Final Presentation for Error-Correcting Codes: Between Algebraic and Combinatorial

Dr. Shashank Srivastava Mursalin Habib Arushi Srinivasan

University of Maryland, College Park

DIMACS

Institute for Advanced Study

DIMACS REU, July 16 2025

Table of Contents

Table of Contents

Models of Error-Correction

- motivation stems from complexity theory/cryptography
- We are given some knowledge about the channel
- Consider messages as elements of $S = \{0,1\}^n$ for now
- e.g. $\forall m \in S$, every bit of m is flipped independently with probability $p \in [0,1]$ – randomized model
- $\forall m \in S$, at most t bits can be flipped – adversarial model
- what if we can send messages in a way that can "withstand" corruption? This leads us to encryption
- We will use the adversarial model throughout this project

Table of Contents

Terminology

- alphabet Σ: code $C \subset \Sigma^n$ (n = **blocklength**)
  - Consider $C \subset \Sigma^n$, $|C| = M$: $C : [M] \to \Sigma^n$: injective mapping, **encoding function**
  - Let $C \subset \Sigma^n$, $D : \Sigma^n \to [|C|]$ is the **decoding function**
- $\forall f, g \in C$, $\Delta(f, g) = |\{x \in \{1, ..., n\} \mid f[x] \neq g[x]\}|$: **hamming distance**
  - $\delta(f, g) = \frac{\Delta(f,g)}{n}$, **relative hamming distance**
- **t-Error Channel** $Ch : \Sigma^n \to \Sigma^n$ satisfying $\Delta(v, Ch(v)) \leq t \ \forall v \in \Sigma^n$
- **Error Correcting Code** Given $C \subset \Sigma^n$ and $t \in Z^+$, C is a $t-$error correcting code if $\exists$ decoding function D s.t. $\forall m \in [M]$, $\forall t-$error channels Ch, we have $D(Ch(C(m))) = m$

Properties of Linear Codes

- $C \subset \Sigma^n$ is a linear code iff it is a *linear subspace* of Σ^n
- Equivalently, C is a space of vectors of F_q^n where q is a prime power
- a code is (ρ, L)-**list decodable**: $\forall y \in \Sigma^n$, $|\{c \in C \mid \Delta(y, c) \leq \rho n\}| \leq L$
- **rate** $R = \frac{k}{n}$
- **distance** $\delta = \min_{f,g \in C} \delta(f, g)$
- (R, δ) negatively correlated and by the Singleton bound, $\delta + R \leq 1 + \frac{1}{n}$

Reed-Solomon (RS) Codes

- $RS_{n,k}(\alpha_1, ..., \alpha_n) \stackrel{\text{def}}{=} \{(f(\alpha_1), ..., f(\alpha_n)) \in F_q^n : f \in F_q[X], \deg(f) < k\}$
- Choose each element of $\{\alpha_1, ..., \alpha_n\}$ from F_q (ideally all distinct)
- Code words are now bounded degree polynomials in $F_q[X]$ evaluated at $\{\alpha_1, ..., \alpha_n\}$: note that these evaluations lie in F_q
- Given a bounded degree poly. f, send $(f(\alpha_1), ..., f(\alpha_n))$ across the channel: at most t evaluations are changed
- want $k << n$: in order to decode the message, you want fewer repeated roots of a polynomial (hence $p_1 - p_2$ has roots at fewer evaluation points)
- use the above to solve a linear system of equations and recover the message w.h.p. (note: RS codes are linear)
- optimal for unique decoding

Concatenated Codes

- Given an outer-code (and a corresponding decoder) $C_{out} \subset \Sigma^n$ an inner-code $C_{in} \subset {\Sigma_0}^d$
- $\Phi : \Sigma \to {\Sigma_0}^d$, an injective mapping over Σ
- C_{concat}: given $s \in \Sigma^n$, apply Φ to every component of s
- when $|\Sigma_0|, d$ are constant, use brute force decoding (check if a string $s' \in {\Sigma_0}^d$ is within distance $\frac{\delta}{2}$ to recover corresponding c.w.) to decode each component of $s \in (\Sigma_0^d)^n$, use $\Phi^{-1} \to$ reduction to outer code decoding
- $\rho = \rho_{out} * \rho_{in}$, $\delta \geq \delta_{out} * \delta_{in}$

AEL (Alon, Edmonds, Luby '96) Codes

- Given $G = (L, R, d)$, $|L| = |R| = n$ d-regular bipartite (expander) graph, encode a labeling of the left vertices via some C_{out}
  - G is a d-regular λ (spectral) bipartite expander iff $\forall X \subseteq L,\ Y \subseteq R$, $|E(X, Y) - \frac{d}{n} * |X| * |Y|| \leq \lambda dn$
  - can explicitly construct such an expander for any $\lambda > 0$, pseudorandomness
  - Idea: d constant ($d = n \rightarrow \lambda = 0$) but can set λ arb. small
- C_{in}, $\Phi : \Sigma \rightarrow {\Sigma_0}^d$: Φ maps each left vertex to a labeling of its incident edges
- given a labeling of E, C_{AEL} sets the label of each right vertex to the tuple-labeling of its incident edges
- Want large rate and distance, by the Singleton bound $\rho + \delta \leq \frac{1}{blocklength} \rightarrow$ use the expander mixing lemma to get $\delta_{AEL} \geq \delta_{in} - \frac{\lambda}{\delta_{out}}$

List Decoding Utility

- decoding a problem: given a labeling in ${\Sigma_0}^{dn}$ of the right vertices and that in the original encoding of L, q fraction of pos'ns were corrupted, recover the original encoding
- can do this for $q \leq \frac{\delta_{out}}{2}$: AEL codes are optimal for unique decoding
- List decoding: given ρ, C and $y \in \Sigma^n$, o/p $\tilde{L} = \{c \in C \mid \delta(c, y) \leq \rho\}$
- works for higher error fractions

Table of Contents

Initial Approaches

- Alrabiah, Guruswami, Li 2024: generalize to (ρ, L) list decoding (unique decoding sets $L = 1$)
  - Can choose $\{\alpha_1, ..., \alpha_n\}$ arbitrarily: they select each one u.a.r. from F_q: collisions occur with low probability (randomly punctured RS codes)
  - some results:
    - Theorem 1.3: $\forall L \geq 1$, $\epsilon \in (0, 1)$, *a random linear code over alphabet size* $q \geq 2^{\frac{10L}{\epsilon}}$ *with n suff. lg. is w.h.p.* $(\frac{L}{L+1}(1 - R - \epsilon), L)$*-average-radius-list-decodable*
    - for any distinct $L + 1$ codewords $(c^1, ..., c^{L+1})$ and for any $y \in F_q^n$, $\frac{1}{L+1} \sum_{j=1 \to L+1} \Delta(y, c^j) \geq \frac{L}{L+1}(1 - R - \epsilon)n$
    - Corollary 1.4: $\forall \epsilon > 0$, *a random linear code over alphabet size* $q \geq 2^{O(\frac{1}{\epsilon^2})}$ *and n suff. lg. is w.h.p.* $(1 - R - \epsilon, O(\frac{1}{\epsilon}))$*-average-radius-list decodable*
  - initial problem: derandomizing the AGL code (proved to be too hard for the REU timespan)

"Progress"

- objective: what properties does the AGL code have that we can explicitly construct to get list decodability?
- Folding, interleaving have previously been useful $\rightarrow$ reconstructed BCDZ '25 proofs on equivalence of interleaved RS/RS over a subfield
- Proved that given a bijection from $F_{q^2} \rightarrow F_q^2$, linearity over F_q^2 leads to linearity over F_{q^2}
- Tried to generalize this to other linear codes, investigated expander-based codes as well
- The explicit constructions of expander graphs caused us to pivot to AEL codes

$|\tilde{L}|$ Bounds

- Known: AEL is optimal for u.d., what happens when you want to list-decode up to $\frac{k}{k+1}(1-R)$ error fraction as $k \to \infty$
- Srivastava thesis: when $k = 2$, $|\tilde{L}| = 2$, but by Plotkin-bound style arguments, these bounds quickly blow up when $k = 4$ or 5 and eventually become tower bounds
- Srivastava, Tulsiani (ST) '25: doubly exponential bounds via graph regularity partitions (solving the cut norm problem and taking the common refinement of a regularity partition, improving Alon-Naor's approach $\to$), but this cost is already paid at $k = 2$
  - randomized near-linear time
- Employed a proof by contradiction to show that the thesis bound of 2 for $\frac{2}{3}(1-R)$ errors is tight
- Question: Can we asymptotically get the ST bounds while improving the (local) bounds for smaller k?

Approaches

- S '24: proved that for decoding $\frac{k}{k+1}(1-R)$ errors for folded RS codes, $|\tilde{L} \cap \tilde{H}| \leq k$
  - combinatorial proof relying on pruning the search space to H: similar to our regularity partitions
  - $\tilde{H} = \{f_0 + \alpha f_1 \ | \alpha \in F_q\}$ where $f_0, f_1 \in C$, $S = \{h \in \tilde{H} | f_1 \neq 0\}$
- ST'25 relies on enumerating all cells in the common refinement $\rightarrow$ large cost for small k
- Avoid the enumeration step for larger k: proved that for $k = 2$, you pay a smaller cost by only enumerating over 4 cells as opposed to 16
- Generalize to larger k
- Introduced erasure decoding algorithm in ST'25 $\rightarrow$ more robust, investigate error vs. erasure tradeoff (correcting both)
- Conjecture: erasure and error decoding prune the search space

Table of Contents

Acknowledgements

- Dr. Shashank Srivastava (DIMACS, IAS)
- Mursalin Habib (DIMACS)
- DIMACS REU
  - Professor Lazaros Gallos (DIMACS)
  - Larry Frolov (Rutgers Math, DIMACS)
- NSF grant CCF-2447342

Table of Contents

Sources

- *Essential Coding Theory* by Guruswami, Rudra, Sudan
- Alrabiah, O., Guruswami, V., Li, R. (2024). Randomly punctured Reed–Solomon codes achieve list-decoding capacity over linear-sized fields. arXiv. https://doi.org/10.48550/arXiv.2304.09445
- Brakensiek, J., Chen, Y., Dhar, M., Zhang, Z. (2025). Unique decoding of Reed–Solomon and related codes for semi-adversarial errors (arXiv:2504.10399). arXiv. https://doi.org/10.48550/arXiv.2504.10399
- Srivastava, S., Tulsiani, M. (2025). List decoding expander-based codes up to capacity in near-linear time (arXiv:2504.20333). arXiv. https://doi.org/10.48550/arXiv.2504.20333
- Jeronimo, F.G., Mittal, T., Srivastava, S., Tulsiani, M. (2025). Explicit codes approaching generalized Singleton bound using expanders (arXiv:2502.07308). arXiv. https://doi.org/10.48550/arXiv.2502.07308

Sources pt. 2

- Srivastava, S. (2024). Continuous optimization for decoding errors (arXiv:2408.14652). arXiv. https://doi.org/10.48550/arXiv.2408.14652
- Srivastava, S. (2024, October 11). Improved List Size for Folded Reed–Solomon Codes (arXiv:2410.09031). arXiv. https://doi.org/10.48550/arXiv.2410.09031
- Vadhan, S.P. (2012). Pseudorandomness [Lecture notes].Harvard University. Foundations and Trends in Theoretical Computer Science, 7(1–3), 1–336. https://people.seas.harvard.edu/ salil/pseudorandomness/expanders.pdf