* in the hypergraph product construction, which breaks the $\sqrt{n}$ barrier inherent in the untwisted scheme. Soon afterwards, a minimum distance of $n/\log n$ was achieved [7]: this work essentially used the same construction as that of the fiber bundle code paper, but with a more efficient framework for analysing the minimum distance, and also using as base classical code a Tanner code, rather than an all-purpose LDPC code as in [6].

Very soon afterwards, Panteleev and Kalachev finally solved the longstanding problem of constructing asymptotically good quantum LDPC codes with their follow-up work [8]. This entailed refining the twisted product ideas of [6, 7] and solving many related issues. The introduction of *Quantum Tanner codes* [9, 10] can both be seen as a simplification of [8], and a break from the product strategy, relying instead on square complexes, that were previously used to construct classical locally testable codes [11]. The construction [12] reverted to a product strategy and can be thought of as a dual variant of [8]. Finally, the very recent construction [13] relies on the construction of lossless expanders that come with a group action, and is the first construction of asymptotically good LDPC codes that does not use the constant size inner codes of the Tanner coding paradigm.

References

[1] A. Y. Kitaev, *Quantum computations: algorithms and error correction*, Uspekhi Mat. Nauk **52** (1997), 53–112.

[2] M. H. Freedman, D. A. Meyer, and F. Luo, *Z_2-systolic freedom and quantum codes*, Mathematics of quantum computation, Comput. Math. Ser. (2002), 287–320.

[3] S. Evra, T. Kaufman, and G. Zémor, *Decodable quantum LDPC codes beyond the $\sqrt{n}$ distance barrier using high-dimensional expanders*, SIAM Journal on Computing, **53** (2024), FOCS20–276–FOCS20–316.

[4] T. Kaufman and R. J. Tessler, *New cosystolic expanders from tensors imply explicit quantum LDPC codes with $\Omega(\sqrt{n}\log^k n)$ distance*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC (2021), 1317–1329.

[5] J.-P. Tillich and G. Zémor, *Quantum LDPC codes with positive rate and minimum distance proportional to the square root of the blocklength*, IEEE Transactions on Information Theory **60** (2014), 1193–1202.

[6] M. B. Hastings, J. Haah, and R. O'Donnell, *Fiber bundle codes: breaking the n1/2 polylog(n) barrier for quantum LDPC codes*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, STOC (2021), 1276–1288.

[7] P. Panteleev and G. Kalachev, *Quantum LDPC codes with almost linear minimum distance*, IEEE Transactions on Information Theory **68** (2022), 213–229.

[8] P. Panteleev and G. Kalachev, *Asymptotically good quantum and locally testable classical LDPC codes*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC (2022), 375–388.

[9] A. Leverrier and G. Zémor, *Quantum Tanner codes*, 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science, FOCS (2022), 872–883.

[10] A. Leverrier and G. Zémor, *Decoding quantum Tanner codes*, IEEE Transactions on Information Theory **69** (2023), 5100–5115.

[11] I. Dinur, S. Evra, R. Livne, A. Lubotzky, and S. Mozes, *Locally testable codes with constant rate, distance, and locality*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, STOC (2022), 357–374.

[12] I. Dinur, M.-H. Hsieh, T.-C. Lin, and T. Vidick, *Good quantum LDPC codes with linear time decoders*, Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC (2023), 905–918.

[13] J.-T. Hsieh, A. Lubotzky, S. Mohanty, A. Reiner, and R. Y. Zhang, *Explicit lossless vertex expanders*, **arXiv:2504.15087** (2025).

Decoded Quantum Interferometry

Alexander Schmidhuber

(joint work with Stephen P. Jordan, Noah Shutty, Mary Wootters, Adam Zalcman, Robbie King, Sergei V. Isakov, Tanuj Khattar, and Ryan Babbush)

This talk is based on the recent work [1]. I firstly described Decoded Quantum Interferometry (DQI), a quantum algorithm for reducing classical optimization problems to classical decoding problems by exploiting structure in the Fourier spectrum of the objective function. DQI reduces sparse max-XORSAT to decoding LDPC codes, which can be achieved using powerful classical algorithms such as Belief Propagation (BP). As an initial benchmark, we compared DQI using belief propagation decoding against classical optimization via simulated annealing. In this setting we presented evidence that, for a certain family of max-XORSAT instances, DQI with BP decoding achieves a better approximation ratio on average than simulated annealing, although not better than specialized classical algorithms tailored to those instances. We also analyzed a combinatorial optimization problem corresponding to finding polynomials that intersect the maximum number of points. There, DQI efficiently achieves a better approximation ratio than any polynomial-time classical algorithm known to us, thus realizing an apparent exponential quantum speedup. Finally, we showed that the problem defined by Yamakawa and Zhandry in order to prove an exponential separation between quantum and classical query complexity is a special case of the optimization problem efficiently solved by DQI.

I further presented the ongoing work on generalizing DQI from classical cost functions to general non-diagonal and non-commuting Hamiltonians. In this context, I introduced Hamiltonian Decoded Quantum Interferometry (HDQI), a quantum algorithm that utilizes coherent Bell measurements and the symplectic representation of the Pauli group to reduce Gibbs sampling and Hamiltonian optimization to classical decoding. The decoding problem inherits structural properties of H; in particular, local Hamiltonians map to LDPC codes. Preparing the pilot state is always efficient for commuting Hamiltonians, but highly non-trivial for non-commuting Hamiltonians. Nevertheless, we proved that this state admits an efficient matrix product state representation for a class of nearly commuting Pauli Hamiltonians whose anti-commutation graph decomposes into connected components of logarithmic size. At the end, we discussed avenues and potential future directions for applying HDQI to Hamiltonians appearing in quantum optimization, quantum chemistry, and beyond.

References

[1] S.P. Jordan, N. Shutty, M. Wootters, A. Zalcman, A. Schmidhuber, R. King, S.V. Isakov, T. Khattar, and R. Babbush. *Optimization by decoded quantum interferometry*, **arXiv:2408.08292** (2024).

(Almost) Good Quantum Codes via Tensor Products

Louis Golowich
(joint work with Venkatesan Guruswami)

Quantum low-density parity-check (qLDPC) codes provide one of the most promising means for achieving efficient quantum fault-tolerance. Such codes are defined to support error detection and correction via sparse queries (i.e. stabilizer measurements) to code states. Each such query involves, and can therefore only propagate errors across, a small number of code components. However, this qLDPC condition has proven difficult to achieve, and asymptotically optimal qLDPC codes were only recently constructed following a line of breakthrough works [1, 2, 3, 4, 5, 6]. The latter three of these works use nearly identical techniques, and still provide the only known asymptotically good qLDPC codes, meaning that the code dimension and distance scale linearly in the block length, and the stabilizers have constant weight. It remains an open question to find new constructions of qLDPC codes with good parameters. Such alternative constructions may yield better practical parameters or have properties useful for fault-tolerant computation, and would also be of independent mathematical interest.

In this work, we develop a new construction of qLDPC codes with close-to-linear dimension and distance based on *homological products* (i.e. tensor products of chain complexes), a well-known construction from homological algebra that generalizes classical tensor codes. To do so, we provide a new method of taking homological products that preserve almost-linear distance by using *subsystem codes*, which only encode messages into a subspace of the entire logical code space. Our analysis of this method builds on techniques of [12]. As a result, we obtain a new construction of qLDPC codes of close-to-linear dimension and distance with constant stabilizer weight using an iterative construction, which is based on iterative homological products of a constant-sized code. We provide an informal statement below:

Theorem 1. *For every $\epsilon > 0$, there exists a constant-sized code $\mathcal{C}$ and an infinite sequence of qLDPC subsystem codes $(Q_i)_{i\in\mathbb{N}}$ with parameters*

$$[[N_i, N_i^{1-\epsilon}, N_i^{1-\epsilon}]]_2$$

and with constant stabilizer weight $O(1)$ (independent of ϵ), such that each Q_i is obtained by applying the stabilizer-weight-reduction transformation of [10] *to the homological product of Q_{i-1} with $\mathcal{C}$.*

The weight-reduction step in Theorem 1 is needed to keep the stabilizer weight constant in each iteration, as in general the stabilizer weight of a homological product code grows as the sum of the stabilizer weights of the input codes.

Most prior homological product code constructions had distance at most $\tilde{O}(\sqrt{N})$. We exceed this bound and obtain almost-linear distance in Theorem 1 by using subsystem codes. Specifically, we observe that the "$\tilde{O}(\sqrt{N})$ barrier" only applies to certain logical operators (i.e. codewords) within the code space. Therefore, we show that appropriate homological product codes will still have good distance for a large subspace of the logical operators. Thus appropriate subsystem codes of these products will still have good distance. Our distance bound proofs are inspired by the techniques of [12], which showed a related result, but that did not use subsystem codes.

To prove Theorem 1, we set $\mathcal{C}$ to be a constant-sized quantum locally testable code (qLTC; see the full version) such as one given by [9, 11]; such a qLTC (which we emphasize is just constant-sized) is needed for our subsystem distance bound described above. We apply our distance bound to show that the distance of each Q_i remains close-to-linear (i.e. $\geq N_i^{1-\epsilon}$) in the block length N_i.

This iterative construction is reminiscent of a line of work in classical pseudorandomness, coding theory, and complexity theory, which iteratively builds larger objects with properties of interest from smaller or weaker ones. For instace, [7] constructed classical locally testable codes by iterative tensoring of a small classical code; at a high level, Theorem 1 uses a related approach to construct qLDPC codes. Other notable iterative constructions in the literature include explicit spectral expanders from the zig-zag product [14], Dinur's proof of the PCP theorem [8], and undirected connectivity in logarithmic space [13]. Theorem 1 can therefore be viewed as an analogue of these results for qLDPC codes. However, our construction does need to start with a strong constant-sized object, namely, a constant-sized qLTC with sufficiently good parameters. It is an interesting open question whether there exists an iterative construction with a weaker starting object.

Our work also raises multiple additional open questions. For instance, the iterative product-based structure of our codes appears to be more flexible than prior balanced product constructions. Indeed, we could use different constant-sized codes $\mathcal{C}$ in different iterations of the product, and still obtain almost-good parameters for the resulting codes. It would be interesting to see if this additional flexibility could be leveraged, for instance in applications to fault-tolerant computation.

Furthermore, our iterative product paradigm provides one alternative to balanced products. It remains an open problem to find additional, fundamentally different constructions of qLDPC codes with good or almost good parameters. One interesting construction towards this goal is given by [10], which obtains qLDPC codes of distance $\tilde{\Omega}(N^{2/3})$ (which is beyond the $\tilde{O}(\sqrt{N})$ barrier) by applying weight-reduction to an appropriate random quantum code. Perhaps similar techniques could obtain even closer to linear distance, without using any sort of product.

References

[1] Matthew B. Hastings, Jeongwan Haah, and Ryan O'Donnell, *Fiber bundle codes: breaking the n1/2 polylog(n) barrier for Quantum LDPC codes*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (2021), 1276–1288.

[2] Pavel Panteleev and Gleb Kalachev, *Quantum LDPC Codes With Almost Linear Minimum Distance*, IEEE Transactions on Information Theory **68(1)** (2022), 213–229.

[3] Nikolas P. Breuckmann and Jens N. Eberhardt, *Balanced product quantum codes*, IEEE Transactions on Information Theory **67(10)** (2021), 6653–6674.

[4] Pavel Panteleev and Gleb Kalachev, *Asymptotically good Quantum and locally testable classical LDPC codes*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (2022), 375–388.

[5] Anthony Leverrier and Gilles Zémor, *Quantum Tanner codes*, Proceedings of the IEEE 63rd Annual Symposium on Foundations of Computer Science (2022), 872–883.

[6] Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick, *Good Quantum LDPC Codes with Linear Time Decoders*, Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC (2023), 905–918.

[7] Eli Ben-Sasson and Madhu Sudan. *Robust locally testable codes and products of codes*, Proceedings of the International Workshop on Randomization and Approximation Techniques in Computer Science (2004), 286–297.

[8] Irit Dinur, *The PCP theorem by gap amplification*, Journal of the ACM **54(3)** (2007), 12–es.

[9] Irit Dinur, Ting-Chun Lin, and Thomas Vidick, *Expansion of higher-dimensional cubical complexes with application to quantum locally testable codes, April 2024*, **arXiv:2402.07476** (2024).

[10] M. B. Hastings, *On quantum weight reduction*, **arXiv:2102.10030** (2023).

[11] Gleb Kalachev and Pavel Panteleev, *Maximally extendable product codes are good coboundary expanders*, **arXiv:2501.01411** (2025).

[12] Tali Kaufman and Ran J. Tessler, *New cosystolic expanders from tensors imply explicit Quantum LDPC codes with Omega(sqrt{n} log^k n) distance.* Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (2021), 1317–1329.

[13] Omer Reingold, *Undirected connectivity in log-space*, Journal of the ACM **55(4)** (2008), 17:1–17:24.

[14] Omer Reingold, Salil Vadhan, and Avi Wigderson, *Entropy waves, the Zig-Zag graph product, and new constant-degree expanders*, Annals of Mathematics **155(1)** (2002), 157–187.

Survey on Coding for Interactive Communiction (Surver Talk)

Klim Efremenko

Error Correcting Codes (ECCs) address a fundamental question: how can we ensure reliable data transmission over unreliable channels? Since Shannon's pioneering 1948 work, seven decades of research have created a rich theoretical framework with practical impacts across diverse fields.

Today, communication systems are no longer just about transmitting information. Instead, they often involve interactive processes requiring multiple exchanges between participants, as seen in cloud computing, cryptographic protocols, and distributed systems.

In this talk, we surveyed recent results on crafting interactive error-correcting codes and uncovering their fundamental limits.

In particular, we talked about the following topics:

(1) adaptive vs non-adaptive channels.
(2) The role of feedback in the coding for interactive communication.
(3) Coding for interactive communication for non interactive tasks

On Deterministic LDCs and Their Application in Distributed Computing

Ran Gelles
(joint work with Keren Censor-Hillel, Orr Fischer, and Pedro Soto)

We examine distributed computations in the “Congested Clique” model [3], in which an $\alpha < 1$ fraction of nodes may crash during the computation. To avoid information loss, data is preserved in the network using locally decodable codes (LDCs). While prior work [2] relied on standard error-correcting codes (ECC), our LDC construction allows access to only the necessary information, thereby reducing overall complexity.

In the model we consider, crashes are equivalent to erasures in the LDC codeword. This allows us to derandomize the scheme and obtain a deterministic LDC decoder that decodes by querying a sufficient number of non-erased indices while maintaining low congestion during multiple simultaneous decodings. To that end, we present a slight variant of LDCs for erasures: the decoder either outputs the correct symbol or reports failure; it never outputs an incorrect symbol. Further exploration of erasure LDCs and their connections to standard LDCs is a promising direction.

References

[1] K. Censor-Hillel, O. Fischer, R. Gelles, P. Soto, *Two for One, One for All: Deterministic LDC–based Robust Computation in Congested Clique*, International Symposium on Distributed Computing (DISC) (2025).

[2] K. Censor-Hillel and P. Soto. *Computing in a faulty congested clique.* **arXiv:2505.11430** (2025).

[3] Z. Lotker, B. Patt-Shamir, E. Pavlov, and D. Peleg. *Minimum-weight spanning tree construction in* $O(\log\log n)$ *communication rounds*, SIAM J. on Computing **35(1)** (2005), 120–131.

Representational Methods in Coding Theory (Survey Talk)

Wolfgang Willems

If $C \subseteq \mathbb{F}^n$ is a linear code of dimension k over the finite field $\mathbb{F}$, then with a very high probability the automorphism group of C is trivial. However, almost all codes we know and are dealing with have a nontrivial automorphism group $G = \mathrm{Aut}(C)$. Thus C is not only a vector space, but a G-module. Often it is a submodule of the group algebra $\mathbb{F}G$ which we call a G-code or a group code. Thus in order to analyze C we may use the full machinery of representation theory of finite groups.

For group codes $C \subseteq \mathbb{F}G$, a crucial connection between representation theory and coding theory is given by the $\mathbb{F}G$-isomorphism $\mathbb{F}G/C^\perp \cong C^*$ where C^* denotes the dual module of C. As a consequence, a self-dual composition factor in $\mathbb{F}G$ must have an even multiplicity in $\mathbb{F}G$. Using this fact and Maschke’s theorem we get the following.

Theorem. If $C \subseteq \mathbb{F}G$ is a self-dual G-invariant code, then the characteristic of $\mathbb{F}$ is 2 and 2 divides the order of G.

Using the obvious fact that, for a cyclic 2-group C_{2^n} and a field $\mathbb{F}$ of characteristic 2, the group algebra $\mathbb{F}C_{2^n}$ has only one self-dual code, we get an easy representational proof of the following result due to Sloane and Thompson.

Theorem. If the Sylow 2-subgroup of G is cyclic, then a self-dual group code in $\mathbb{F}_2G$ is never doubly even.

Finally, we extend a result of Yang and Massey on LCD (linear complementary dual) codes from cyclic codes to general group codes by applying the fact that an LCD group code is always a projective module for the underlying group algebra, hence generated by an idempotent.

Theorem. A group code $C \leq \mathbb{F}G$ is an LCD code if and only if $C = e\mathbb{F}G$, $e = e^2 = \hat{e}$ where $\hat{e}$ is the adjoint of the idempotent e.

These and some other results enlighten the powerful methods of using representation theory in coding theory.

For more information on group codes and references to the above results we refer to the article “Codes in group algebras”, Chapter 16 in Concise Encyclopedia in Coding Theory, Chapmann and Hall/CRC 2021, 363–384.

Soft-Decision Decoding of Recursive Plotkin Constructions using Hidden Code Words

Martin Bossert

The Plotkin construction [1] from 1960 combines two codes to obtain a code of doubled length. It can be applied recursively. In [2] the hidden code words are introduced which are used in the first step of decoding of these recursive constructions. These hidden code words can be uncovered by adding particular parts of the overall code word. The main idea is to use more than one decoding variant where each variant starts with the decoding of a different hidden code word. Given the decoding of the first hidden code word is correct the number of errors is reduced for the remaining decoding steps. As final decoding decision the best of the decisions of the used variants is choosen. Using more variants the performance gets closer to the maximum-likelihood (ML) decoding performance. The class of Reed–Muller (RM) codes is a particular example for the use of this decoding with hidden code words.

The Plotkin construction uses two codes $\mathcal{C}_0(n, k_0, d_0)$ and $\mathcal{C}_1(n, k_1, d_1)$. These codes are combined to a code $\mathcal{C}$ of double length by choosing two code words $\mathbf{u}_0 \in \mathcal{C}_0$ and $\mathbf{u}_1 \in \mathcal{C}_1$ and appending the code word $\mathbf{u}_0 + \mathbf{u}_1$ to the code word $\mathbf{u}_0$ which results in the code word $\mathbf{c} = (\mathbf{u}_0|\mathbf{u}_0 + \mathbf{u}_1)$ of the new code $\mathcal{C}(2n, k_0 + k_1, \min\{2d_0, d_1\})$. The double Plotkin construction uses four component codes $\mathcal{C}_0$,$\mathcal{C}_1$,$\mathcal{C}_2$, and $\mathcal{C}_3$ and the two Plotkin constructions $(\mathbf{u}_0|\mathbf{u}_0 + \mathbf{u}_1)$ and $(\mathbf{u}_2|\mathbf{u}_2 + \mathbf{u}_3)$.

Table 1. Soft and Hard Join-Two of the Four Blocks

$\mathbf{y}_0 \bowtie \mathbf{y}_1 = \mathbf{x}_1 + \mathbf{z}'$	$\mathbf{u}_1 + \mathbf{e}_0 + \mathbf{e}_1$
$\mathbf{y}_0 \bowtie \mathbf{y}_2 = \mathbf{x}_2 + \mathbf{z}'$	$\mathbf{u}_2 + \mathbf{e}_0 + \mathbf{e}_2$
$\mathbf{y}_0 \bowtie \mathbf{y}_3 = \mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 + \mathbf{z}'$	$\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \mathbf{e}_0 + \mathbf{e}_3$
$\mathbf{y}_1 \bowtie \mathbf{y}_2 = \mathbf{x}_1\mathbf{x}_2 + \mathbf{z}'$	$\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{e}_1 + \mathbf{e}_2$
$\mathbf{y}_1 \bowtie \mathbf{y}_3 = \mathbf{x}_2\mathbf{x}_3 + \mathbf{z}'$	$\mathbf{u}_2 + \mathbf{u}_3 + \mathbf{e}_1 + \mathbf{e}_3$
$\mathbf{y}_2 \bowtie \mathbf{y}_3 = \mathbf{x}_1\mathbf{x}_3 + \mathbf{z}'$	$\mathbf{u}_1 + \mathbf{u}_3 + \mathbf{e}_2 + \mathbf{e}_3$

If we apply the Plotkin construction for these two codes we get a code of length $4n$

$$(\mathbf{u}_0|\mathbf{u}_0 + \mathbf{u}_1|\mathbf{u}_0 + \mathbf{u}_2|\mathbf{u}_0 + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3) = (\mathbf{a}_0|\mathbf{a}_1|\mathbf{a}_2|\mathbf{a}_3).$$

If the component codes have the property $\mathcal{C}_0, \mathcal{C}_2 \subseteq \mathcal{C}_1$, and $\mathcal{C}_3 \subset \mathcal{C}_2$ then the uncovered code word $\mathbf{a}_0 + \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3$ is from code $\mathcal{C}_3$. The two uncovered code words $\mathbf{a}_0 + \mathbf{a}_2$ and $\mathbf{a}_1 + \mathbf{a}_3$ are from the code $\mathcal{C}_2$ and the four uncovered code words $\mathbf{a}_0 + \mathbf{a}_1,\ \mathbf{a}_0 + \mathbf{a}_3,\ \mathbf{a}_1 + \mathbf{a}_2,\ \mathbf{a}_2 + \mathbf{a}_3$ are from the code $\mathcal{C}_1$. In case of $\mathcal{C}_1 = \mathcal{C}_2$, all six code words are from this code. If additionally $\mathcal{C}_1 \subset \mathcal{C}_0$, all blocks $\mathbf{a}_i$ are from code $\mathcal{C}_0$.

For soft-decision decoding we use binary phase shift keying (BPSK) and the usual mapping of the binary code symbols $u_i = 0 \leftrightarrow x_i = 1$ and $u_i = 1 \leftrightarrow x_i = -1$, and for vectors $\mathbf{u} \leftrightarrow \mathbf{x}$. In the additive white Gaussian noise (AWGN) channel we receive $y_i = x_i + z_i$ where z_i denotes the Gaussian noise. The addition of two binary code words is a component-wise multiplication of the BPSK-modulated code words $\mathbf{u}_0 + \mathbf{u}_1 \leftrightarrow \mathbf{x}_0\mathbf{x}_1$, where

$$\mathbf{x}_0\mathbf{x}_1 = (x_{0,0}x_{1,0}, x_{0,1}x_{1,1}, \dots, x_{0,n-1}x_{1,n-1})$$

with $x_{i,j} = \pm 1$.

Transmitting the modulated code word over a Gaussian channel we receive $(\mathbf{y}_0 = \mathbf{x}_0 + \mathbf{z}_0|\mathbf{y}_1 = \mathbf{x}_0\mathbf{x}_1 + \mathbf{z}_1|\mathbf{y}_2 = \mathbf{x}_0\mathbf{x}_2 + \mathbf{z}_2|\mathbf{y}_3 = \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 + \mathbf{z}_3)$. For uncovering the hidden code words we define the join operation $\bowtie$ by

$$y_i \bowtie y_j = \operatorname{sign}(y_i y_j) \min\{|y_i|, |y_j|\}$$

which is sometimes called minsum approximation. The join operation for vectors is done component wise and it is commutative and associative. In Table 1 examples of the uncovering of several hidden code words are shown, on the left side the Gaussian case and right the hard-decision case with errors.

Table 2 shows the operations used in the variants for decoding the complete douple Plotkin construction. The error cancelation is due to the fact that the coherent addition of two received values increase the signal ernergy by a factor of four while the variance of the addition of two Gaussian variables is only increasing by a factor of two.

As an example we describe one decoding variant, namely $\mathrm{V}(\mathbf{y}_0 \bowtie \mathbf{y}_1)$ where the decoding is started according the first of the six possible uncoverings from

Table 2. Join and Add Operations of the Four Blocks

join-four	$\mathbf{y}_0 \bowtie \mathbf{y}_1 \bowtie \mathbf{y}_2 \bowtie \mathbf{y}_3 \to \mathbf{x}_3 + \mathbf{z}'$
add-two	$\mathbf{y}_0 + \mathbf{y}_2\mathbf{x}_2 \to 2\mathbf{x}_0 + \mathbf{z}'$
add-four	$\mathbf{y}_0 + \mathbf{y}_1\mathbf{x}_1 + \mathbf{y}_2\mathbf{x}_2 + \mathbf{y}_3\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3 \to 4\mathbf{x}_0 + \mathbf{z}'$
join-add	$(\mathbf{y}_0 \bowtie \mathbf{y}_1) + (\mathbf{y}_2 \bowtie \mathbf{y}_3\mathbf{x}_3) \to 2\mathbf{x}_1 + \mathbf{z}'$
add-join	$(\mathbf{y}_0 + \mathbf{y}_2\mathbf{x}_2) \bowtie (\mathbf{y}_1 + \mathbf{y}_3\mathbf{x}_2\mathbf{x}_3) \to \mathbf{x}_1 + \mathbf{z}'$

Table 1. The join-two operation $\mathbf{y}_0 \bowtie \mathbf{y}_1$ is a noisy version of $\mathbf{x}_1$. With a decoder for the code $\mathcal{C}_1$ we get the estimate $\hat{\mathbf{x}}_1$. Knowing the estimate of $\hat{\mathbf{x}}_1$ the join-two operation $\mathbf{y}_2 \bowtie \mathbf{y}_3\hat{\mathbf{x}}_1$ is a noisy version of $\mathbf{x}_3$ and with a decoder for the code $\mathcal{C}_3$ we get the estimate $\hat{\mathbf{x}}_3$. Now we have the estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_3$ and use the add-join operation $(\mathbf{y}_0 + \mathbf{y}_1\hat{\mathbf{x}}_1) \bowtie (\mathbf{y}_2 + \mathbf{y}_3\hat{\mathbf{x}}_1\hat{\mathbf{x}}_3)$ which is a noisy version of $\mathbf{x}_2$. With a decoder for the code $\mathcal{C}_2$ we get the estimate $\hat{\mathbf{x}}_2$. The add-four operation $\mathbf{y}_0 + \mathbf{y}_1\hat{\mathbf{x}}_1 + \mathbf{y}_2\hat{\mathbf{x}}_2 + \mathbf{y}_3\hat{\mathbf{x}}_1\hat{\mathbf{x}}_2\hat{\mathbf{x}}_3$ is a noisy version of $\mathbf{x}_0$ and with a decoder for the code $\mathcal{C}_0$ we get the estimate $\hat{\mathbf{x}}_0$ which completes the decoding. The other variants start with other hidden code words according Table 1.

The RM code $\mathcal{C}(128, 64, 16)$ can be constructed as follows. Take the codes $\mathcal{C}(8, 8, 1)$, the parity-check code $\mathcal{C}(8, 7, 2)$, the extended Hamming code $\mathcal{C}(8, 4, 4)$, and the repetition code $\mathcal{C}(8, 1, 8)$. With these codes we can construct with the double Plotkin construction three codes. Using $\mathcal{C}_0''(8, 8, 1)$, $\mathcal{C}_1''(8, 7, 2) = \mathcal{C}_2''$, and $\mathcal{C}_3''(8, 4, 4)$ we get the code $\mathcal{C}_0'(32, 26, 4)$. Similarly, with $\mathcal{C}_0''(8, 7, 2)$, $\mathcal{C}_1''(8, 4, 4) = \mathcal{C}_2''$, and $\mathcal{C}_3''(8, 1, 8)$ we get $\mathcal{C}_1'(32, 16, 8)$. Choosing $\mathcal{C}_0''(8, 4, 4)$,$\mathcal{C}_1''(8, 1, 8) = \mathcal{C}_2''$, and for $\mathcal{C}_3''$ the all-zero code word we get $\mathcal{C}_3'(32, 6, 16)$. The double Plotkin construction of these three codes gives $\mathcal{C}(128, 64, 16)$ where $\mathcal{C}_0'(32, 26, 4)$,$\mathcal{C}_1'(32, 16, 8) = \mathcal{C}_2'$,$\mathcal{C}_3'(32, 6, 16)$. The decoding of this code of length 128 and dimension 64 is done by decoding only the four codes of length 8 multiple times. Note that the complexity for decoding a code of length 8 is very small. However, using enough decoding variants which start with different hidden code words the decoding performance is the same as for ML decoding.

An interesting observation with this decoding of double Plotkin constructions is for first-order RM codes. It can be explaind by the example of the $\mathcal{C}(8, 4, 4)$ code. The construction uses the codes $\mathcal{C}_0(2, 2, 1)$, $\mathcal{C}_1(2, 1, 2) = \mathcal{C}_2$, and $\mathcal{C}_3$ is the all-zero code word. The received vector is $(\mathbf{y}_0 = \mathbf{x}_0 + \mathbf{z}_0 | \mathbf{y}_1 = \mathbf{x}_0\mathbf{x}_1 + \mathbf{z}_1 | \mathbf{y}_2 = \mathbf{x}_0\mathbf{x}_2 + \mathbf{z}_2 | \mathbf{y}_3 = \mathbf{x}_0\mathbf{x}_1\mathbf{x}_2 + \mathbf{z}_3) = (y_0, y_1, y_2, y_3, y_4, y_5, y_6, y_7)$. Since the two repetition codes have only two possibilities we have only four add-four (from Table 2) possibilities for decoding the code $\mathcal{C}_0(2, 2, 1)$. The two symbols of this code are denoted by (w_0, w_1). Then the four possibilities for (w_0, w_1) are $(y_0 + y_2 + y_4 + y_6, y_1 + y_3 + y_5 + y_7)$, $(y_0 - y_2 + y_4 - y_6, y_1 - y_3 + y_5 - y_7)$, $(y_0 + y_2 - y_4 - y_6, y_1 + y_3 - y_5 - y_7)$, or $(y_0 - y_2 - y_4 + y_6, y_1 - y_3 - y_5 + y_7)$, depending on $\mathbf{x}_1$ and $\mathbf{x}_2$ which can be $(1, 1)$ or $(-1, -1)$. Note that several additions appear twice in these equations and thus, have to be calculated only once. For example $y_0 + y_2$ or $y_1 + y_3$. Choosing the maximum of $|w_0| + |w_1|$ gives the ML decision. Surprisingly this method needs less additions than using the Fast-Hadamard transform for ML decoding.

References

[1] M. Plotkin, *Binary codes with specific minimum distances*, IEEE Trans. on Inf. Theory **6** (1960), 445–450.

[2] M. Bossert, *Soft Decision Decoding of Recursive Plotkin Constructions Based on Hidden Code Words*, IEEE Trans. on Inf. Theory **71** (2025), 4228–4249.

Codes for Computationally Bounded Channels (Survey Talk)

Jad Silbak

(joint work with Ronen Shaltiel and Daniel Wichs)

This talk is based on a series of works with Ronen Shaltiel and Daniel Wichs. We consider error-correcting codes for channels that are computationally bounded and may corrupt up to a p-fraction of the codeword's bits. These channels are considerably more powerful than Shannon's binary symmetric channel (which flips each bit independently with probability p), yet weaker than Hamming's worst-case channels (which can flip any p-fraction of bits without computational limits).

Lipton [1] and Guruswami and Smith [2], argued that the error induced in real-world applications is typically not fully adversarial, and could often be simulated by weak classes of channels. Thus, for most practical applications, it is often sufficient to construct codes for computationally bounded adversaries. Moreover, this restriction on the adversary makes it possible to construct codes with a rate that is strictly better than what is possible when the adversary is information theoretic.

In recent years, there has been a growing body of work that aims to construct codes against channels that are computationally bounded (e.g., bounded memory channels, channels that run in a fixed polynomial time, and channels that can run in any polynomial time). In this talk, we will survey these results, focusing on channel capacities and the techniques used to obtain explicit, uniquely decodable codes that surpass the information-theoretic rate limitations.

We highlight two main research directions:

(1) *Channels weaker than the encoder/decoder:* Following Guruswami and Smith, one line of work considers adversaries that are weaker than the encoding and decoding algorithms. Here, we focus on space-bounded channels and those restricted to fixed polynomial time. We will discuss techniques developed by Shaltiel and Silbak [7, 8, 9] that achieve unique decoding with a better rate than what is possible information theoretically. A central notion here is the evasiveness of codes which is concerned with whether a decoding algorithm for say, binary symmetric channels, rejects a word that is obtained when a computationally bounded channel induces few errors to a uniformly chosen (or pseudorandom) string.

(2) *Channels with full polynomial-time power:* A second line of work, initiated by Micali, Peikert, Sudan, and Wilson [3], studies channels that can run

any polynomial-time algorithm—including the encoder and decoder themselves. In this stronger setting, codes are achievable only under cryptographic assumptions, which are shown to be necessary. We present recent results in this direction. In particular, Silbak and Wichs [4] give a simple construction over a large constant alphabet that yields uniquely decodable codes with rates matching those of list-decodable codes. This construction relies on one-way functions, which they also show to be necessary. Finally, we will explain how to obtain explicit binary codes in this setting under standard cryptographic assumptions [5, 6], using multi-input correlation-intractable hash functions, a powerful cryptographic primitive.

References

[1] Richard J. Lipton, *A new approach to information theory*, 11th Annual Symposium on Theoretical Aspects of Computer Science (1994), 699–708.

[2] Venkatesan Guruswami and Adam Smith, *Optimal rate code constructions for computationally simple channels*, Journal of the ACM (JACM) **63(4)** (2016), 1–37.

[3] Silvio Micali, Chris Peikert, Madhu Sudan, and David A. Wilson, *Optimal error correction for computationally bounded noise*, IEEE Trans. Information Theory **56(11)** (2010), 5673–5680.

[4] Jad Silbak and Daniel Wichs, *Detecting and correcting computationally bounded errors: A simple construction under minimal assumptions*, 16th Innovations in Theoretical Computer Science Conference (ITCS 2025), 88–1.

[5] Jad Silbak and Daniel Wichs, *Binary codes for error detection and correction in a computationally bounded world*, Annual International Conference on the Theory and Applications of Cryptographic Techniques (2025), 186–211.

[6] G. Lu, J. Silbak, and D. Wichs, *Binary codes for computationally bounded errors under standard crypto assumptions*, Proceedings of the 66th Annual Symposium on Foundations of Computer Science (FOCS 2025).

[7] Ronen Shaltiel and Jad Silbak, *Explicit uniquely decodable codes for space bounded channels that achieve list-decoding capacity*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (2021), 1516–1526.

[8] R. Shaltiel and J. Silbak, *Error correcting codes that achieve bsc capacity against channels that are poly-size circuits*, Proceedings of the 63th Annual Symposium on Foundations of Computer Science (FOCS) (2022).

[9] Ronen Shaltiel and Jad Silbak, *Explicit codes for poly-size circuits and functions that are hard to sample on low entropy distributions*, Proceedings of the 56th Annual ACM Symposium on Theory of Computing (2024), 2028–2038.

Codes' Redundancy from The Lovász Local Lemma

Yonatan Yehezkeally

The probabilistic method has no shortage of applications in combinatorics (for an excellent overview and many examples, see [1]). Even rudimentary 1st-moment calculation methods often yield profound results, but at times these may not be sufficient. Unsurprisingly, the method finds applications in coding theoretic problems as well; specifically, we observed in this talk several examples where it can be useful for attaining the redundancy of a particular family of codes.

Our focus was codes in some arbitrary finite space Ω, that may naturally be described in the fashion $\mathcal{C} = \Omega \setminus \bigcup_{\alpha\in I} A_\alpha$, for a given index-set I and a collection $\{A_\alpha : \alpha \in I\}$. An application of the probabilistic method for redundancy calculation follows then from equipping Ω with the uniform probability measure, resulting in $\mathrm{red}(\mathcal{C}) \triangleq \log(|\Omega|) - \log(|\mathcal{C}|) = -\log \Pr(\mathcal{C})$. (We tacitly ignore the logarithm base in what follows.)

In these settings, the simplest scenario would be for mutually independent collection $\{A_\alpha : \alpha \in I\}$, where one readily has $\mathrm{red}(\mathcal{C}) = -\sum_{\alpha\in I} \log \Pr(\Omega \setminus A_\alpha)$. However, in many interesting scenarios there cannot easily be found a collection of events satisfying this requirement for $\mathcal{C}$. On the other hand, one can always apply the union bound to derive $\mathrm{red}(\mathcal{C}) = \log\bigl(1 - \sum_{\alpha\in I} \Pr(A_\alpha)\bigr)$, but this solution flat-out fails in the (fairly common) case that $\sum_{\alpha\in I} \Pr(A_\alpha) > 1$.

We note that there nevertheless exist a qualitative difference between these two approaches. In the case that for all $\alpha \in I$ there exists $\Gamma_\alpha \subseteq I$ satisfying that A_α is mutually independent from $\{A_\beta : \beta \notin \Gamma_\alpha\}$, and if $|\Gamma_\alpha| << |I|$ for all (or 'most') $\alpha \in I$ (a precise notion could be $|\Gamma_\alpha| = o(|I|)$ in some asymptotic regime), one could perhaps be justified in expecting $\mathrm{red}(\mathcal{C})$ to be better approximated by the former expression, rather than the latter.

Indeed, the Lovász local lemma [2] gives precise meaning to the last statement. In its general (so-called *asymmetric*) formulation it states (relying on notations above)

Theorem 1 ([3]). *If there exists constants* $\{f_\alpha : \alpha \in I\} \subseteq (0,1)$ *satisfying for all* $\alpha \in I$ *that* $\Pr(A_\alpha) \le f_\alpha \prod_{\beta\in\Gamma_\alpha}(1 - f_\beta)$*, then* $\Pr(\mathcal{C}) \ge \prod_{\alpha\in I}(1 - f_\alpha)$*.*

Frequently, at least some sets in the collection $\{A_\alpha : \alpha \in I\}$ resemble one another, or are of 'the same *type*'. It is therefore quite useful to use the following well-known *symmetric* version of the local lemma, derived simply by assigning $f_\alpha \equiv \frac{1}{d+1}$:

Corollary 2. *If there exist constants* $p, d > 0$ *such that* $\Pr(A_\alpha) \le p$ *and* $|\Gamma_\alpha| \le d$ *for all* $\alpha \in I$*, and if* $ep(d+1) \le 1$*, then* $\Pr(\mathcal{C}) \ge \bigl(1 + \frac{1}{d}\bigr)^{-|I|}$*.*

Since in some applications one may find it more useful to derive an expression in terms of p rather than d, the following can similarly be derived using the althernative assignment $f_\alpha \equiv \frac{ep}{1+ep}$.

Corollary 3. *If there exist constants* $p, d > 0$ *such that* $\Pr(A_\alpha) \le p$ *and* $|\Gamma_\alpha| \le d$ *for all* $\alpha \in I$*, and if* $ep(d+1) \le 1$*, then* $\Pr(\mathcal{C}) \ge (1 + ep)^{-|I|}$*.*

In the talk, we saw that Theorem 3 can directly be applied to gain the redundancy of *Repeat-free Sequences* [4], a constrained system generalising de Bruijn sequences shown to be efficient and redundancy-optimal for data reconstruction from shot-gun sequencing, a technology used to read transcribe DNA molecules (with applications for DNA-based data storage schemata). It is also applied to derive the redundancy of approximate-hairpin-avoiding strings (a constraint useful for preventing the formation of secondary structures in DNA oligos) in [5]. In yet another example, it can be used to tightly bound the redundancy of Run-length-limited constrained sequences [6], in regimes applicable to novel storage media where the constraint length scales with block-length (see, e.g., [7, 8, 9]). Critically, it can be seen that the results produced by applying the local lemma in all these examples are asymptotically equivalent to those that could have been derived, had the collection $\{A_\alpha : \alpha \in I\}$ been mutually independent, answering the question posed above.

Next, we noted that the usefulness of the symmetric formulation of the local lemma is somewhat limited, as in some foreseeable situations one cannot tightly approximate $\Pr(A_\alpha)$ and $|\Gamma_\alpha|$ by the same constants, for all $\alpha \in I$. Nevertheless, a slightly more fine-grained approach can be taken, by partitioning these sets into a finite (small) number of types. Denoting $I = I_1 \cup I_2 \cup \cdots \cup I_\ell$ and for all $\alpha \in I$, $\Gamma_\alpha = \Gamma_{\alpha,1} \cup \Gamma_{\alpha,2} \cup \cdots \cup \Gamma_{\alpha,\ell}$ where $\Gamma_{\alpha,s} \subseteq I_s$ for all $s = 1, \dots, \ell$, and assume that constants p_s and $d_{s,u}$ exist such that for all $\alpha \in I_s$ and $u = 1, \dots, \ell$ it holds that $\Pr(A_\alpha) \le p_s$ and $|\Gamma_{\alpha,u}| \le d_{s,u}$. Then, one can derive the following two analogues:

Corollary 4 (Semi-symmetric case). *If* $\max\Big\{p_s + \sum_{u=1}^{\ell} d_{s,u} p_u : 1 \le s \le \ell\Big\} \le 1/e$, *then* $\Pr(\mathcal{C}) \ge \prod_{s=1}^{\ell} \big(1 + e p_s\big)^{-|I_s|}$.

Proof. By applying the assignment $f_\alpha \triangleq \frac{e p_s}{e p_s + 1}$ where $\alpha \in I_s$. □

Corollary 5 (Case 2). *If* $\max\Big\{p_s(1 + \sum_{u=1}^{\ell} d_{s,u}) : 1 \le s \le \ell\Big\} \le 1/e$, *then* $\Pr(\mathcal{C}) \ge \prod_{s=1}^{\ell} \big(1 + \frac{1}{\sum_{u=1}^{\ell} d_{s,u}}\big)^{-|I_s|}$.

Proof. By applying the assignment $f_\alpha \triangleq \frac{1}{1 + \sum_{u=1}^{\ell} d_{s,u}}$ where $\alpha \in I_s$. □

Through another example, we saw how Theorem 4 was used in [5] to derive the redundancy of *Resilient-repeat-free sequences*, utilised for the same purpose as repeat-free sequences under relaxed assumptions, namely allowing the presence of errors prior to sequencing.

In summary, the talk demonstrated how the celebrated result of the Lovás local lemma can still be used to derive meaningful results in contemporary literature.

References

[1] N. Alon and J. H. Spencer, *The Probabilistic Method (4th edition)*, John Wiley & Sons Ltd (2016).

[2] P. Erdős and L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, 10. Infinite and finite sets, ser. Colloquia Mathematica Societatis **2** (1975), 609–627.

[3] J. Spencer, *Asymptotic lower bounds for Ramsey functions*, Discrete Math. **20** (1977), 69–76.

[4] O. Elishco, R. Gabrys, M. Médard, and E. Yaakobi, *Repeat-free codes*, IEEE Trans. Inf. Theory **67(9)** (2021), 5749–5764.

[5] Y. Yehezkeally and N. Polyanskii, *On codes for the noisy substring channel*, IEEE Trans. Mol., Bio. and Multi-Scale Commun. **10(2)** (2024), 368–381.

[6] S. X. Wang and A. M. Taratorin, *Magnetic Information Storage Technology*, Academic Press (1999).

[7] M. Levy and E. Yaakobi, *Mutually uncorrelated codes for DNA storage*, IEEE Trans. Inf. Theory **65(6)** (2019), 3671–3691.
[8] D. Bar-Lev, S. Marcovich, E. Yaakobi, and Y. Yehezkeally, *Adversarial torn-paper codes*, IEEE Trans. Inf. Theory **69(10)** (2023), 6414–6427.
[9] F. Walter and Y. Yehezkeally, *Coding for strand breaks in composite dna*, Proc. IEEE Int. Symp. Inf. Theory (2025).

An Information Theory of Query Optimization

Hsin-Po Wang
(joint work with Yu-Ting Lin)

In a large database, joining multiple relations is basic yet costly to execute. A query optimizer wants to know if one execution plan, say $(R \wedge S) \wedge T$, is better than the other, say $R \wedge (S \wedge T)$. Therefore, estimating the cardinalities of intermediate relations helps planning [7, 6, 5]. This work seeks upper bounds on the join cardinality that are tight and cheap. This is called *pessimistic cardinality estimation* in literature.

Let us consider an example: Let A be the set of sellers, B be the set of merchandises, and C is the set of buyers. Between any pair of attributes there can be relations. For instance, say $R(A, B)$ records if a seller sells a merchandise; it is a table whose rows are pairs $(a, b) \in A \times B$ where a sells b. Between B and C there can a relation $S(B, C)$ that records if a buyer wants a merchandise. And between C and A can be a relation $T(C, A)$ that records if a buyer lives near a seller. Now, say we are interested in the join query $R(A, B) \wedge S(B, C) \wedge T(C, A)$. This is a table whose rows are triples (a, b, c) where a sells b that c wants and they live near each other. The challenge here is to estimate the number of rows

$$\#_\Delta := |R(A,B) \wedge S(B,C) \wedge T(C,A)| = \left| \left\{ (a,b,c) : \substack{R(a,b) \text{ and} \\ S(b,c) \text{ and} \\ T(c,a)} \right\} \right|$$

before actually forming the table.

1. The Old Dexterous Framework

The starting point is that the number of triangles is at most the number of free triples (which is not tight)

$$\#_\Delta \le |A| \cdot |B| \cdot |C|. \tag{1}$$

Second, one notices that a triangle is determined by two edges. Hence,

$$\#_\Delta \le |R(A,B)||S(B,C)| \ , \ |S(B,C)||T(C,A)| \ , \ |T(C,A)||R(A,B)| \ . \tag{2}$$

Atserias, Grohe, and Marx [1] generalized (2). The new estimated bound, called AGM bounds, reads

$$\#_\Delta \le |R(A,B)|^u \cdot |S(B,C)|^v \cdot |T(C,A)|^w \tag{3}$$

where the weights (say u and v) associated to an attribute (say B) sum to 1 or more.

The next breakthrough is *chain bounds* [2] and *PANDA bounds* [3]. They harvest the fact that the two edges must share the same vertex, and hence choosing the second edge is constrained by the degree of the shared vertex:

$$\#_\Delta \le |R(A,B)| \cdot \max_{a\in A} \deg_T(a)\ ,\ |R(A,B)| \cdot \max_{b\in B} \deg_S(b)\ ,\ |S(B,C)| \cdot \max_{b\in B} \deg_R(b)\ ,$$

$$|S(B,C)| \cdot \max_{c\in C} \deg_T(c)\ ,\ |T(C,A)| \cdot \max_{c\in C} \deg_S(c)\ ,\ |T(C,A)| \cdot \max_{a\in A} \deg_R(a)\ . \tag{4}$$

Here, $\deg_R(a)$ is the number of $b \in B$ that satisfy $R(a,b)$, and other degrees are defined similarly.

Finally, Abo Khamis, Nakos, Olteanu, and Suciu [4] made huge progress by generalizing (1)–(4) to an infinite family of bounds. Two new instances include

$$\#_\Delta^3 \le \sum_{a\in A} \deg_R(a)^2 \cdot \sum_{b\in B} \deg_S(b)^2 \cdot \sum_{c\in C} \deg_T(c)^2, \tag{5}$$

$$\#_\Delta^6 \le \sum_{a\in A} \deg_R(a)^3 \cdot \sum_{c\in C} \deg_S(c)^3 \cdot |T(C,A)|^5 \tag{6}$$

Note that the right-hand sides of (1)–(6) all have one thing in common: $|A| = \sum_{a\in A} 1$ is the 0-norm, $|R(A,B)| = \sum_{a\in A} \deg_R(a)$ is the 1-norm, $\sum_{a\in A} \deg_R(a)^2$ and $\sum_{a\in A} \deg_R(a)^3$ are the 2-norm squared and the 3-norm cubed, and finally $\max_{a\in A} \deg_R(a)$ is the ∞-norm of the *degree sequence* $\{\deg_R(a)\}_{a\in A}$. Abo Khamis et al. [4] unified all these bounds using a single building block.

The meta reason degree sequences are useful here is their relation to join cardinalities via entropy inequalities. For instance, for any random pair $(X,Y) \in R(A,B)$,

$$H(X) \le \ln |A|, \qquad H(X,Y) \le \ln |R(A,B)|, \qquad H(Y|X) \le \ln \max_{a\in A} \deg_R(a). \tag{7}$$

Abo Khamis et al. [4] unified them as

$$H(X) + pH(Y|X) \le \ln \sum_{a\in A} \deg_R(a)^p \tag{8}$$

for any $p \ge 0$, making (7) special cases at $p = 0$, 1, and ∞, respectively.

2. Our New Ambidextrous Bound

The main contribution of our work is to introduce *bivariate moments*

$$_p\big\|R(A,B)\big\|_q \coloneqq \sum_{(a,b)\in R} \deg_R(a)^{p-1} \deg_R(b)^{q-1}$$

of a *bi-degree sequence* $\{(\deg_R(a), \deg_R(b))\}_{(a,b)\in R(A,B)}$. This generalizes (7)–(8) to

$$pH(Y|X) + I(X;Y) + qH(X|Y) \le \ln {_p\big\|R(A,B)\big\|_q} \tag{9}$$

for all $p, q \ge 1$, making (8) a special case at $q = 1$. This provides new building blocks to bound $\#_\Delta$, for instance

$$\#_\Delta^3 \le {_{4/3}\big\|R(A,B)\big\|_{5/3}} \cdot {_{4/3}\big\|S(B,C)\big\|_{5/3}} \cdot {_{4/3}\big\|T(C,A)\big\|_{5/3}}. \tag{10}$$

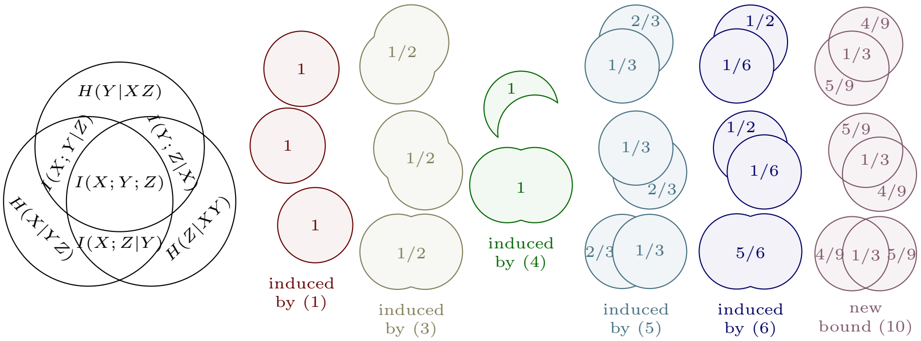

Figure 1. Bounding $\#_\Delta$ can be viewed as a fractional covering problem on the entropy Venn diagram. The building blocks are ○, ⬭, ☾, and ⚭, which correspond to (7), and (8), respectively. Our contribution can be viewed as inventing a new building block ⚭, which corresponds to (9).

We call (8) *dexterous* bounds for them counting claws "$\in$". We call the new bounds (9) *ambidextrous* for them counting claw pairs "$\ni\!\!-\!\!\in$". Ambidextrous upper bounds are provably tighter (or at least not looser) due to Hölder's inequality

$$ {}_{p}\|R\|_1^{w} \cdot {}_{1}\|R\|_q^{1-w} \geq {}_{wp+(1-w)}\|R\|_{w+(1-w)q}. \tag{11}$$

As a bonus contribution, we show that finding the best bound, dexterous or ambidextrous, is a convex optimization problem. This means that we can take advantages of existing black-box algorithms.

References

[1] Martin Grohe and Dániel Marx, *Constraint Solving via Fractional Edge Covers*, ACM Transactions on Algorithms **11(1)** (2014), 1–20.

[2] Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu, *Computing Join Queries with Functional Dependencies*, Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (2016), 327–342.

[3] Mahmoud Abo Khamis, Hung Q. Ngo, and Dan Suciu, *What Do Shannon-type Inequalities, Submodular Width, and Disjunctive Datalog Have to Do with One Another?* Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (2017), 429–444.

[4] Mahmoud Abo Khamis, Vasileios Nakos, Dan Olteanu, and Dan Suciu, *Join Size Bounds using l_p-Norms on Degree Sequences*, Proceedings of the ACM on Management of Data **2(2)** (2024), 1–24.

[5] Jan Kossmann, Thorsten Papenbrock, and Felix Naumann, *Data dependencies for query optimization: A survey*, The VLDB Journal **31(1)** (2022), 1–22.

[6] Hai Lan, Zhifeng Bao, and Yuwei Peng, *A Survey on Advancing the DBMS Query Optimizer: Cardinality Estimation, Cost Model, and Plan Enumeration*, Data Science and Engineering **6(1)**(2021), 86–101.

[7] Yeonsu Park, Seongyun Ko, Sourav S. Bhowmick, Kyoungmin Kim, Kijae Hong, and Wook-Shin Han, *G-CARE: A Framework for Performance Benchmarking of Cardinality Estimation Techniques for Subgraph Matching*, Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (2020), 1099–1114.

Decreasing Norm-Trace Codes

Hiram Lopez
(joint work with Cícero Carvalho and Gretchen L. Matthews)

We studied the decreasing norm-trace codes, which depend on the evaluation of certain polynomials on the affine points of the norm-trace curve. We see how the Gröbner basis of the vanishing ideal of the affine points and their indicator functions help us to find the dual of the code. We show that the trace function helps recover an erased entry of a codeword using partial information from the rest of the entries. This recovery property has applications for distributed storage systems. This talk is based on the works [1, 2, 3, 4, 5].

References

[1] M. Bras-Amorós, M.E. O'Sullivan, *Extended norm-trace codes with optimized correction capability*, Applied Algebra, Algebraic Algorithms and Error-Correcting Codes (2007), 337–346.

[2] E. Camps, H. H. López, G.L. Matthews, E. Sarmiento, *Polar decreasing monomial-cartesian codes*, IEEE Trans. Inf. Theory **67(6)** (2021), 3664–3674.

[3] J.L.D. Cox, D. O'Shea, *Ideals, Varieties, and Algorithms*, Undergraduate Texts in Mathematics (2008).

[4] O. Geil, T. Høholdt, *Footprints or generalized Bezout's theorem*, IEEE Trans. Inf. Theory **46(2)** (2000), 635–641.

[5] H.H. López, I. Soprunov, R.H. Villarreal, *The dual of an evaluation code*, Des. Codes Cryptogr. **89(7)** (2021), 1367–1403.

Reporter: Hedongliang Liu

Participants

Anisha Banerjee
Lehr- und Forschungseinheit für
Nachrichtentechnik
TU München
Theresienstraße 90
München 80333
GERMANY

Prof. Dr. Alexander Barg
Department of Electrical and
Computer Engineering
University of Maryland
College Park MD 20742
UNITED STATES

Prof. Dr. Martin Bossert
Institut für Nachrichtentechnik
Universität Ulm
Albert-Einstein-Allee 43
89081 Ulm
GERMANY

Kuan Cheng
School of Computer Science
Peking University
Beijing 100871
CHINA

Prof. Dr. Mahdi Cheraghchi
University of Michigan
Department of EECS
2260 Hayward St., Room 3603
Ann Arbor, MI 48109-1043
UNITED STATES

Dr. Gil Cohen
Department of Computer Science
Tel Aviv University
6939539 Tel Aviv
ISRAEL

Dr. Roni Con
Department of Computer Science
Technion - Israel Institute of
Technology
Haifa 32000
ISRAEL

Yotam Dikstein
School of Mathematics
IAS
Einstein Drive 1
Princeton 08540
UNITED STATES

Dean Doron
Computer Science Department
Ben Gurion University of the Negev
PO BOX 653
84105 Beer-Sheva
ISRAEL

Klim Efremenko
Computer Science, Ben-Gurion
University of the Negev
Beer-Sheva 69978
ISRAEL

Prof. Dr. Ran Gelles
Faculty of Engineering
Bar-Ilan University
52 900 Ramat-Gan
ISRAEL

Louis Golowich
Department of Computer Science
University of California, Berkeley
Berkeley CA 94720-3840
UNITED STATES

Assist. Prof. Dr. Fernando Granha Jeronimo
Thomas M. Siebel Center
for Computer Science
University of Illinois at Urbana
201 N. Goodwin Avenue, MC-258
Urbana IL 61801
UNITED STATES

Zeyu Guo
Department of Computer Science and
Engineering
The Ohio State University
2015 Neil Ave
Columbus OH 43210-1174
UNITED STATES

Prof. Dr. Venkatesan Guruswami
Departments of EECS and Mathematics
Simons Institute for the Theory of
Computing
University of California, Berkeley
625 Soda Hall
Berkeley CA 94720-1770
UNITED STATES

Prof. Dr. Prahladh Harsha
Tata Institute of Fundamental Research
Homi Bhabha Road, Colaba
400 005 Mumbai
INDIA

Prof. Dr. Sihuang Hu
School of Cyber Science and Technology
Shandong University
72 Binhai Road
266237 Jinan, Shandong
CHINA

Gillat Kol
Princeton University
Princeton 08540
UNITED STATES

Prof. Dr. Swastik Kopparty
Department of Mathematics
University of Toronto
40 St. George Street
Toronto ON M5S 2E4
CANADA

apl. Prof. Dr. Sascha Kurz
Mathematisches Institut
Universität Bayreuth
Postfach 101251
95447 Bayreuth
GERMANY

Ray Li
Dept. of Mathematics and Computer
Science, Santa Clara University
Santa Clara University
500 El Camino Real
Santa Clara 95050
UNITED STATES

Dr. Xin Li
Department of Computer Science
Johns Hopkins University
Baltimore, MD 21218-2689
UNITED STATES

Dr.-Ing Hedongliang Liu
Karlsruhe Institute for Technology (KIT)
76133 Karlsruhe
GERMANY

Dr. Siqi Liu
Department of Computer Science
Duke University
Durham NC 27708
UNITED STATES

Dr. Hiram H. López
460 McBryde Hall (550)
Department of Mathematics
Virginia Polytechnic Institute and
State University
225 Stanger St
Blacksburg 24061
UNITED STATES

Peter Manohar
School of Mathematics
The Institute for Advanced Study
Simonyi Hall
Princeton, NJ 08540
UNITED STATES

Dr. Gretchen Matthews
Department of Mathematics
Virginia Tech
Blacksburg VA 24061
UNITED STATES

Jonathan Mosheiff
Dept. of Computer Science
Ben Gurion University of the Negev
84105 Beer-Sheva
ISRAEL

Aaron Putterman
Science and Engineering Complex
Department of Computer Science
Harvard University
Science and Engineering Complex
150 Western Ave
Boston 02134
UNITED STATES

Dr. Nicolas Resch
Informatics Institute
University of Amsterdam
Science Park 900
Amsterdam 1098 XH
NETHERLANDS

Prof. Dr. Noga Ron-Zewi
Department of Computer Science
University of Haifa
Haifa 32000
ISRAEL

Dr. Shubhangi Saraf
Department of Mathematics
University of Toronto
40 St. George Street
Toronto ON M5S 2E4
CANADA

Dr. Hugo Sauerbier Couvée
Associate Professorship of Coding
and Cryptography
Technical University of Munich (TUM)
Theresienstraße 90
80333 München
GERMANY

Alexander Schmidhuber
Center for Theoretical Physics
Massachusetts Institute of
Technology
77 Massachusetts Avenue
Cambridge, MA 02139-4307
UNITED STATES

Dr. Ronen Shaltiel
Dept. of Computer Sciences
University of Haifa
Mount Carmel
Haifa 31905
ISRAEL

Dr. Chong Shangguan
Research Center for Mathematics and
Interdisciplinary Sciences, Shandong
University, University Campus Qingdao,
Room E208, Huagang Bldg.
72 Binhai Road
Qingdao 266237, Shangdong
CHINA

Prof. Dr. Amir Shpilka
Department of Computer Science
Tel Aviv University
Tel Aviv 69978
ISRAEL

Dr. Jad Silbak
Khoury College of Computer Sciences
Northeastern University
Boston MA 02115-5000
UNITED STATES

Shashank Srivastava
Center for Discrete Math and
Theoretical Computer Science
Rutgers University
96 Frelinghuysen Road Piscataway
New Brunswick NJ 08854-8019
UNITED STATES

Prof. Dr. Madhu Sudan
John A. Paulson School of Engineering
and Applied Sciences
Harvard University
150 Western Avenue, SEC 3.434
Boston MA 02134
UNITED STATES

Prof. Dr. Itzhak Tamo
Department Electrical-Engineering -
Systems
Tel Aviv University
Ramat Aviv, Tel Aviv 69978
ISRAEL

Prof. Dr. Amnon Ta-Shma
Department of Computer Science
Tel Aviv University
Tel Aviv 69978
ISRAEL

Dr. Madhur Tulsiani
Toyota Technological Institute
at Chicago
6045 S Kenwood Av
Chicago, IL 60637
UNITED STATES

Prof. Dr. Hsin-Po Wang
Department of Electrical Engineering
National Taiwan University
No. 1, Sec. 4, Roosevelt Rd.
Taipei 106
TAIWAN

Prof. Dr. Wolfgang Willems
Fakultät für Mathematik
Otto-von-Guericke-Universität
Magdeburg
Universitätsplatz 2
39106 Magdeburg
GERMANY

Dr. Mary Wootters
Departments of Computer Science and
Electrical Engineering
Stanford University
1455 California Ave
94304 Palo Alto
UNITED STATES

Tal Yankovitz
Department of Mathematics
School of Mathematical Sciences
Tel Aviv University
Ramat Aviv, Tel Aviv 69978
ISRAEL

Dr. Yonatan Yehezkeally
School of Computing
Newcastle University UK
1 Science Square
Newcastle upon Tyne NE4 5TG
UNITED KINGDOM

Prof. Dr. Gilles Zemor
Institut de Mathématiques de Bordeaux
(IMB)
Université de Bordeaux
351 Cours de la Liberation
33400 Talence
FRANCE

Rachel Zhang
Computer Science & Artificial
Intelligence Laboratory
MIT
Cambridge, MA 02139
UNITED STATES

Zihan Zhang
Department of Computer Science and
Engineering
The Ohio State University
395 Dreese Laboratories
Columbus, OH 43210-1277
UNITED STATES