AG Codes Achieve List Decoding Capacity over Constant-Sized Fields

AG Codes Achieve List Decoding Capacity over Constant-Sized Fields*

Joshua Brakensiek
jbrakens@cs.stanford.edu
Independent Researcher
USA

Manik Dhar
dmanik@mit.edu
Massachusetts Institute of Technology
Cambridge, Massachusetts, USA

Sivakanth Gopi
sigopi@microsoft.com
Microsoft Research
Redmond, Washington, USA

Zihan Zhang
zhang.13691@osu.edu
The Ohio State University
Columbus, Ohio, USA

ABSTRACT

The recently-emerging field of higher order MDS codes has sought to unify a number of concepts in coding theory. Such areas captured by higher order MDS codes include maximally recoverable (MR) tensor codes, codes with optimal list-decoding guarantees, and codes with constrained generator matrices (as in the GM-MDS theorem).

By proving these equivalences, Brakensiek-Gopi-Makam showed the existence of optimally list-decodable Reed-Solomon codes over exponential sized fields. Building on this, recent breakthroughs by Guo-Zhang and Alrabiah-Guruswami-Li have shown that randomly punctured Reed-Solomon codes achieve list-decoding capacity (which is a relaxation of optimal list-decodability) over linear size fields. We extend these works by developing a formal theory of relaxed higher order MDS codes. In particular, we show that there are two inequivalent relaxations which we call lower and upper relaxations. The lower relaxation is equivalent to relaxed optimal list-decodable codes and the upper relaxation is equivalent to relaxed MR tensor codes with a single parity check per column.

We then generalize the techniques of Guo-Zhang and Alrabiah-Guruswami-Li to show that both these relaxations can be constructed over constant size fields by randomly puncturing suitable algebraic-geometric codes. For this, we crucially use the generalized GM-MDS theorem for polynomial codes recently proved by Brakensiek-Dhar-Gopi. We obtain the following corollaries from our main result:

Randomly punctured algebraic-geometric codes of rate R are list-decodable up to radius $\frac{L}{L+1}(1-R-\epsilon)$ with list size L over fields of size $\exp(O(L/\epsilon))$. In particular, they achieve list-decoding capacity with list size $O(1/\epsilon)$ and field size $\exp(O(1/\epsilon^2))$. Prior to this work, AG codes were not even known to achieve list-decoding capacity.

By randomly puncturing algebraic-geometric codes, we can construct relaxed MR tensor codes with a single parity check per column over constant-sized fields, whereas (non-relaxed) MR tensor codes require exponential field size.

CCS CONCEPTS

• **Mathematics of computing** → **Coding theory**; • **Theory of computation** → **Error-correcting codes**.

KEYWORDS

coding theory, MDS codes, list-decoding, Reed-Solomon codes, Algebraic Geometry codes

ACM Reference Format:
Joshua Brakensiek, Manik Dhar, Sivakanth Gopi, and Zihan Zhang. 2024. AG Codes Achieve List Decoding Capacity over Constant-Sized Fields. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3618260.3649651

1 INTRODUCTION

MDS (maximum distance separable) codes meet the Singleton bound [35], which is the optimal rate-distance tradeoff for codes over large alphabet. It states that an (n,k)-code has distance $d \le n-k+1$. Reed-Solomon codes [31] are a simple and explicit construction of linear MDS codes over fields of linear size (i.e., $q = O(n)$). Due to their optimality, MDS codes are extensively used in practice such as in communication, data storage, cryptography etc.. An (n,k)-code is MDS iff any k columns of its generator matrix are linearly independent.[1] Higher order MDS codes were recently introduced by Brakensiek-Gopi-Makam ([6]) as a natural generalization of MDS codes.

Definition 1.1. *An (n,k)-code with generator matrix $V_{k\times n}$ is order-ℓ higher order MDS, denoted by* $\mathrm{MDS}(\ell)$*, if for every ℓ subsets $A_1, A_2, \dots, A_\ell \subset [n]$, we have* $\dim(V_{A_1} \cap V_{A_2} \cap \cdots \cap V_{A_\ell}) = \dim(W_{A_1} \cap W_{A_2} \cap \cdots \cap W_{A_\ell})$ *where $W_{k\times n}$ is a generic $k \times n$ matrix.*

Here V_A is the span of the columns indexed by A. In other words, ℓ subspaces spanned by subsets of columns should intersect as minimally as possible. When $\ell \le 2$, MDS(1) and MDS(2) codes are equivalent to MDS codes ([6]). Higher order MDS codes have

*Full version: [5].

[1] In this paper, we will only focus on linear codes. Unless explicitly mentioned, all codes are linear.

STOC '24, June 24–28, 2024, Vancouver, BC, Canada

ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649651

since been shown to be equivalent to many concepts independently studied in coding theory ([7]). These include maximally recoverable (MR) tensor codes, codes with optimal list-decoding guarantees, and codes with constrained generator matrices (as in the GM-MDS theorem). We refer the reader to [7] for a detailed survey of these connections. [7] showed that random Reed-Solomon codes over exponentially large fields are higher order MDS codes with high probability. A major challenge is to give (explicit) constructions of higher order MDS codes over small fields. But in a recent work, [4] showed that even MDS(3) codes of constant rate require exponential field size, in sharp contrast to MDS codes (or MDS(2) codes) for which we have explicit constructions over linear size fields (for example Reed-Solomon codes). Therefore, it is natural to look for some kind of relaxation of higher order MDS codes which would allow for constructions over small fields. In this work, we introduce two ways to relax higher order MDS codes called *lower* and *upper* relaxation (these are defined in Section 3). We then establish equivalences between these relaxations and suitable relaxations of optimal list-decodable codes and MR tensor codes. In particular, the lower relaxation is equivalent to relaxed version of optimal list-decodable codes and the upper relaxation is equivalent to the relaxation of MR tensor codes with a single parity check per column. Finally we show that one can indeed construct such (upper and lower) relaxed higher order MDS codes over small fields. We start by briefly defining some of these concepts and discuss prior work along the way.

Optimal List-decodable Codes (LD-MDS($\leq L$)). List-decoding is an important concept in coding theory which allows for correction beyond half the minimum distance [13, 39]. An (n, k) code C is (ρ, L)-list decodable if any Hamming ball of radius ρn contains at most L codewords. The generalized Singleton bound due to [16, 32, 34] generalizes the well-known Singleton bound to the setting of list-decoding. The generalized Singleton bound states that for every (n, k) code C that is (ρ, L)-list decodable, we have

$$\rho \leq \frac{L}{L+1}(1 - k/n). \tag{1}$$

Note that when $L = 1$, we recover the Singleton bound. The same bound also holds for average-radius list-decoding which is a strengthening of list-decoding. An (n, k)-code C is (ρ, L)-average-radius list-decodable if for any $y \in \mathbb{F}^n$ and any $L + 1$ codewords $c_0, c_1, \ldots, c_L \in C$, we have that $\frac{1}{L+1}\sum_{i=0}^{L} \mathrm{wt}(y - c_i) \leq \rho n$. We now define the notion of optimal list-decodable codes.

Definition 1.2. *We say that C is* LD-MDS(L) *if it is (ρ, L)-average-radius list-decodable with radius $\rho = \frac{L}{L+1}(1-k/n)$, matching the generalized Singleton bound (1).*

A code is LD-MDS($\leq L$) if it is LD-MDS(ℓ) for all list sizes $\ell \leq L$. We would like to have explicit constructions of LD-MDS($\leq L$) codes over small alphabets. [7] showed that random Reed-Solomon codes over exponentially large fields are LD-MDS($\leq L$), i.e., optimally list-decodable for all list sizes L, proving a conjecture of [34]. Meanwhile, as discussed before, [4] showed that even LD-MDS(≤ 2) codes of constant rate, i.e., optimally list-decodable with list size 2, require exponential field size. Thus it is natural to look at relaxations of optimal list-decodability if we want constructions over small fields. A natural relaxation is to ask that a rate R code is (ρ, L)-average-radius list-decodable for

$$\rho = \frac{L}{L+1}(1 - R - \epsilon) \tag{2}$$

for some $\epsilon > 0$. We call such codes *relaxed* LD-MDS *codes* (see Section 3 for a formal definition) and denote them by rLD-MDS. We show that these relaxed LD-MDS codes are equivalent to the lower relaxation of higher order MDS codes.

Guo-Zhang ([21]) were the first to observe that such a relaxation leads to much improved constructions. They showed that random Reed-Solomon codes over fields of size $O_{L,\epsilon}(n^2)$ are relaxed LD-MDS codes with ρ given by (2).[2] Alrabiah-Guruswami-Li ([2]) further improved the field size to $O_{L,\epsilon}(n)$ for random Reed-Solomon codes. They also showed that random linear codes achieve the same relaxed list-decoding radius in (2) with $\exp(O(L/\epsilon))$ field size. In subsequent work [1], they also showed a lower bound that any (not necessarily linear) code which achieves the bound in (2) must have field size at least $\exp(\Omega_{L,R}(1/\epsilon))$. For average-radius list-decoding, they obtain a better lower bound of $\exp(\Omega_R(1/\epsilon))$ independent of list size $L \geq 2$.

One can further relax the list-decodability requirement to get what is known as *list-decoding capacity achieving codes*. Note that as $L \to \infty$, the list-decoding radius $\rho \to 1 - R$ where $R = k/n$ is the rate of the code, i.e., any non-trivial list-decoding cannot be done beyond $\rho = 1 - R$. This is the called the *list-decoding capacity*. We say that a code family achieves list-decoding capacity if for every $\epsilon > 0$ and $R \in (0, 1)$, there exists a code C from this family of rate R which is (ρ, L)-list decodable with $\rho = 1-R-\epsilon$ and $L = O_\epsilon(1)$.[3] In addition, we would also want the alphabet size of the code to be as small as possible. Random codes of rate R are $(1 - R - \epsilon, O(1/\epsilon))$-list-decodable (with alphabet size $2^{O(1/\epsilon)}$, see [23]), this was recently shown to hold for random linear codes as well (with larger alphabet size $2^{O(1/\epsilon^2)}$) in [2]. There is a long line of work trying to construct explicit code families achieving list-decoding capacity. After an initial breakthrough by [30], [22] constructed the first known family of codes achieving list-decoding capacity called Folded Reed-Solomon codes. Though the initial list size and alphabet size were of the form $n^{O_\epsilon(1)}$, later works have reduced the alphabet size and list size to $\exp(poly(1/\epsilon))$ [12, 20, 24–26, 28]. See [20] for an explicit construction matching these bounds. They obtain this code by starting with an AG code over $\mathbb{F}_{q^m}$ evaluated only on points of $\mathbb{F}_q$. Then they take a subcode of this by restricting the messages to a BTT evasive subspace (where BTT stands for Block Triangular Toeplitz). Similar to folded Reed-Solomon codes, their code is not linear over $\mathbb{F}_{q^m}$ which is the alphabet–it is only linear over the base field $\mathbb{F}_q$. They also give an efficient list-decoding algorithm. [26] give a randomized construction of codes achieving list-decoding capacity with alphabet size $\exp(\tilde{O}(1/\epsilon^2))$ and a much better list size of $O(1/\epsilon)$. They obtain this by folding AG codes using automorphisms of the underlying function field. The message space is also restricted to a hierarchical subspace-evasive set (of which we don't have explicit constructions, so they give a psuedorandom

[2] Here random Reed-Solomon codes over $\mathbb{F}_q$ refers to choosing the n evaluation points at random from $\mathbb{F}_q$. Alternatively, one can think of it as the code of length n obtained by randomly puncturing Reed-Solomon code of length q evaluated at all points of $\mathbb{F}_q$.
[3] In some works, even getting $L = n^{O_\epsilon(1)}$ is considered enough.

construction) to obtain the list size of $O(1/\epsilon)$. In particular, their codes are non-linear because of this restriction. But they give an efficient list-decoding algorithm for their codes.

We show that by simple random puncturing of an AG code, we can achieve relaxed LD-MDS codes over constant size fields. In fact, the field size is also nearly optimal and matches the lower bound shown in [1]. This result can be thought of as a partial derandomization of random linear codes which achieve the same bounds.

Theorem 1.3. *(Informal - See Theorem 4.22 and Corollary 4.23) By randomly puncturing suitable AG codes, with high probability, we get a code of rate R which is (ρ, L)-list-decodable with $\rho = \frac{L}{L+1}(1-R-\epsilon)$ (as in (2)) over fields of size $\exp(O(L/\epsilon))$. The same result also holds for the stronger average-radius list-decoding.*

We summarize the prior results and our work in Table 1.

Maximally Recoverable Tensor Codes. An (m, n, a, b)-tensor code is a linear code formed by the tensor product of two codes, an $(n, n-b)$-code C_{row} (called the row code) and an $(m, m-a)$-code C_{col} (called the column code), i.e., $C = C_{col} \otimes C_{row}$. Equivalently, the codewords of C are $m \times n$ matrices whose rows belong C_{row} and columns belong to C_{col}. Such a code satisfies 'a' parity checks per column and 'b' parity checks per row. Tensor codes have good *locality*, which means that we can recover an erased symbol by reading a small number of remaining symbols (called a repair group). They also have good *availability* which means that there are two such disjoint repair groups, one along the row and one along the column. Thus tensor codes are well-suited in coding for distributed storage. For example, Facebook's (now Meta) f4 storage architecture uses an $(m = 3, n = 14, a = 1, b = 4)$ tensor code to store data.

An (m, n, a, b)-tensor code is called maximally recoverable (MR) if it can recover from any erasure pattern that is information theoretically possible to recover from (for that particular field characteristic). Thus MR tensor codes are optimal codes in terms of their ability to recover from erasure patterns. The notion of maximal recoverability is introduced by [9, 19] to design optimal codes for distributed storage. Because of their optimality, MR codes are being used in large scale distributed storage s such as in Microsoft's data centers [27]. MR tensor codes were first studied in the work of [17], with special emphasis on the case of $a = 1$. When $a = 1$, the column code is a simple parity check code. In this case, [17] give an explicit condition called *regularity* to check when an erasure pattern is correctable.[4] [6] showed that an $(m, n, a = 1, b)$-tensor code being MR is equivalent to the row code C_{row} being higher order MDS of order m, i.e., MDS(m). Thus MR tensor codes in the regime of $a = 1$ are exactly equivalent to higher order MDS codes.

In distributed storage applications, having small field size is extremely important as encoding and decoding involves finite field arithmetic. Therefore constructions of MR tensor codes over small fields is a very important problem. Unfortunately, recent lower bounds due to [4] imply that MR tensor codes with $a = 1$ and just three rows (i.e., $m = 3$) already require exponential field size. Thus it is again natural to look for relaxations of MR tensor codes and hope that we can construct them over smaller fields. In fact, there is a very natural relaxation for MR tensor codes. We say that an (m, n, a, b)-tensor code is (a', b')-*relaxed MR* if it can correct every erasure pattern that an (m, n, a', b')-MR tensor code can correct (here $a' \le a$ and $b' \le b$). In this work, since we are focusing only on MR tensor codes with $a = 1$, we will also only look at $(a' = 1, b')$-relaxed MR tensor codes. We show that such codes are equivalent to the upper relaxation of higher order MDS codes. We then show that one can indeed construct such codes over small fields by randomly puncturing Reed-Solomon codes or AG codes. Our main result is as follows:

Theorem 1.4. *(Informal) Let C_{col} be a simple $(m, m-1)$-parity check code where m is some fixed constant and let C_{row} be an $(n, n-b)$ code sampled as follows for the two different regimes.*

1. *Row code has constant rate $R \in (0, 1)$, i.e., $b = (1-R)n$: In this case, setting C_{row} to be a randomly punctured AG code over an alphabet of size $\exp(O(m/\epsilon))$ will make $C_{col} \otimes C_{row}$ into an $(1, (1-\epsilon)b)$-relaxed $(m, n, 1, b)$-MR tensor code with high probability. (Theorem 4.24)*
2. *Row code has small codimension, i.e., $b \ll n$: In this case, setting C_{row} to be a randomly punctured Reed-Solomon code over an alphabet of constant size $n^{O(m/\epsilon)}$ will make $C_{col} \otimes C_{row}$ into an $(1, (1-\epsilon)b)$-relaxed $(m, n, 1, b)$-MR tensor code with high probability. (Corollary 4.15)*

The lower bounds from [4] imply that in regime (1) of Theorem 1.3, (non-relaxed) MR tensor codes require fields of size $\exp(\Omega(n))$, whereas the theorem gives constant field size. While in regime (2) of Theorem 1.3, (non-relaxed) MR tensor codes require fields of size $n^{\Omega(b)}$, whereas the theorem gives polynomial field size independent of codimension b.

1.1 Technical Overview

We now give an overview of the main techniques needed to prove our main results. We start with motivating the relaxed notions of higher order MDS codes we use to abstract the important properties of our constructions. Then, we discuss how to adapt the state-of-the-art techniques for constructing list-decoding capacity-achieving codes to build more general techniques for constructing relaxed higher order MDS codes.

1.1.1 Relaxed Higher Order MDS Codes. As established in [7], there is an equivalence between higher order MDS codes, codes with optimal average-case list decoding guarantees, and maximally recoverable tensor codes (in the $a = 1$ regime). Thus, in a theory of *relaxed* higher order MDS codes, one would hope that natural relaxations of each of these quantities is also equivalent. However, such an equivalence is not possible (see discussion in Section 3.1.3). Instead, we define two relaxations of Definition 1.1, which we call the "lower" and "upper" relaxations of higher order MDS codes.

Consider a $k \times n$ matrix V. Informally, we say that V is the *d-dimensional lower relaxation* of an MDS(ℓ) code, which we denote by rMDS$_d(\ell)$, if it behaves like a $(n, k-d)$-MDS(ℓ) code. Likewise we say that V is a *d-dimensional upper relaxation* of an MDS(ℓ) code, which we denote by rMDS$^d(\ell)$, if it behaves like a $(n, k+d)$-MDS(ℓ) code. The precise definition of "behaves like" requires a bit of care.

[4] Giving such a condition for general a, b is still open.

Table 1: Summary of results on nearly optimal list-decodable codes. Here R is the rate and n is the length of the code.

Code Family	List Size	Radius (ρ)	Field Size	Construction	Reference
Random non-linear code	$O(1/\epsilon)$	$1-R-\epsilon$	$\exp(O(1/\epsilon))$	randomized	[13, 39]
Random linear code	L	$\frac{L}{L+1}(1-R-\epsilon)$	$\exp(O(L/\epsilon))$	randomized	[2]
Random Reed-Solomon	L	$\frac{L}{L+1}(1-R-\epsilon)$	$\exp(O(L/\epsilon))\cdot n$	randomized	[2]
Folded AG subcode + hierarchical evasive sets (non-linear code)	$O(1/\epsilon)$	$1-R-\epsilon$	$\exp(\tilde{O}(1/\epsilon^2))$	randomized	[26]
AG codes with subfield evaluation + BTT evasive subspace	$\exp(poly(1/\epsilon))$	$1-R-\epsilon$	$\exp(\tilde{O}(1/\epsilon^2))$	explicit	[20]
Random AG code	L	$\frac{L}{L+1}(1-R-\epsilon)$	$\exp(O(L/\epsilon))$	randomized	Theorem 4.22

As observed in [6, 37], computing the intersections of spaces is closely related to the following block matrix (see Proposition 2.1).

$$\mathcal{G}_{A_1,\dots,A_\ell}[V] := \begin{pmatrix} I_k & V|_{A_1} & & & \\ I_k & & V|_{A_2} & & \\ \vdots & & & \ddots & \\ I_k & & & & V|_{A_\ell} \end{pmatrix}.$$

We suppress the sets A_i in the notation if it is clear from context what they are. In fact, an alternative definition of a higher order MDS code is that rank $\mathcal{G}[V]$ is always equal to rank $\mathcal{G}[W]$, where W is a generic $k \times n$ matrix (see Corollary 2.2). Further, one need not check this rank equality for all $A_1, \dots, A_\ell$, but rather only needs to check when $\mathcal{G}[W]$ has full column rank (Corollary 2.7) or full row rank (Proposition 2.9). These alternative definitions of higher order MDS codes lead to our definitions of relaxed MDS codes.

- We say that a $k \times n$ matrix V is a *lower* relaxation of a higher order MDS code if $\mathcal{G}_{A_1,\dots,A_\ell}[V]$ has full column rank whenever $\mathcal{G}_{A_1,\dots,A_\ell}[W]$ has full column rank, where W is a generic $(k-d) \times n$ matrix. (Definition 3.1)
- We say that a $k \times n$ matrix V is an *upper* relaxation of a higher order MDS code if $\mathcal{G}_{A_1,\dots,A_\ell}[V]$ has full row rank whenever $\mathcal{G}_{A_1,\dots,A_\ell}[W]$ has full row rank, where W is a generic $(k+d) \times n$ matrix. (Definition 3.4)

In the process of building the theory of these two relaxations, we prove two equivalence theorems: Theorem 3.11 and Theorem 3.15. Theorem 3.11 shows that a code C is rLD-MDS if and only if its dual code $C^\perp$ is a lower relaxed higher order MDS code. This is a direct generalization of the equivalence theorem of [7]. Likewise, Theorem 3.15 shows that, if C' is the dual of a code with a single parity check involving all coordinates, $C' \otimes C$ is an relaxed MR tensor code if and only if C is an upper relaxed higher order MDS code. This generalized an equivalence theorem of [6].

The proofs of these equivalences are relatively straightforward, and mostly mimic arguments in [6] and [7]. However, some care is needed in the proofs as we can no longer assume that the matrix V is MDS. The main utility of these equivalencies is that the notions of rLD-MDS and rMR can be checked using the relatively simple matrix conditions used to check the lower and upper MDS conditions–see Corollary 3.12 and Definition 2.8, respectively. This simplicity helps in proving our main results.

1.1.2 Constructions Using Punctured Codes Coming from Polynomials and Varieties. The GM-MDS theorem [29, 40] using the equivalence in [7] shows that generic Reed-Solomon (RS) codes are higher order MDS codes (and hence LD-MDS as well). Using this insight [21] and [2] showed that a randomly punctured RS code can achieve list decoding capacity over a linear sized field.

The key idea in [21] was that the average list decoding condition was equivalent to checking the rank of a 'Reduced intersection matrix' M. As [7] showed that RS codes achieve list decoding capacity, it was known that M had the right rank for a generic RS code. List decoding close to capacity was then shown to translate to a lower rank condition on M. If an RS code was randomly initialized then M not having the right rank would need many 'faults' (as generically M had a much higher rank). This lead to a significant advantage in calculating the failure probability and [21] showed that randomly punctured RS codes achieve list decoding capacity over quadratic sized fields. [2] improved this analysis at a few key technical points to get linear field sizes.

Our main result vastly generalizes this and gives results for both the upper and lower relaxations (giving us close to capacity list decodable and relaxed MR-tensor codes). To initialize the [2, 21] strategy we need the GM-MDS theorem to hold in a more general setting. In the paper [3], it was shown that the GM-MDS theorem holds for any code where the points are generically from an irreducible variety which contains no hyperplanes through the origin (a particular example would be anything generated by linearly independent polynomials). Using the relaxations discussed earlier we get two matrix conditions for the two relaxations. They generically have much higher rank by the generalized GM-MDS theorem and we are able to follow the argument of [21] and [2] using simple theorems from algebraic geometry (for instance using Bezout's theorem instead of Schwartz-Zippel [33, 41]).

Note, the argument of [21] and [2] cannot be applied directly to the setting of AG codes. Their arguments use the fact that Reed Solomon and Random linear codes are images of polynomial maps. A failure to achieve list decoding capacity implied certain matrices having low rank which gave an algebraic condition on the generator matrix of a code. Substituting the polynomial maps generating the code then allows us to use Schwartz-Zippel to control the number of failures caused by bad columns. AG codes cannot be represented as an image of polynomial maps so we use an algebraic geometric perspective to adapt the arguments of [21] and [2].We use facts proven using the Reimann-Roch theorem (which is also needed to prove classical distance bounds for AG codes) to show that the

columns of an AG code can be treated as points on an irreducible variety of controllable degree. We can then use Bezout's theorem to control the number of bad columns causing an algebraic condition for list-decoding to vanish. This perspective gives us a general statement which works for any code whose columns are sampled from an irreducible variety and recovers the statement for random Reed Solomon and random linear codes from [2].

1.2 Open Problems

We conclude the introduction with a few open problems.

- One question that remains open is if we can achieve (even existentially) the parameters of the random non-linear code for getting ϵ-close to list-decoding capacity, i.e., a field size of $\exp(O(1/\epsilon))$ and list size of $O(1/\epsilon)$, with a linear code. This will match the known lower bound of $\exp(\Omega(1/\epsilon))$ on the field size required to even get polynomial list size [13, 39]. We note that this question is closely related to the current gap in the optimal field size for (n, k)-MDS(L) codes whose currently field size lower/upper bounds are of the form $\exp(\Omega(n))$ and $\exp(O(nL))$, respectively, in the regime that k/n converges to a constant $R \in (0, 1)$ [4, 6].
- Do random puncturings of Reed-Solomon or AG codes have efficient list decoding algorithms up to list decoding capacity? As discussed in [7], current hardness results for list decoding do not apply to randomly punctured Reed-Solomon or AG codes.
- Can we generalize the results in this paper to the duals of punctured AG codes? In particular, does the dual of a punctured AG code achieve list decoding capacity over small fields? Does the dual of a punctured AG code give relaxed MR tensor code with suitable parameters over small fields? Though the generalized GM-MDS theorem of [3] applies to duals of polynomial codes, there are further technical difficulties in generalizing the approach in this paper.
- The upper relaxation of a higher order MDS code corresponds to a relaxed MR(a, b) code in the $a = 1$ case. To what extent can we generalize these results for larger a? Seemingly a road block is a combinatorial characterization of which patterns are correctable by an MR(m, n, a, b) tensor code when $a, b \geq 2$. See [3] for further discussion. But a more approachable open problem is to show that if C_{row} and C_{col} are both random linear codes of constant rate, then $C_{col} \otimes C_{row}$ is a $((1-\epsilon)a, (1-\epsilon)b)$-relaxed (m, n, a, b)-MR tensor code over small fields, maybe even fields of size $\exp(O_\epsilon(1))$ independent of m, n, a, b.

1.3 Organization

In Section 2 we go over essential background on higher order MDS codes. In Section 3, we define various relaxations of higher order MDS codes and show that some of these are equivalent. In Section 4 we show that randomly-punctured algebraic-geometry codes satisfy these relaxed notions over small fields.

2 PRELIMINARIES

In this section, we discuss various background material needed to prove our main results.

2.1 Notation

Given a matrix $M \in \mathbb{F}^{m \times n}$, we say that M has *full row rank* if $\operatorname{rank} M = m$ and that M has *full column rank* if $\operatorname{rank} M = n$. When discussing the properties of higher order MDS codes and their relaxations, we need to consider the ranks of various block matrices (e.g., [6]). Here, we introduce some succinct notation to discuss these matrices. Given a matrix $V \in \mathbb{F}^{k \times n}$ and sets $A_1, \ldots, A_\ell \subseteq [n]$, we define the following (affine) operators:

$$\mathcal{G}_{A_1,\ldots,A_\ell}[V] := \begin{pmatrix} I_k & V|_{A_1} & & & \\ I_k & & V|_{A_2} & & \\ \vdots & & & \ddots & \\ I_k & & & & V|_{A_\ell} \end{pmatrix}, \tag{3}$$

and

$$\mathcal{H}_{A_1,\ldots,A_\ell}[V] := \begin{pmatrix} I_n|_{\overline{A}_1} & I_n|_{\overline{A}_2} & \cdots & I_n|_{\overline{A}_\ell} \\ V|_{\overline{A}_1} & & & \\ & V|_{\overline{A}_2} & & \\ & & \ddots & \\ & & & V|_{\overline{A}_\ell} \end{pmatrix}, \tag{4}$$

where $\overline{A}_i := [n] \setminus A_i$.

2.2 Properties of Higher Order MDS Codes

The proofs and more detailed discussion is in the full version [5].

Proposition 2.1 ([37], as stated in [3]). *Let V be a $k \times n$-matrix. For any $A_1, \ldots, A_\ell \subseteq [n]$, we have that*

$$\dim(V_{A_1} \cap \cdots \cap V_{A_\ell}) = k + \sum_{i=1}^{\ell} \dim(V_{A_i}) - \operatorname{rank} \mathcal{G}_{A_1,\ldots,A_\ell}[V]. \tag{5}$$

We note the following corollary.

Corollary 2.2. *Let V be a $k \times n$-matrix and W a generic $k \times n$-matrix. For all $\ell \geq 2$, we have that V is* MDS(ℓ) *if and only if for all $A_1, \ldots, A_\ell \subseteq [n]$,*

$$\operatorname{rank} \mathcal{G}_{A_1,\ldots,A_\ell}[V] = \operatorname{rank} \mathcal{G}_{A_1,\ldots,A_\ell}[W]. \tag{6}$$

We now state what we believe is a novel adaptation of Proposition 2.1 for parity-check matrices. Note that this formula also relates the rank of $\mathcal{G}(G)$ to the rank of $\mathcal{H}(G^\perp)$.

Lemma 2.3. *Let C be an (n, k)-code (not necessarily MDS) with generator matrix G and parity-check matrix H. For any $A_1, \ldots, A_\ell \subseteq [n]$, we have that*

$$\dim(H_{A_1} \cap \cdots \cap H_{A_\ell}) = n - k + \sum_{i=1}^{\ell} \operatorname{rank}(G|_{\overline{A}_i}) - \operatorname{rank} \mathcal{H}_{A_1,\ldots,A_\ell}[G]. \tag{7}$$

2.2.1 Null Intersection. An important structural property of higher order MDS codes is characterizing when the intersection of spaces is the null space.

Definition 2.4 ([6]). *We say that $A_1, \ldots, A_\ell \subseteq [n]$ have the k-dimensional null intersection property if for any generic $k \times n$ matrix W, we have that $W_{A_1} \cap \cdots \cap W_{A_\ell} = 0$.*

Combinatorial characterizations of the null intersection property were found in [6, 7].

Proposition 2.5 ([7]). *The sets $A_1, \dots, A_\ell \subseteq [n]$ with $|A_i| \le k$ have the k-dimensional null intersection property if and only if for all $i \in [n]$ and for all partitions $P_1 \cup \cdots \cup P_s = [\ell]$, we have that*

$$\sum_{i=1}^{s} \left| \bigcap_{j \in P_i} A_j \right| \le (s-1)k \tag{8}$$

We note that it suffices to verify Definition 1.1 for $A_1, \dots, A_\ell$ has the k-dimensional null intersection property.

Proposition 2.6 ([6]). *For $\ell \ge 2$, a generator matrix $V \in \mathbb{F}^{k \times n}$ with every column nonzero is $\mathrm{MDS}(\ell)$ if and only if for all $A_1, \dots, A_\ell \subseteq [n]$ with the k-dimensional null intersection property and $|A_i| \le k$ we have $V_{A_1} \cap \cdots \cap V_{A_\ell} = 0$.*

We also note the following simple observation will help motivate the lower relaxation of higher order MDS codes.

Corollary 2.7. *Let V be a $k \times n$-matrix and W a generic $k \times n$-matrix. For all $\ell \ge 2$, we have that V is $\mathrm{MDS}(\ell)$ if and only if for all $A_1, \dots, A_\ell \subseteq [n]$, such that $\mathcal{G}_{A_1,\dots,A_\ell}[W]$ has full column rank, we have that (6) holds.*

2.2.2 Saturation. As noted, the null intersection property is closely related to $\mathcal{G}$ having full column rank. We now consider the "dual" situation in which $\mathcal{G}$ has full row rank. We call this the *saturation property*. This property is useful for motivating the upper relaxation of higher order MDS codes.

Definition 2.8. *Let $V \in \mathbb{F}^{k \times n}$. We say that $A_1, \dots, A_\ell \subseteq [n]$ are* V-saturated *if*

$$\operatorname{rank} \mathcal{G}_{A_1,\dots,A_\ell}[V] = \ell k,$$

i.e., $\mathcal{G}_{A_1,\dots,A_\ell}[V]$ has full row rank. We say the sets have the k-dimensional saturation property *if they are W-saturated, where W is a $k \times n$ generic matrix.*

By Proposition 2.1, we have that $A_1, \dots, A_\ell$ are V-saturated if and only if

$$\dim(V_{A_1} \cap \cdots \cap V_{A_\ell}) = \sum_{i=1}^{\ell} \dim(V_{A_i}) - (\ell - 1)k. \tag{9}$$

We note that Proposition 2.6 for null intersection has a analogous result for saturation.

Proposition 2.9. *For $\ell \ge 2$, a generator matrix $V \in \mathbb{F}^{k \times n}$ is $\mathrm{MDS}(\ell)$ if and only if for all $A_1, \dots, A_\ell \subseteq [n]$ with the k-dimension saturation property we have that $A_1, \dots, A_\ell$ are V-saturated.*

2.2.3 MR Tensor Codes. Let $\mathcal{E}^{m,n}_{a,b,p}$ be the set of all subsets $E \subseteq [m] \times [n]$ which are correctable by a (m, n, a, b)-MR-tensor code over a field of characteristic p. If the base field is clear from context, we drop mention of p. It is easy to see that this family is monotone in a and b: for all $a' \le a$ and $b' \le b$, we have that $\mathcal{E}^{m,n}_{a',b'} \subseteq \mathcal{E}^{m,n}_{a,b}$.

We also let $\mathcal{P}^{m,n}_{a,b} \subset \mathcal{E}^{m,n}_{a,b}$ be the set of maximal patterns. In particular, if $m \ge a$ and $n \ge b$, then $\mathcal{P}^{m,n}_{a,b} = \{E \in \mathcal{E}^{m,n}_{a,b} : |E| = bm + an - ab\}$. We recall the regularity theorem of [18].

Theorem 2.10 ([17]). *For $E \subseteq [m] \times [n]$, we have that $E \in \mathcal{E}^{m,n}_{a,b}$ only if for all $S \subseteq [m]$ with $|S| \ge a$ and $T \subseteq [n]$ with $|T| \ge b$, we have that*

$$|E \cap (S \times T)| \ge b|S| + a|T| - ab.$$

Further, this description of $\mathcal{E}^{m,n}_{a,b}$ is complete if $a = 1$.

The following restate some results in [6], their proofs are in our full version [5]. Correctable erasure patterns in the case of $a = 1$ are related to sets with the saturation property.

Proposition 2.11. *Let C_1 be a $(m, m-a)$-code with generator matrix U. Likewise, let C_2 be an $(n, n-b)$-code with generator matrix V. Let $E \subseteq [m] \times [n]$. Also let $A_1, \dots, A_m \subseteq [n]$ such that $\bar{E} = \bigcup_{i=1}^{m} \{i\} \times A_i$ the following are equivalent:*

(a) *E is a correctable pattern.*
(b) *$\sum_{(i,j) \in \bar{E}} \langle U_i \otimes V_j \rangle = \mathbb{F}^{m-a} \otimes \mathbb{F}^{n-b}$.*
(c) *If $a = 1$ and C_1 is MDS, $\operatorname{rank} \mathcal{G}_{A_1,\dots,A_m}[V] = m(n-b)$, i.e., $A_1, \dots, A_m$ are V-saturated.*

Remark 2.1. *Likewise, from the results of [6], one can prove that the vectors $\{U_i \otimes V_j : (i,j) \in \bar{E}\}$ are independent if and only if $G_{A_1,\dots,A_m}[V]$ has full column rank, i.e., $V_{A_1} \cap \dots \cap V_{A_\ell} = 0$ and $\dim V_{A_i} = |A_i|$ for all $i \in [m]$.*

Thus, sets with the null intersection property are related to independent sets in the tensor product matroid, i.e., $A_1, A_2, \dots, A_m$ have k-dimensional null-intersection property if generically $\{U_i \otimes V_j : j \in A_i\}$ are independent in $\mathbb{F}^{m-1} \otimes \mathbb{F}^k$. Likewise, sets with the saturation property are related to spanning sets in the tensor product matroid, i.e., $A_1, A_2, \dots, A_m$ have k-dimensional saturation property if to $\{U_i \otimes V_j : j \in A_i\}$ span $\mathbb{F}^{m-1} \otimes \mathbb{F}^k$.

Lemma 2.12 (Implicit in [6]). *Let $C_1 \subseteq \mathbb{F}^m$ and $C_2 \subseteq \mathbb{F}^n$ be codes with distance at least $a+1$ and $b+1$ respectively. Then, there exists a set $\mathcal{P}^{m,n}_{a,b} \subseteq \mathcal{E}^{m,n}_{a,b}$ such that $C_1 \otimes C_2$ is (a,b)-*MR *if and only if for all $P \in \mathcal{P}^{m,n}_{a,b}$ we have that $C_1 \otimes C_2$ can correct erasure pattern P and*

$$|\mathcal{P}^{m,n}_{a,b}| \le \min\left[2^{mn}, \binom{n}{\le b(m-a)} \binom{bm(m-a)}{\le b(a+1)(m-a)}, \right.$$
$$\left. \binom{m}{\le a(n-b)} \binom{an(n-b)}{\le a(b+1)(n-b)} \right].$$

3 RELAXED HIGHER ORDER MDS CODES

In this section, we define lower and upper relaxations of higher order MDS codes. We then show that lower relaxation is equivalent (up to duality) to relaxed version of optimal list-decodable codes (LD-MDS), and upper relaxation is equivalent to relaxed version of MR tensor codes (with single column parity check).

3.1 Lower and Upper Relaxations of Higher Order MDS Codes

As discussed in Section 2, there are many equivalent criteria for checking if a code is a higher order MDS code, including Corollary 2.7 and Proposition 2.9. Each of these criteria is a logical conjunction of many algebraic conditions (e.g., a matrix is of a specific rank, etc.). As such, these formulations of a higher order MDS code

can be relaxed by only imposing that a (suitable) subset of the algebraic conditions hold. The precise relaxation needs to be carefully chosen so that many "natural" properties of higher order MDS codes continue to hold in these relaxed versions.

More precisely, our *lower* relaxation (Definition 3.1) relaxes Corollary 2.7 so that our code looks like a higher order MDS code of *smaller* rate; while our *upper* relaxation (Definition 3.4) relaxes Proposition 2.9 to look like a higher order MDS code of *larger* rate.

3.1.1 Lower Relaxation. The lower relaxation checks the condition of Corollary 2.7 on a smaller number of families of sets.

Definition 3.1 (Lower Relaxation). *Let $\ell \geq 2$ and $n \geq k \geq d \geq 0$. Let W be a $(k-d) \times n$ generic matrix. We say that a (n,k)-code C with generator matrix G is $\mathrm{rMDS}_d(\ell)$ if $\mathcal{G}_{A_1,\ldots,A_\ell}[G]$ has full column rank whenever $\mathcal{G}_{A_1,\ldots,A_\ell}[W]$ has full column rank.*

We now note an equivalent definition, whose equivalence we prove in the full version [5] of this paper.

Proposition 3.2. *Let $\ell \geq 2$ and $n \geq k \geq d \geq 0$. We say that a (n,k)-code C with generator matrix G is $\mathrm{rMDS}_d(\ell)$ if and only if (1) every column of G is nonzero and (2) for all $A_1, \ldots, A_\ell \subseteq [n]$ of size at most $k-d$ with the $k-d$-dimensional null intersection property we have that $G_{A_1} \cap \cdots \cap G_{A_\ell} = 0$.*

We now state a few basic properties of this relaxation, many of which are analogous to results in [6, 32]. These are proved in our full version [5].

Proposition 3.3. *Let C be an (n,k)-code with generator matrix G. Let $\ell \geq 2$.*

(a) *C is $\mathrm{rMDS}_0(\ell)$ if and only if C is $\mathrm{MDS}(\ell)$.*
(b) *If C is $\mathrm{rMDS}_d(\ell)$, then C is $\mathrm{rMDS}_{d'}(\ell)$ for all $d' \in \{d+1, \ldots, k\}$.*
(c) *If C is $\mathrm{rMDS}_d(\ell)$, then C is $\mathrm{rMDS}_d(\ell')$ for all $\ell' \in \{2, \ldots, \ell-1\}$.*
(d) *C is $\mathrm{rMDS}_d(2)$ if and only if every $k-d$ columns of G are linearly independent.*

3.1.2 Upper Relaxation. We now define the upper relaxation of a higher order MDS code based on the saturation property (Definition 2.8).

Definition 3.4 (Upper Relaxation). *Let $\ell \geq 2$. We say that a (n,k)-code C with generator matrix G is $\mathrm{rMDS}^d(\ell)$ with $0 \leq d \leq n-k$ if for all $A_1, \ldots, A_\ell \subseteq [n]$ with the $k+d$-dimensional saturation property, we have that $A_1, \ldots, A_\ell$ are G-saturated.*

We note that unlike the lower relaxation, some columns of G may equal zero. The following properties are proved in our full version [5].

Proposition 3.5. *Let C be an (n,k)-code with generator matrix G. Let $\ell \geq 2$.*

(a) *C is $\mathrm{rMDS}^0(\ell)$ if and only if C is $\mathrm{MDS}(\ell)$.*
(b) *If C is $\mathrm{rMDS}^d(\ell)$, then C is $\mathrm{rMDS}^{d'}(\ell)$ for all $d' \geq d+1$.*
(c) *If C is $\mathrm{rMDS}^d(\ell)$, then C is $\mathrm{rMDS}^d(\ell')$ for all $\ell' \in \{2, \ldots, \ell-1\}$.*
(d) *C is $\mathrm{rMDS}^d(2)$ if and only if every $k+d$ columns of G span $\mathbb{F}^k$.*

3.1.3 Comparison Between Relaxations. A natural question is if the distinction between upper and lower relaxations is truly necessary. Note that one cannot directly compare the upper and lower relaxations: if C is an (n,k)-code. There are $k+1$ values of d for which one can ask whether C is an $\mathrm{rMDS}_d(\ell)$ code, but there are $n-k+1$ values of d for which one can ask whether C is an $\mathrm{rMDS}^d(\ell)$ code. For this comparison to "type check" we need to compare whether C is $\mathrm{rMDS}_d(\ell)$ with whether the dual code $C^\perp$ is $\mathrm{rMDS}^d(\ell)$. From this perspective, these notions are equivalent for $\ell = 2$, but not necessarily for large ℓ. We first demonstrate the equivalence part of this statement.

Proposition 3.6. *Let C be an (n,k)-code. We have that C is $\mathrm{rMDS}_d(2)$ if and only if $C^\perp$ is $\mathrm{rMDS}^d(2)$.*

However, for larger ℓ this equivalence breaks down.

Proposition 3.7. *The following statements hold.*

(a) *There exists an $\mathrm{rMDS}_d(4)$ code C such that $C^\perp$ fails to be $\mathrm{rMDS}^d(4)$.*
(b) *There exists an $\mathrm{rMDS}^d(4)$ code C such that $C^\perp$ fails to be $\mathrm{rMDS}_d(4)$.*

Proof. These equivalences break down due to counterexamples appeared in [6, Appendix B.3]. □

3.2 Relaxed LD-MDS Codes

To understand codes achieving a list-decoding capacity over small fields, Guo and Zhang [21] used a relaxed Singleton bound. We abstract that into a definition.

Definition 3.8 (Relaxed LD-MDS Codes). *Given $d \geq 0$, let $\rho_{n,k,d}(L) := \frac{L}{L+1}\frac{n-k-d}{n}$. We say that a (n,k)-code C is $\mathrm{rLD\text{-}MDS}_d(L)$ if it is $(\rho_{n,k,d}(L), L)$ average-radius list-decodable. We say a code is $\mathrm{rLD\text{-}MDS}_d(\leq L)$ if it is $\mathrm{rLD\text{-}MDS}_d(\ell)$ for all $\ell \in \{1, \ldots, L\}$.*

Note that $\mathrm{rLD\text{-}MDS}_0(L)$ coincides with $\mathrm{LD\text{-}MDS}(L)$. However, when $d \geq 1$, a $\mathrm{rLD\text{-}MDS}_d(1)$ code may not be MDS. However, we can prove such a code has large distance (analogous to a result of [32]).

Proposition 3.9. *Let C be an (n,k)-code. Then, C is $\mathrm{rLD\text{-}MDS}_d(1)$ if and only if its distance is at least $n-k-d+1$.*

As a corollary, we have that $\mathrm{rLD\text{-}MDS}_d(1)$ is dual to $\mathrm{rMDS}_d(2)$.

Corollary 3.10. *Let C be an (n,k)-code. Then, C is $\mathrm{rLD\text{-}MDS}_d(1)$ if and only if $C^\perp$ is $\mathrm{rMDS}_d(2)$.*

We now generalize this duality for all values of L. In particular, our result is a generalization to relaxed higher order MDS codes of the main theorem from [7] that $\mathrm{LD\text{-}MDS}(\leq L)$ is dual to $\mathrm{MDS}(L+1)$.

Theorem 3.11. *Let C be an (n,k)-code. For any $d \in \{0, \ldots, n-k\}$, we have that C is $\mathrm{rLD\text{-}MDS}_d(\leq L)$ if and only if $C^\perp$ is $\mathrm{rMDS}_d(L+1)$.*

The follow corollary will be useful for analyzing the list decoding properties of algebraic-geometry codes. This result is essentially equivalent to Lemma 3.5 of [21].

Corollary 3.12. *Let C be a (n,k)-code. We have that C is $\mathrm{LD\text{-}MDS}_d(\leq L)$ if and only if for all $\ell \in \{2, \ldots, L+1\}$, and for all*

$A_1, \dots A_\ell \subseteq [n]$ with the $n-k-d$-dimensional null intersection property we have that

$$\operatorname{rank} \mathcal{H}_{A_1,\dots,A_\ell}[G] = n + (\ell - 1)k. \tag{10}$$

For applications to list decoding, we need to understand how (10) changes when some columns of G are deleted.

Proposition 3.13. *Let C be a (n,k)-code with generator matrix G. Let $A_1, \dots A_\ell \subseteq [n]$. For all $B \subseteq \overline{A}_1 \cup \cdots \overline{A}_\ell$, we have that*

$$\operatorname{rank} \mathcal{H}_{A_1,\dots,A_\ell}[G] \geq |B| + \operatorname{rank} \mathcal{H}_{A_1 \setminus B,\dots,A_\ell \setminus B}[G|_{\overline{B}}]. \tag{11}$$

3.3 Relaxed MR Tensor Codes

Definition 3.14 (Relaxed MR Tensor Codes). *Let $0 \leq a' \leq a \leq m$ and $0 \leq b' \leq b \leq n$. Let C_{col} be a $(m, m-a)$-code and C_{row} be a $(n, n-b)$-code. We say that $C_{col} \otimes C_{row}$ is a (a', b')-relaxed (m, n, a, b)-MR tensor code if $C_{col} \otimes C_{row}$ can correct any erasure pattern $E \in \mathcal{E}^{m,n}_{a',b'}$.*

Unlike for MR-tensor codes, C_{col} and C_{row} need not be MDS codes. We show for $a = 1$ this condition can be made equivalent to a suitable upper relaxation of a higher order MDS code.

Theorem 3.15. *Let C be an $(n, n-b)$-code. Let $d \in \{0, 1, \dots, b\}$. Let C' be any $(m, m-1)$-MDS-code. The following are equivalent*

(a) *C is $\mathrm{MDS}^d(m)$.*
(b) *$C' \otimes C$ is $(1, b-d)$-rMR$(m, n, 1, b)$.*

We first observe the following corollary of Proposition 2.11

Proposition 3.16. *Consider $E \subseteq [m] \times [n]$. Let $A_1, \dots, A_m \subseteq [n]$ be such that $\bar{E} = \bigcup_{i=1}^m \{i\} \times A_i$. The following are equivalent.*

(1) *$E \in \mathcal{E}^{m,n}_{1,b}$.*
(2) *$A_1, \dots, A_m$ have the $(n-b)$-dimensional saturation property.*

Proof. Apply Proposition 2.11 (a) and (c) with for a generic $(n-b) \times n$ matrix W. □

Proof of Theorem 3.15. Consider $E \subseteq [m] \times [n]$. Let $A_1, \dots, A_m \subseteq [n]$ be such that $\bar{E} = \bigcup_{i=1}^m \{i\} \times A_i$. Let G be a generator matrix for C.

By Proposition 2.11, we have that E is correctable for $C' \otimes C$ if and only if $A_1, \dots, A_m$ is G-saturated. By Proposition 3.16, we have that $C' \otimes C$ corrects every pattern in $\mathcal{E}^{m,n}_{1,b}$ if and only if every $A_1, \dots, A_m$ with the $n-b$-dimensional saturation property is G-saturated. □

The following lemma is useful in proving that relaxed MR tensor codes exist over small fields.

Lemma 3.17. *Let $E \in \mathcal{E}^{m,n}_{a,b}$. Pick $A \subseteq [m]$ and $B \subseteq [n]$. Let $E' = E \cup A \times [n] \cup [m] \times B$. Then, $E' \in \mathcal{E}^{m,n}_{a',b'}$, where $a' = \min(a + |A|, m)$ and $b' = \min(b + |B|, n)$.*

As a corollary, we get the following fact about the saturation property.

Corollary 3.18. *Let $A_1, \dots, A_m \subseteq [n]$ have the $k+d$-dimensional saturation property. Then, for any $B \subseteq [n]$ of size at most d, we have that $A_1 \setminus B, \dots, A_m \setminus B$ have the k-dimensional saturation property.*

Proof. Let $E = \bigcup_{i=1}^m \{i\} \times \overline{A}_i$ and $E' = \bigcup_{i=1}^m \{i\} \times (\overline{A}_i \cup B)$. By Proposition 3.16, it suffices to prove that if $E \in \mathcal{E}^{m,n}_{1,n-k-d}$ then $E' \in \mathcal{E}^{m,n}_{1,n-k}$. This follows form Lemma 3.17. □

4 CONSTRUCTIONS FROM RANDOMLY PUNCTURED ALGEBRAIC CODES

In this section, we will show that by randomly puncturing algebraic codes (such as Reed-Solomon codes or AG codes or even random linear codes), we can construct both upper and lower relaxed higher order MDS codes over small fields. We first recall some standard facts from algebraic geometry similar to [3].

4.1 Algebraic Preliminaries

Let $\mathbb{F}$ be an algebraically closed field. An ideal $I \subset \mathbb{F}[x_1, \dots, x_k]$ is a subset of polynomials closed under addition and multiplication with arbitrary polynomials in $\mathbb{F}[x_1, \dots, x_k]$. The set of common zeros of a collection of polynomials in $\mathbb{F}[x_1, x_2, \dots, x_k]$ is called a *variety*. The variety $V(I)$ denotes the common zeros of polynomials in an ideal I. The set of all polynomials vanishing on a subset $V \subset \mathbb{F}^k$ is an ideal in $\mathbb{F}[x_1, x_2, \dots, x_k]$ and is denoted by $I(V)$. An ideal $I \subseteq \mathbb{F}[x_1, \dots, x_k]$ is said to be radical if $f^r \in I$ for any r implies $f \in I$. By Hilbert's Nullstellensatz, there is a one-to-one correspondence between radical ideals and varieties, i.e., for a variety X, $I(X)$ is a radical ideal and $V(I(X)) = X$. An irreducible variety X is a variety for which $I(X)$ is a prime ideal.

A Zariski open set over a variety X is a set of the form $U_f = \{x | f(x) \neq 0, x \in X\}$, where $f \in \mathbb{F}[x_1, \dots, x_k]$. The Zariski topology over a variety X is generated by the Zariski open sets over X. Note that X is irreducible if and only if no subset of X can be partitioned into two non-empty Zariski open sets. If U_f is Zariski dense in X, then we say that f is generically non-vanishing. For irreducible X, it is equivalent check that f is nonzero for one point of X or that $f \notin I(X)$. We now state some simple facts about varieties.

Lemma 4.1 (Product of varieties (See 9.4.E in [38])). *If $X_1 \subseteq \mathbb{F}^{k_1}$ and $X_2 \subseteq \mathbb{F}^{k_2}$ are two varieties then $X_1 \times X_2 \subseteq \mathbb{F}^{k_1+k_2}$ is also a variety. Furthermore, if X_1 and X_2 are irreducible then so is $X_1 \times X_2$.*

A non-empty variety X is said to have dimension r if intersecting X with r generic hyperplanes gives a finite set of points. If we count the number of points projectively, that is take $\mathbb{F}_q^n$ as a subset of $\mathbb{PF}_q^n$ and consider hyperplanes in projective space, then this number of points is generically constant and is said to be the degree of X. Bezout's theorem for curves states two irreducible curves of degree d_1 and d_2 intersect in at most $d_1 d_2$ points. The following is a generalization to varieties.

Lemma 4.2 (Bezout's theorem (see 18.6.K in [38])). *If X is an irreducible curve of degree d_1 cut out by I and f is a polynomial of degree d_2 not vanishing on it then the f has at most $d_1 d_2$ zeros on X.*

In general, if X is an irreducible variety of degree d_1 and f is a polynomial of degree d_2 not vanishing on it then $\{x | f(x) = 0\} \cap X$ has degree at most $d_1 d_2$.

4.2 Generalized GM-MDS Theorem for Polynomial Codes

We will be crucially using the generalized GM-MDS theorem for polynomial codes due to Brakensiek-Dhar-Gopi ([3]). We first give the relevant definitions from their work.

Definition 4.3 (MDS(ℓ) property for varieties, [3]). *For an algebraic closed $\mathbb{F}$, we say a variety $X \subseteq \mathbb{F}^k$ satisfies the* MDS(ℓ)

property if and only if for every $k \times n$ matrix V with columns of V initialized from X, V is generically a $[n,k]$-MDS(ℓ) code.

Concretely this means for every $A_1, \ldots, A_\ell \subseteq [n]$ with $|A_i| \le k$ and $|A_1| + \ldots + |A_\ell| = (\ell-1)k$ satisfies $W_{A_1} \cap \ldots \cap W_{A_\ell} = 0$ (equivalently $A_1, \ldots, A_\ell$ satisfies the k-dimensional null intersection property) if and only if $\det(\mathcal{G}_{A_1,\ldots,A_\ell}(V))$ is generically non-vanishing on $X^{(\ell-1)k}$, where W is a $k \times k(\ell-1)k$ generic matrix (that is a matrix whose every entry is an independent formal variable).

Since MDS(2) is equivalent to the usual MDS condition, we call an MDS(2) variety just an MDS variety. We will use the following simple characterization of MDS varieties.

Proposition 4.4 (MDS varieties are not contained in hyper-planes passing through the origin (Theorem 6.3 in [3])). *A variety $X \subseteq \mathbb{F}^k$ is* MDS *if and only if each of its irreducible components is not contained in any hyper-plane passing through the origin. In particular, an irreducible variety is* MDS *if and only if it is not contained in any hyperplane passing through the origin.*

Definition 4.5 (LD-MDS$(\le \ell)$ property for varieties, [3]). *For an algebraic closed $\mathbb{F}$, we say a variety $X \subseteq \mathbb{F}^k$ satisfies the* LD-MDS$(\le \ell)$ *property if and only if for ever $k \times n$ matrix V with columns of V initialized from X, the dual of V is generically a $[n, n-k]$-MDS$(\ell+1)$ code.*

Concretely this means for any $A_1, \ldots, A_{\ell+1} \subseteq [n]$ with $|A_i| \le n-k$ and $|A_1| + \ldots + |A_{\ell+1}| = \ell(n-k)$, satisfies $W_{A_1} \cap \ldots \cap W_{A_{\ell+1}} = 0$ if and only if $\det(\mathcal{G}_{A_1,\ldots,A_\ell}(W^\perp))$ is generically non-vanishing on X^n where W is a $n-k \times n$ generic matrix (that is a matrix whose every entry is an independent formal variable).

The original GM-MDS theorem first conjectured by [11] and proved independently by [29, 40] is equivalent to saying that generic Reed-Solomon codes are higher order MDS for all ℓ, which was shown in [7]. We can now restate the original GM-MDS theorem as follows.

Theorem 4.6 (GM-MDS [7, 11, 29, 40]). *The curve $X = \{(1, t, t^2, \ldots, t^{k-1}) : t \in \mathbb{F}\}$ is* MDS(ℓ) *and* LD-MDS$(\le \ell)$ *for all $\ell \ge 1$.*

[3] generalized the GM-MDS theorem for any irreducible MDS variety X, i.e., as long as X is not contained in a hyperplane through the origin.

Theorem 4.7 (Generalized GM-MDS [3] (Theorems 6.4 and 6.5 in [3])). *For an an algebraic closed $\mathbb{F}$, if an irreducible X is* MDS *(i.e., X is not contained in a hyperplane through the origin) then it is* MDS(ℓ) *for all $\ell \ge 1$ and also* LD-MDS$(\le \ell)$ *for all $\ell \ge 1$.*

4.3 Adapting Previous Works

Extending the argument of [21] and [2], we want to prove that random puncturing of AG codes give us codes which satisfy relaxed versions of the higher order MDS property (as seen earlier the lower relaxation gives us list decoding and upper relaxation imply results for MR tensor codes). The main theorems we want to prove are the following. For the proof of the following theorems (Theorem 4.8 and Theorem 4.9), we may refer to our full version [5].

Theorem 4.8 (Lower Relaxation). *Let $X \subseteq \overline{\mathbb{F}}_q^k$ be an irreducible* MDS *variety and S be a set of points on $X \cap \mathbb{F}_q^k$. Let G be $k \times n$ matrix with columns initialized uniformly at random from S. Then G is* rLD-MDS$_r(\le L)$ *(equivalently $G^\perp$ is a* rMDS$_r(L+1)$*-code) with probability at least*

$$1 - 2^{(L+1)n}\binom{n}{r/2}2^{r(L+1)/2}\left(\frac{P_q(X,L)}{|S|}\right)^{r/2},$$

when repeated columns are allowed in G, or

$$1 - 2^{(L+1)n}\binom{n}{r/2}2^{r(L+1)/2}\left(\frac{P_q(X,L)}{|S|-n}\right)^{r/2},$$

when repeated columns are **not** *allowed in G, where $P_q(X,L)$ is the maximum number of zeros a degree L polynomial generically non-vanishing on X can have over $X \cap \mathbb{F}_q^k$.*

Theorem 4.9 (Upper Relaxation). *Let $X \subseteq \overline{\mathbb{F}}_q^k$ be an irreducible* MDS *variety and S be a set of points on $X \cap \mathbb{F}_q^k$. Let G be $k \times n$ matrix with columns initialized uniformly at random from S and C be the code spanned by it. C is* rMDS$^r(L)$ *with probability at least*

$$1 - C_{n,k,r,L}\binom{n}{r/2}2^{rL/2}\left(\frac{P_q(X,L-1)}{|S|}\right)^{r/2},$$

when repeated columns are allowed in G, or

$$1 - C_{n,k,r,L}\binom{n}{r/2}2^{rL/2}\left(\frac{P_q(X,L-1)}{|S|-n}\right)^{r/2},$$

when repeated columns are **not** *allowed in G, where $P_q(X,\ell)$ is the maximum number of zeros a degree ℓ polynomial generically non-vanishing on X can have over $X \cap \mathbb{F}_q^k$ and*

$$C_{n,k,r,L} = \min\left[2^{Ln}, \binom{(n-k-r)L(L-1)}{\le 2(n-k-r)(L-1)}\binom{n}{\le (n-k-r)(L-1)} + \binom{n}{k+r}\right].$$

Note that C being MDS$^r(L)$ *is equivalent to $C' \otimes C$ being a $(1, n-k-r)$-*rMR$(L, n, 1, n-k)$ *code where C' is a fixed $(L, L-1)$ parity check code.*

Note when puncturing a code (that is G does not have repeated columns), for the theorems to make sense, we need $|X \cap \mathbb{F}_q^k| \gg n$, i.e., there should sufficiently many $\mathbb{F}_q$ rational points on the variety X. For our applications to Reed-Solomon codes (which recovers the result of [2]) and AG codes, X will be a curve (that is a dimension 1 irreducible variety). In this case, we have $P_q(X,\ell) \le \deg(X)\ell$ by Bezout's Theorem (Lemma 4.2). Another application would be to take $X = \overline{\mathbb{F}_q^{-k}}$ which would give $P_q(\overline{\mathbb{F}}_q^k, \ell) \le \ell q^{k-1}$ by the Schwartz-Zippel lemma. In general one could use Bezout's Theorem and explicit Lang-Weil bounds (like the ones here [8, 36]) to bound this for general varieties.

We now recover the Reed-Solomon and Random linear codes result of [2] (we get a slight improvement because in our argument we allowed for constructing the generator matrix of the code with replacement). The proofs can be found in the full version [5].

Corollary 4.10 (punctured RS codes are relaxed LD-MDS over linear field sizes). *For any $\epsilon > 0$, a $[q,k]$ Reed-Solomon code for $q \ge n + k2^{10L/\epsilon}$ on random puncturing to a $[n,k]$ code will be* rLD-MDS$_{\epsilon n}(\le L)$ *with probability at least $1 - 2^{-Ln}$.*

With a similar calculation we get a slight improvement if we allow the generator to have repeated columns.

Corollary 4.11 (Random RS codes are relaxed LD-MDS over linear field sizes). *For any $\epsilon > 0$, a $[n,k]$ Reed-Solomon code with generator G for $q \geq k2^{10L/\epsilon}$ with each column of G initialized randomly as $(1, \alpha, \ldots, \alpha^{k-1})$ with $\alpha \in \mathbb{F}_q$ will be* rLD-MDS$_{\epsilon n}(\leq L)$ *with probability at least $1 - 2^{-Ln}$.*

Corollary 4.12 (random linear codes are relaxed LD-MDS over constant field sizes). *For any $\epsilon > 0$, rate $R > 0$, $q \geq 2^{10L/\epsilon}$, and n large enough, a random linear code will be* rLD-MDS$_{\epsilon n}(\leq L)$ *with probability at least $1 - 2^{-Ln}$.*

We use similar calculations as above but with Theorem 4.9 and Lemma 2.12 to prove the following two corollaries. This extends the results of [2, 21] to the relaxed MR setting (via Theorem 3.15).

Corollary 4.13 (punctured RS codes give relaxed MR codes over linear field sizes (constant rate setting)). *For any $\epsilon > 0$, a $[q,k]$ Reed-Solomon for $q \geq k2^{10L/\epsilon} + n$ on random puncturing to a $[n,k]$ code will be* rMDS$^{\epsilon n}(L)$ *with probability at least $1 - 2^{-Ln}$.*

If we allow repetitions in the generator matrix then a $[n,k]$ Reed-Solomon code with generator G for a field size greater than $k2^{10L/\epsilon}$ with each column of G initialized randomly as $(1, \alpha, \ldots, \alpha^{k-1})$ with $\alpha \in \mathbb{F}_q$ will be rMDS$^{\epsilon n}(L)$ *with probability at least $1 - 2^{-Ln}$.*

Corollary 4.14 (random linear codes give relaxed MR codes over constant field sizes (constant rate setting)). *For any $\epsilon > 0$, a random linear code for $q \geq 2^{10L/\epsilon}$ will be* rMDS$^{\epsilon n}(L)$ *with probability at least $1 - 2^{-Ln}$.*

Unlike the LD-MDS setting, the relaxed MR setting also makes sense for constant/very low co-dimension codes (we skip stating the versions for G with repeated columns as we do not get significant savings here).

Corollary 4.15 (punctured RS codes give relaxed MR codes over polynomial field sizes (low co-dimension setting)). *For any $\epsilon > 0$ and $k = n - b$, a $[q,k]$ Reed-Solomon for $q \gg_{L,b} kn^{4(L-1)/\epsilon}$, on random puncturing to a $[n,k]$ code will be* rMDS$^{b\epsilon}(L)$ *(leading to a $(1,(1-\epsilon)b)$-*rMR$(L,n,1,b)$ *code) with probability at least $1 - n^{-L}$.*

The striking feature about the above bound is how the exponent does not depend on b but only on ϵ and L. By a similar calculation one can show that in this setting random linear codes can shave off a linear factor.

Corollary 4.16 (random linear codes give relaxed MR codes over polynomial field sizes(constant rate setting)). *For any $\epsilon > 0$ and $k = n - b$, a random linear code for $q \gg_{L,b} n^{4(L-1)/\epsilon}$ will be* rMDS$^{\epsilon b}(L)$ *with probability at least $1 - n^{-L}$.*

4.4 Application to AG Codes

We first define evaluation codes. Algebraic geometric codes are instances of this.

Definition 4.17 (Evaluation codes). *Let $F = \{f_1, \ldots, f_k\}$ be a set of rational functions in $\mathbb{F}_q(x_1, x_2, \ldots, x_r)$ and $P = (p_1, \ldots, p_n) \in \mathbb{F}_q^r$ such that the functions in F do not have a pole over the points in P. (In other words substituting any point in P in any function in F gives us an element in $\mathbb{F}_q$.)*

We define the Eval(F,P) *evaluation code as the image of a linear map from the $\mathbb{F}_q^k$ to $\mathbb{F}_q^n$ where we map $(y_1, \ldots, y_k)$ to $(\sum_{i=1}^k y_i f_i(p_1), \ldots, \sum_{i=1}^k y_i f_i(p_n))$.*

Reed Solomon and Gabidulin (c.f., [14]) codes are examples where F are monomials in one variable.

A simple corollary of Theorem 4.7 will imply that a large family of evaluation codes can attain list decoding capacity: if $p_1, \ldots, p_n$ are chosen generically over a MDS irreducible variety X and the $f_1, \ldots, f_k$ are chosen to be linearly independent rational functions over X.

Corollary 4.18. *Let X be an irreducible variety cut out by the prime ideal $I \subseteq \overline{\mathbb{F}}_q[x_1, \ldots, x_r]$. If $F = \{f_1, \ldots, f_k\}$ are linearly independent elements in the fraction field of $\overline{\mathbb{F}}_q[x_1, \ldots, x_r]/I$ then for all integers $n > 0$,* Eval$(F, (p_1, \ldots, p_n))$ *is $[n,k] -$* MDS(ℓ) *and $[n,k] -$* LD-MDS$(\leq \ell)$ *for a generic choice of $p_1, \ldots, p_n$ in X^n.*

In other words the tuple of points $p_1, \ldots, p_n \in X$ for which Eval$(F, (p_1, \ldots, p_n))$ *is* **not** MDS(ℓ) *and* LD-MDS$(\leq \ell)$ *is cut out by a non-vanishing ideal over X^n.*

Algebraic-Geometric codes are evaluations codes where F is chosen to be the spanning set of rational functions on an irreducible (projective) curve X with some constraints on the multiplicity of poles and zeros on a fixed finite subset of X.

To state our theorems we will give a brief description. More details can be found in the following survey [10]. Here we will be interested in projective irreducible curves C. Projective curves are cut out by ideals of homogeneous polynomials in a finite number of variables and are irreducible if the ideal is prime. $\mathbb{P}\overline{\mathbb{F}}_q^n$ is the n dimensional projective space and corresponds to the 0 ideal in $n+1$ variables. If we remove any transverse hyperplane from an irreducible projective curve, we will get an irreducible affine curve so Theorem 4.7 still applies.

We assume C is irreducible throughout for convenience. If C is cut out by an ideal of homogeneous polynomials in $\mathbb{F}_q[x_1, \ldots, x_n]$ then rational functions on C are ratios of homogeneous polynomials of the same degree such that in its reduced form the denominator is non-vanishing on C. A homogeneous polynomial f can be assigned a multiplicity of vanishing on a point. f is said to vanish on x with multiplicity $\text{mult}(f,x) = n$ if all Hasse derivatives of f of order at most n vanish on x. Each rational function f/g where f and g are homogeneous polynomials can now be assigned multiplicity $\text{mult}(f,x) - \text{mult}(g,x)$ at x.

A **divisor** is a finite formal sum of points on C with integer coefficients. A divisor D is said to be $D \geq 0$ if it has only non-negative coefficients. The degree of a divisor is simply the sum of its coefficients.

Given a rational function r we define $\text{div}(r)$ to be the divisor $\sum_{x \in C} \text{mult}(r,x)x$. Such divisors are always degree 0 (geometrically speaking a rational function has the same number of poles and zeros).

Let $D = \sum_{i=1}^t n_i p_i$ be a divisor where $p_1, \ldots, p_n$ are $\mathbb{F}_q$ points on C. $\mathcal{L}(D)$ is defined as a $\mathbb{F}_q$ vector space of rational functions r which satisfy $\text{div}(r) + D \geq 0$.

We are now ready to define AG codes (more details can be found in this survey [10]).

Definition 4.19 (Algebraic Geometry codes). *Let C be a projective irreducible curve and $D = \sum_{i=1}^{t} n_i p_i$ be a divisor where $p_1, \ldots, p_n$ are $\mathbb{F}_q$ points on C. Given a set S of $\mathbb{F}_q$ points disjoint from the support of D, we define* $\mathrm{AG}(C, D, S)$ *to be evaluation code* $\mathrm{Eval}(\mathcal{L}(D), S)$.

We note that by Corollary 4.18 $\mathrm{Eval}(\mathcal{L}(D), S)$ up to a generic choice of S will be MDS. Hence by Corollary 4.18 over the algebraic closure for a generic choice of S algebraic-geometric codes will be MDS(ℓ) and LD-MDS(ℓ). But we want to work with constant sized fields over which generic arguments will not naively apply. We remedy this in the next subsection.

Every projective curve C has an invariant associated to it called its genus g. Algebraically, it can be defined as one minus the constant term of the Hilbert polynomial of the curve (geometrically for complex projective curves it counts the number of 'holes'). The main theorem about AG codes in the literature is the following.

Theorem 4.20 (Theorem 21 from [10]). *For a projective curve C of genus g, let D be a divisor of $\mathbb{F}_q$ points of degree d and S a set of $\mathbb{F}_q$ points disjoint from the support of D such that $d < |S| = n$.* $\mathrm{AG}(C, D, S)$ *will be a* $[n, k = \dim_{\mathbb{F}_q} \mathcal{L}(D)]$ *code with the following properties,*

1. $k \geq d + 1 - g$,
2. *if* $2g - 2 < d$ *then* $k = d + 1 - g$,
3. *the minimum distance of the code is at least* $n - d$.

To prove our theorems about punctured AG codes we need two more algebraic geometric preliminaries and a few lemmas.

First we need the fact that for a divisor D of large enough degree the evaluation $\mathcal{L}(D)$ over a generic point gives us a closed embedding of the curve in projective space (for experts we want $\mathcal{L}(D)$ to be very ample).

Lemma 4.21 (See 15.3.F, 18.4.1, 18.6.I, and 19.2.D in [38]). *Given a projective irreducible curve C of genus g and a divisor D of degree d on it, if $d \geq 2g + 1$ then image of evaluating $\mathcal{L}(D)$ on points in C defines an irreducible projective curve C' of degree d in $\mathbb{P}\overline{F}_q^{d-g}$.*

The above lemma allows us to treat the columns of an AG code as points in an irreducible projective curve allowing us to apply Theorem 4.7. We get a point in $d - g$ projective space because our space of functions is $d - g + 1$ dimensional by Theorem 4.20.

Next we apply a union bound for a particular family of curves to prove our list decoding theorem. First we define and state the properties of the AG code we will use.

The Garcia-Stichtenoth Tower. We briefly summarize some facts, more details can be found in [15]. Let p be a prime power and $q = p^2$. We define a series of curves using the relations,

$$x_{i+1}^p + x_{i+1} = \frac{x_i^p}{x_i^{p-1} + 1}, i = 1, \ldots, t - 1.$$

We let the projectivized curve be $G_{p,t}$. There is only one $\mathbb{F}_q$ point at infinity. The affine curve has $N_q(G_{p,t}) = p^{t-1}(p^2 - p)$, $\mathbb{F}_q$ rational points. The genus $g(G_{p,t})$ is

$$(p^{t/2} - 1)^2 \text{ if } t \text{ is even},$$
$$(p^{(t+1)/2} - 1)(p^{(t-1)/2} + 1) \text{ if } t \text{ is odd}.$$

We will use the bound $p^t + p^{(t+1)/2} - p^{(t-1)/2} - 1 \geq g(G_{p,t}) \geq p^t - 2p^{t/2} + 1$.

Consider the divisor sP_∞ with $s \geq 2g + 1$. By Theorem 4.20, $\mathcal{L}(sP_\infty)$ is $s - g(G_{p,t}) + 1$ dimensional. We use the $N_q(G_{p,t})$ $\mathbb{F}_q$ rational points as evaluation points. Therefore, $G_{p,e}$ with the divisor sP_∞ defines a $[N_q(G_{p,t}), s - g(G_{p,t}) + 1]$ code with distance $N_q(G_{p,t}) - s$. We call this code $GS(p, t, s)$. We also note by Lemma 4.21 the columns of the generating matrix of this code are coming from an irreducible projective curve of degree s. By Corollary 4.18 this curve is MDS because the functions being evaluated are linearly independent.

Theorem 4.22. *Let $\epsilon > 0, R \in (0, 1)$ be rational numbers. If*

$$p \geq 1 + 4/R + 12eL\epsilon^{-1}2^{5(L+2)/\epsilon}, \quad \text{(do } \textbf{not} \text{ allow repetitions in } G)$$
$$p \geq 1 + 12eL\epsilon^{-1}2^{5(L+2)/\epsilon}, \quad \text{(allow repetitions in } G)$$

then for any large enough $n > 0$, such that there exists an integer t satisfying $Rn/2 \geq p^t \geq Rn/4$, a $[n, nR]$ code whose generator matrix is constructed by sampling the columns of the generator of $GS(p, t, s)$ with $s = Rn + g(G_{p,t}) - 1$ will give a $\mathrm{rLD\text{-}MDS}_{\epsilon n}(\leq L)-$ *code over a field of size p^2 with probability at least at least $1 - 1/2^{Ln}$.*

We note the theorem is only meaningful if an infinite sequence of such n can be found. That is true and easy to check. We also note that consecutive such values of n are at most a p multiplicative factor apart from each other.

By choosing the right list size we now show that punctured AG codes can achieve list decoding capacity.

Corollary 4.23. *Let $\epsilon, R \in (0, 1)$ be rationals. Let $L = \lceil 2\frac{1-R}{\epsilon} \rceil$ and p be a prime such that*

$$\log_2 p \gg_R 1/\epsilon^2,$$

any large enough $n > 0$, such that there exists an integer t satisfying $Rn/2 \geq p^t \geq Rn/4$, a random puncturing of $GS(p, t, s)$ with $s = Rn + g(G_{p,t}) - 1$ will give a $(1 - R - \epsilon, L)$ average-radius list-decodable $[n, nR]$-code over a field of size p^2 with probability at least $1 - 1/2^{Ln}$.

Proof. Apply the previous theorem for $\epsilon/2$ and list size $L = \lceil 2\frac{1-R}{\epsilon} \rceil$. □

A similar calculation now applies to the upper MDS relaxation (or equivalently the relaxed MR tensor code setting).

Theorem 4.24. *Let $\epsilon > 0, R \in (0, 1)$ be rational numbers. If*

$$p \geq 1 + 4/R + 12e(L - 1)\epsilon^{-1}2^{5(L+1)/\epsilon},$$

then for any large enough $n > 0$, such that there exists an integer t satisfying $Rn/2 \geq p^t \geq Rn/4$ and $s = Rn + g(G_{p,t}) - 1$ a random puncturing of $GS(p, t, s)$ will give a $\mathrm{rMDS}^{\epsilon n}(L) - [n, nR]$*-code over a field of size p^2 with probability at least $1 - 1/2^{Ln}$.*

We note that we can not prove a similar result in the constant co-dimension setting. This is because we need Rn/p^t to be bounded away from 0 which is not possible if p is already polynomial in n.

ACKNOWLEDGMENTS

We would like to thank Omar Alrabiah, Zeyu Guo, Venkatesan Guruswami and Ray Li for discussions that inspired this work. We also thank Zeyu Guo for clarifying that the constructions of [20] and [26] are non-linear.

REFERENCES

[1] Omar Alrabiah, Venkatesan Guruswami, and Ray Li. 2023. AG codes have no list-decoding friends: Approaching the generalized Singleton bound requires exponential alphabets. *arXiv preprint arXiv:2308.13424* (2023).

[2] Omar Alrabiah, Venkatesan Guruswami, and Ray Li. 2023. Randomly punctured Reed–Solomon codes achieve list-decoding capacity over linear-sized fields. *arXiv preprint arXiv:2304.09445* (2023).

[3] Joshua Brakensiek, Manik Dhar, and Sivakanth Gopi. 2023. Generalized GM-MDS: Polynomial Codes are Higher Order MDS. *arXiv preprint arXiv:2310.12888* (2023).

[4] Joshua Brakensiek, Manik Dhar, and Sivakanth Gopi. 2023. Improved field size bounds for higher order MDS codes. In *2023 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 1243–1248.

[5] Joshua Brakensiek, Manik Dhar, Sivakanth Gopi, and Zihan Zhang. 2023. AG codes achieve list decoding capacity over constant-sized fields. *arXiv e-prints* (2023), arXiv–2310.

[6] Joshua Brakensiek, Sivakanth Gopi, and Visu Makam. 2022. Lower Bounds for Maximally Recoverable Tensor Codes and Higher Order MDS Codes. *IEEE Trans. Inf. Theory* 68, 11 (2022), 7125–7140. https://doi.org/10.1109/TIT.2022.3187366

[7] Joshua Brakensiek, Sivakanth Gopi, and Visu Makam. 2023. Generic Reed-Solomon codes achieve list-decoding capacity. (2023), 1488–1501.

[8] Antonio Cafure and Guillermo Matera. 2006. Improved explicit estimates on the number of solutions of equations over a finite field. *Finite Fields and Their Applications* 12, 2 (2006), 155–185. https://www.sciencedirect.com/science/article/pii/S1071579705000201

[9] Minghua Chen, Cheng Huang, and Jin Li. 2007. On the maximally recoverable property for multi-protection group codes. In *2007 IEEE International Symposium on Information Theory*. IEEE, 486–490.

[10] Alain Couvreur and Hugues Randriambololona. 2020. Algebraic geometry codes and some applications. *ArXiv* abs/2009.01281 (2020). https://api.semanticscholar.org/CorpusID:221669565

[11] Son Hoang Dau, Wentu Song, and Chau Yuen. 2014. On the existence of MDS codes over small fields with constrained generator matrices. In *2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, June 29 - July 4, 2014*. IEEE, 1787–1791. https://doi.org/10.1109/ISIT.2014.6875141

[12] Zeev Dvir and Shachar Lovett. 2012. Subspace evasive sets. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, Howard J. Karloff and Toniann Pitassi (Eds.). ACM, 351–358. https://doi.org/10.1145/2213977.2214010

[13] P Elias. 1957. List Decoding for Noisy Channels. In *IRE WESCON Convention Record, 1957*, Vol. 2. 94–104.

[14] Ernst M Gabidulin. 2021. *Rank codes*. TUM. University Press.

[15] Arnaldo Garcia and Henning Stichtenoth. 1996. On the Asymptotic Behaviour of Some Towers of Function Fields over Finite Fields. *Journal of Number Theory* 61, 2 (1996), 248–273. https://doi.org/10.1006/jnth.1996.0147

[16] Eitan Goldberg, Chong Shangguan, and Itzhak Tamo. 2022. Singleton-type bounds for list-decoding and list-recovery, and related results. (2022), 2565–2570. https://doi.org/10.1109/ISIT50566.2022.9834849

[17] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang, and Sergey Yekhanin. 2017. Maximally recoverable codes for grid-like topologies. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2092–2108.

[18] Parikshit Gopalan, Guangda Hu, Swastik Kopparty, Shubhangi Saraf, Carol Wang, and Sergey Yekhanin. 2017. Maximally Recoverable Codes for Grid-like Topologies. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, Philip N. Klein (Ed.). SIAM, 2092–2108. https://doi.org/10.1137/1.9781611974782.136

[19] Parikshit Gopalan, Cheng Huang, Huseyin Simitci, and Sergey Yekhanin. 2012. On the locality of codeword symbols. *IEEE Transactions on Information theory* 58, 11 (2012), 6925–6934.

[20] Zeyu Guo and Noga Ron-Zewi. 2021. Efficient list-decoding with constant alphabet and list sizes. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 1502–1515.

[21] Zeyu Guo and Zihan Zhang. 2023. Randomly punctured reed-solomon codes achieve the list decoding capacity over polynomial-size alphabets. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 164–176.

[22] Venkatesan Guruswami and Atri Rudra. 2008. Explicit Codes Achieving List Decoding Capacity: Error-Correction With Optimal Redundancy. *IEEE Trans. Inf. Theory* 54, 1 (2008), 135–150. https://doi.org/10.1109/TIT.2007.911222

[23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. 2012. Essential coding theory. *Draft available at http://www. cse. buffalo. edu/ atri/courses/coding-theory/book* (2012).

[24] Venkatesan Guruswami and Chaoping Xing. 2012. Folded codes from function field towers and improved optimal rate list decoding. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, Howard J. Karloff and Toniann Pitassi (Eds.). ACM, 339–350. https://doi.org/10.1145/2213977.2214009

[25] Venkatesan Guruswami and Chaoping Xing. 2013. List decoding reed-solomon, algebraic-geometric, and gabidulin subcodes up to the singleton bound. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, Dan Boneh, Tim Roughgarden, and Joan Feigenbaum (Eds.). ACM, 843–852. https://doi.org/10.1145/2488608.2488715

[26] Venkatesan Guruswami and Chaoping Xing. 2022. Optimal rate list decoding over bounded alphabets using algebraic-geometric codes. *ACM Journal of the ACM (JACM)* 69, 2 (2022), 1–48.

[27] Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin. 2012. Erasure Coding in Windows Azure Storage. In *2012 USENIX Annual Technical Conference, Boston, MA, USA, June 13-15, 2012*, Gernot Heiser and Wilson C. Hsieh (Eds.). USENIX Association, 15–26. https://www.usenix.org/conference/atc12/technical-sessions/presentation/huang

[28] Swastik Kopparty, Noga Ron-Zewi, Shubhangi Saraf, and Mary Wootters. 2018. Improved Decoding of Folded Reed-Solomon and Multiplicity Codes. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, Mikkel Thorup (Ed.). IEEE Computer Society, 212–223. https://doi.org/10.1109/FOCS.2018.00029

[29] Shachar Lovett. 2021. Sparse MDS Matrices over Small Fields: A Proof of the GM-MDS Conjecture. *SIAM J. Comput.* 50, 4, 1248–1262. https://doi.org/10.1137/20M1323345

[30] Farzad Parvaresh and Alexander Vardy. 2005. Correcting Errors Beyond the Guruswami-Sudan Radius in Polynomial Time. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*. IEEE Computer Society, 285–294. https://doi.org/10.1109/SFCS.2005.29

[31] Irving S Reed and Gustave Solomon. 1960. Polynomial codes over certain finite fields. *Journal of the society for industrial and applied mathematics* 8, 2 (1960), 300–304.

[32] Ron M. Roth. 2022. Higher-Order MDS Codes. *IEEE Trans. Inf. Theory* 68, 12 (2022), 7798–7816. https://doi.org/10.1109/TIT.2022.3194521

[33] Jacob T. Schwartz. 1980. Fast Probabilistic Algorithms for Verification of Polynomial Identities. *J. ACM* 27, 4 (1980), 701–717. https://doi.org/10.1145/322217.322225

[34] Chong Shangguan and Itzhak Tamo. 2020. Combinatorial list-decoding of Reed-Solomon codes beyond the Johnson radius. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy (Eds.). ACM, 538–551. https://doi.org/10.1145/3357713.3384295

[35] Richard Singleton. 1964. Maximum distance q-nary codes. *IEEE Transactions on Information Theory* 10, 2 (1964), 116–118.

[36] Kaloyan Slavov. 2023. Nearly sharp Lang–Weil bounds for a hypersurface. *Canad. Math. Bull.* 66, 2 (2023), 654–664. https://doi.org/10.4153/S0008439522000625

[37] Yongge Tian. 2019. Formulas for calculating the dimensions of the sums and the intersections of a family of linear subspaces with applications. *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry* 60, 3 (2019), 471–485.

[38] Ravi Vakil. 2013. *Foundations of algebraic geometry*.

[39] John M Wozencraft. 1958. List decoding. *Quarterly Progress Report* 48 (1958), 90–95.

[40] Hikmet Yildiz and Babak Hassibi. 2019. Optimum Linear Codes With Support-Constrained Generator Matrices Over Small Fields. *IEEE Transactions on Information Theory* 65, 12 (2019), 7868–7875. https://doi.org/10.1109/TIT.2019.2932663

[41] Richard Zippel. 1979. Probabilistic algorithms for sparse polynomials. In *Symbolic and Algebraic Computation, EUROSAM '79, An International Symposiumon Symbolic and Algebraic Computation, Marseille, France, June 1979, Proceedings (Lecture Notes in Computer Science, Vol. 72)*, Edward W. Ng (Ed.). Springer, 216–226. https://doi.org/10.1007/3-540-09519-5_73