

The One-and-a-Half Johnson Bound Is Tight for Proximity Gaps of General Linear Codes

Scott Duke Kominers*

Justin Thaler†

Kai Zhe Zheng‡

September 2026

Abstract

For a linear code $C \subseteq \mathbb{F}_q^n$, we say that C satisfies the *proximity-gaps property* up to distance δ_1 if, for every $\delta_2 > \delta_1$ and every $f, g \in \mathbb{F}_q^n$, at least one of which is δ_2 -far from C in relative Hamming distance, there are only a small fraction (typically at most $\text{poly}(n)/q$) of *exceptional coefficients* $z \in \mathbb{F}_q$ for which $f + zg$ is δ_1 -close to C . The work [BGKS20] shows that every linear code of relative distance δ satisfies proximity gaps up to the one-and-a-half Johnson radius

$$J_{3/2}(\delta) = 1 - (1 - \delta)^{1/3}.$$

We prove that this threshold is tight, at every distance $0 < \delta < 1$, for general linear codes. Specifically, for every $0 < \delta < 1$, we construct a linear code of relative distance arbitrarily close to δ and words f, g which are both $(1 - (1 - \delta)^{4/9})$ -far from the code, but have a constant fraction of coefficients z for which $f + zg$ is nearly $J_{3/2}(\delta)$ -close to the code. Our counterexamples still hold even if one allows a fixed amount of distance-dependent slack.

1 Introduction

Many code-based interactive oracle proofs and SNARKs require simultaneously testing the proximity of several words to a fixed linear code. That is, given words $u_0, \dots, u_{h-1} \in \mathbb{F}_q^n$ and a linear code $C \subseteq \mathbb{F}_q^n$, one wants to test whether each $u_i \in C$, or if at least one u_i is far from C . To avoid running h separate proximity tests, a common approach is to randomly sample coefficients $\gamma_0, \dots, \gamma_{h-1}$ and instead test a single linear combination

$$u^* := \sum_{i=0}^{h-1} \gamma_i u_i.$$

In order for this strategy to be sound, however, one needs the following guarantee, called the *proximity-gaps property*. It says that if at least one u_i is far from C , the random linear combination is still (nearly) just as far from C with high probability.

More formally, we say that the code $C \subseteq \mathbb{F}_q^n$ has the proximity-gaps property up to radius $\rho \in [0, 1]$ with slack $\eta \in [0, 1]$ and error err if, for every $f, g \in \mathbb{F}_q^n$ satisfying $\max\{\delta(f, C), \delta(g, C)\} > \rho + \eta$, the set

$$\mathcal{Z}_\rho(f, g; C) := \{z \in \mathbb{F}_q \mid \delta(f + zg, C) \leq \rho\}$$

*Harvard University and a16z crypto research.

†Georgetown University and a16z crypto research.

‡Simons Institute for the Theory of Computing and IAS. The authors used LLMs to assist with analysis and synthesis in the preparation of this article, in particular GPT-5.6 Sol. The model initially produced one concrete counterexample which the authors generalized to the family of counterexamples.

has size at most $\text{err} \cdot q$. Typically, err should be small enough that $\text{err} \cdot q$ grows at most polynomially in n (ideally linearly in n). In general, err should be bounded above by the target soundness error of the resulting proof system. Here $\delta(\cdot, \cdot)$ denotes the relative Hamming distance, and $\delta(v, C)$ for a word v and code C is the smallest relative Hamming distance between v and any codeword in C .

We study the limitations of the proximity-gaps property for general linear codes. It has long been known [BGKS20] that every linear code C of relative distance δ satisfies the proximity-gaps property at any radius below

$$J_{3/2}(\delta) := 1 - (1 - \delta)^{1/3}.$$

This distance is often called the one-and-a-half Johnson radius and despite much attention, it still stands as the best known radius for the proximity-gaps property of general linear codes (and, as we show in this paper, there is a good reason for this). On the other hand, certain families of codes are known to satisfy much stronger proximity-gaps properties. Reed–Solomon codes have the proximity-gaps property up to the Johnson radius $1 - (1 - \delta)^{1/2}$, rather than only the one-and-a-half Johnson radius [BCI⁺23, BCH⁺25], while random linear codes, folded Reed–Solomon codes, and random Reed–Solomon codes have the proximity-gaps property all the way up to capacity, i.e., radius $1 - R$ for a code of rate R (asymptotically equal to δ for Reed–Solomon codes) [GG25, GGSW26]. The stronger proximity-gaps property of Reed–Solomon codes, along with their optimal rate–distance tradeoff, are the main reason that Reed–Solomon codes are the predominant choice in hash-based SNARKs [BBHR18, ACFY24, ACFY25]. At a high level, these properties allow Reed–Solomon-based SNARKs to achieve the smallest proof sizes among hash-based SNARKs.

On the other hand, several works have proposed hash-based proof systems based on other linear codes in order to obtain desirable properties such as faster encoding or the ability to work over different fields [Ano23, XZS22, BCF⁺25, ZCF24, Ano26]. According to existing analyses, however, the proximity tests in these proof systems are limited to the one-and-a-half Johnson radius leading to larger proof sizes compared to proof systems using Reed–Solomon codes. Many authors have thus considered the possibility that broader classes of linear codes might satisfy stronger, Reed–Solomon-like proximity gaps.¹ In particular, Diamond and Gruen [DG24] conjectured that every linear code of relative distance δ satisfies proximity gaps up to its unique-decoding radius $\delta/2$.

We show in this paper that the one-and-a-half Johnson bound is tight for general linear codes; since for every $0 < \delta < 3 - \sqrt{5}$ our counterexamples exhibit failures at radii between $J_{3/2}(\delta)$ and $\delta/2$, this disproves the Diamond–Gruen conjecture. We discuss these two contributions in more detail below.

1.1 Main results

For a linear code C , write $\delta(C)$ and $R(C)$ for its relative distance and rate. We define

$$\Gamma(\delta) := 1 - (1 - \delta)^{4/9}.$$

Note that for every $0 < \delta < 1$,

$$J_{3/2}(\delta) < \Gamma(\delta) < \delta.$$

For every $\delta \in (0, 1)$ and every $\rho \in (J_{3/2}(\delta), \Gamma(\delta))$, we construct codes of relative distance arbitrarily close to δ and words f, g at distance greater than $\Gamma(\delta)$ from the code, with many linear combinations that are ρ -close to the code. This family of codes shows that the one-and-a-half Johnson bound is tight for proximity gaps of general linear codes. Table 1 compares these thresholds with the unique-decoding and Johnson radii at several representative values of δ .

¹In fact, some proof systems have implicitly conjectured within their security analyses that general linear codes satisfy proximity gaps all the way up to the Johnson bound or even beyond [XZS22].

δ	$J(\delta)$	$\delta/2$	$J_{3/2}(\delta)$	$\Gamma(\delta)$
0.10	0.0513	0.0500	0.0345	0.0457
0.25	0.1340	0.1250	0.0914	0.1200
0.50	0.2929	0.2500	0.2063	0.2651
0.75	0.5000	0.3750	0.3700	0.4600
$3 - \sqrt{5} \approx 0.764$	0.5141	0.3820	0.3820	0.4736
7/8	0.6464	0.4375	0.5000	0.6031

Table 1: Representative values of the unique-decoding radius $\delta/2$, the one-and-a-half Johnson radius $J_{3/2}(\delta)$, the defining-vector threshold $\Gamma(\delta)$, and the Johnson radius $J(\delta) := 1 - \sqrt{1 - \delta}$.

Theorem 1.1. *For every prime p , every $0 < \delta < 1$, every $\rho \in (J_{3/2}(\delta), \Gamma(\delta))$ and every $\eta > 0$, there are constants $A, c > 0$, an integer $D \geq 1$, and an integer $s_0 \geq 1$, depending only on p, δ, ρ, η , with the following property. For every $s \geq s_0$, set $q := p^s$. For every block length N satisfying*

$$As^D \leq N < q,$$

there are a linear code $C \subseteq \mathbb{F}_q^N$ and vectors $f, g \in \mathbb{F}_q^N$ with the following properties.

- **Code parameters.**

$$|\delta(C) - \delta| < \eta, \quad R(C) = 1 - \delta(C) + O_{p, \delta, \rho, \eta}(1/s).$$

- **Distance of the defining vectors.**

$$\delta(f, C), \delta(g, C) > \Gamma(\delta).$$

- **Exceptional coefficients.**

$$|\mathcal{Z}_\rho(f, g; C)| \geq cq.$$

Thus C fails the proximity-gaps property at radius ρ , with every slack $\sigma < \Gamma(\delta) - \rho$ and every error bound $\text{err} < c$. Of particular interest is the setting of [Theorem 1.1](#) in which the one-and-a-half Johnson radius lies below the unique-decoding radius. This is especially interesting because several authors had conjectured that the proximity-gaps property should hold up to the unique-decoding radius; for example, this conjecture is stated explicitly in [\[DG24\]](#). Our results show that this is not the case.

Corollary 1.2. *For every $0 < \delta < 3 - \sqrt{5}$ and every ρ satisfying*

$$J_{3/2}(\delta) < \rho < \min \left\{ \Gamma(\delta), \frac{\delta}{2} \right\},$$

the conclusions of [Theorem 1.1](#) hold and hence there exists linear codes which do not satisfy the proximity-gaps property all the way up to the unique-decoding radius.

Finally, in [Section 4](#) we give a concrete instantiation with block length 2^{36} over the field $\mathbb{F}_{2^{128}}$, relative distance between 0.4419 and 0.4511, rate between 0.5410875 and 0.5502875, counterexample words that are at least 0.22945 far from the code, and more than 2^{54} exceptional coefficients at radius 0.1816. Thus the number of exceptional coefficients exceeds the block length by a factor greater than 2^{18} , and the exceptional radius lies inside the unique-decoding radius. This example is included to illustrate that the counterexamples do not require exorbitant field sizes or block lengths, nor vanishing distances or rates. Specifically, our concrete instantiation is at a parameter setting which is reasonable in hash-based SNARKs.

1.2 Technical overview

Here we give a high level overview of our counterexample construction. We roughly proceed in two steps, starting with a small-distance base construction and then amplifying distance using standard probabilistic techniques to obtain counterexamples at all distances. The bulk of our work goes towards the first step.

The small-distance base construction. Fix a small constant $x > 0$. We first construct a counterexample code with small relative distance $3x > 0$. At this scale, $J_{3/2}(3x) \approx x$ and $\Gamma(3x) \approx 4x/3$. Towards our construction, we use a finite field of size $Q = p^a$. There are some requirements to this Q , but they are not important for the overview, so we ignore them for now. Consider the space $\mathbb{F}_Q^m$ for some dimension m and let X denote the set of lines in $\mathbb{F}_Q^m$. We start by constructing a code

$$D \subseteq \mathbb{F}_p^X$$

which is indexed by the set of lines X in $\mathbb{F}_Q^m$. The code D is constructed as follows. For each codimension two subspace $W \subseteq \mathbb{F}_Q^m$, we let $u_W \in \mathbb{F}_p^X$ be its incidence vector. That is, for a line $\ell \in X$,

$$(u_W)_\ell := \begin{cases} 1 & \text{if } \ell \subseteq W, \\ 0 & \text{otherwise.} \end{cases}$$

Then we define

$$D := \text{span}_{\mathbb{F}_p}(u_W : W \subseteq \mathbb{F}_Q^m \text{ codimension two}).$$

The motivation for considering this D is that we have a good understanding of its small weight codewords. More specifically, using results of Adriaensen and Denaux [AD21] we can obtain the following two useful properties. First, every word of relative weight at most about $3x$ lies in the span of at most two incidence vectors, $u_W, u_{W'}$. Second, every word of relative weight at most about $4x/3$ lies in the span of a single incidence vector. Thus, although D may have large dimension, its sufficiently low-weight words have a very simple structure.

We exploit this structure to obtain both our final code C and the two words $f, g \in \text{span}_{\mathbb{F}_q}(D)$ giving the counterexample by imposing two linear parity checks. Specifically, we look for a linear map

$$\Psi : \text{span}_{\mathbb{F}_q}(D) \rightarrow \mathbb{F}_q^2$$

where $\mathbb{F}_q$ is a sufficiently large extension of $\mathbb{F}_Q$ and define $C := \ker(\Psi)$. Here C is a subcode of $\text{span}_{\mathbb{F}_q}(D)$, i.e. the extension of D to an $\mathbb{F}_q$ code. We construct Ψ so that the incidence vectors u_W have pairwise linearly independent syndromes $\Psi(u_W)$. That is, no two $W, W' \subseteq \mathbb{F}_Q^m$ of codimension two have $\Psi(u_W)$ and $\Psi(u_{W'})$ that are scalar multiples of each other. Finding this Ψ , combined with our understanding of low weight words in D , essentially gives all of the properties we need.

- **Distance:** The distance of C is at least $3x$, because any nonzero word of relative weight at most about $3x$ in $\text{span}_{\mathbb{F}_q}(D)$ would have to be a linear combination of at most two incidence vectors, but no such combination can have zero syndrome due to the linear independence of the incidence syndromes.
- **Counterexample $f, g \in \mathbb{F}_q^X$:** By a straightforward counting argument there exist words $f, g \in \text{span}_{\mathbb{F}_q}(D)$ whose syndromes are not scalar multiples of any incidence vector syndrome. This ensures that $\delta(f, C), \delta(g, C) > 4x/3$ because any word of relative weight at most about $4x/3$ in $\text{span}_{\mathbb{F}_q}(D)$ must be a multiple of a single incidence vector, and hence cannot have a syndrome that is a scalar multiple of $\Psi(f)$ or $\Psi(g)$.

- **Exceptional Coefficients:** Note that $\Psi(f), \Psi(g)$ are linearly independent. Then we can show that each z such that $\Psi(f) + z\Psi(g)$ is a scalar multiple of some $\Psi(u_W)$ is an exceptional coefficient where $f + zg$ is x -close to C . Indeed, if for such z we get that there is some αu_W such that $f + zg - \alpha u_W \in C$, and therefore

$$\delta(f + zg, C) \leq \frac{\text{wt}(u_W)}{|X|} \approx x.$$

Moreover each u_W gives a distinct exceptional coefficient, so we have many of them by counting the number of codimension two subspaces $W \subseteq \mathbb{F}_Q^m$. In particular, we have at least $\Omega(Q^{2m})$ of them. Ultimately we arrange $\Omega(Q^{2m}) = \Omega(q)$ yielding a constant fraction of exceptional coefficients in the final field $\mathbb{F}_q$ of the code C .

Amplifying to all distances. After we obtain a small distance counterexample, we amplify to obtain codes of all distances. Our amplification procedure also reduces the field size and it works in a straightforward way. We first choose a subfield $\mathbb{F}_{q'} \subseteq \mathbb{F}_q$ such that $q = (q')^r$ for some constant r , and view each symbol of $\mathbb{F}_q$ as an r -dimensional vector over $\mathbb{F}_{q'}$. It is straightforward to transform f, g into words over $\mathbb{F}_{q'}$ which still have many exceptional coefficients in $\mathbb{F}_{q'}$, so the remaining issue is to amplify the distance. We then define an amplified code as follows. For each coordinate of the new code, we choose a random t -tuple of coordinates of the original word and apply a random $\mathbb{F}_{q'}$ -linear map from the corresponding t symbols in $\mathbb{F}_q$ to a single symbol in $\mathbb{F}_{q'}$ (so $r \cdot t$ symbols are being combined). This defines a new code over $\mathbb{F}_{q'}$. If an original word has relative weight a , then a new coordinate is nonzero with probability approximately $1 - (1 - a)^t$. Hence, by standard concentration arguments, the relevant distances are amplified according to this transformation $a \rightarrow 1 - (1 - a)^t$. Choosing t appropriately sends the three base distances x , $4x/3$, and $3x$ to values close to their target values $J_{3/2}(\delta)$, $\Gamma(\delta)$, and δ , respectively. We also include one enlargement step to improve the rate of our code to nearly-MDS. This step is also done relatively straightforwardly.

1.3 Related Work

The work [BGKS20] gives a proximity-gaps counterexample at $(\delta, \rho) = (7/8, 1/2)$. That is, a code with distance $7/8$ that does not satisfy the proximity-gaps property for any radius above $J_{3/2}(7/8) = 1/2$. This gives one distance δ where the one-and-a-half Johnson radius is tight, but [BGKS20] does not give any counterexamples at other distances. Diamond and Gruen [DG24] constructed a Reed–Solomon example at the unique-decoding radius with exactly n exceptional coefficients. The presumed tightness of this example (not only among Reed–Solomon codes but also amongst all linear codes) is what motivated their conjecture that all linear codes satisfy proximity gaps up to their unique-decoding radius, with at most n exceptional coefficients.

For the special class of Reed–Solomon codes, a number of recent works have given counterexamples at various distances, but the distances of these counterexamples are near capacity [DG25, Kam26, KKH26]. Their intention is to show the limits of the recent Grand MCA Challenge [ABF26] which conjectures that Reed–Solomon codes satisfy proximity gaps all the way up to capacity (or equivalently, up to their distance δ).

Our work is orthogonal to this line of research on Reed–Solomon counterexamples as we focus on general linear codes.

Finally, we remark that two the works [GKL24, Zei24] show that all linear codes exhibit the stronger properties of correlated agreement and mutual correlated agreement up to the one-and-a-half Johnson radius. Our work focuses on providing a counterexample for the weaker “plain” proximity-gaps property and hence shows that the radius for these results is tight too.

1.4 Organization

In [Section 2](#) we describe preliminary notation and basic facts. [Section 3](#) gives the proof of our main theorem. [Section 4](#) gives a concrete instantiation within a parameter regime that is commonly used in hash-based SNARKs. The sole purpose of this last section is to demonstrate that such counterexamples do not require large field sizes or other extreme parameter choices.

2 Preliminaries

We collect the notation and basic facts used throughout the construction. For a positive integer n , write $[n] := \{1, \dots, n\}$. For $x \in \mathbb{F}_q^n$, let $\text{wt}(x)$ and $\text{supp}(x)$ denote its Hamming weight and support. For a linear subspace $H \leq \mathbb{F}_q^n$, write $\text{supp}(H) := \bigcup_{h \in H} \text{supp}(h)$. For a linear code $C \subseteq \mathbb{F}_q^n$, set

$$d(C) := \min_{0 \neq c \in C} \text{wt}(c), \quad \delta(C) := \frac{d(C)}{n}, \quad R(C) := \frac{\dim C}{n}.$$

For $x \in \mathbb{F}_q^n$, define its relative distance to C by

$$\delta(x, C) := \frac{1}{n} \min_{c \in C} \text{wt}(x - c).$$

For $f, g \in \mathbb{F}_q^n$, define

$$\mathcal{Z}_\rho(f, g; C) := \{z \in \mathbb{F}_q : \delta(f + zg, C) \leq \rho\}.$$

We call the elements of this set the exceptional coefficients of f, g with respect to C at radius ρ . When f, g , and C are clear from context, we simply call them exceptional coefficients. The Singleton defect of C is $n - \dim C - d(C) + 1$, and C is MDS when this quantity is zero. We suppress floors and ceilings in asymptotic constructions; when a support size such as xn is required to be integral, we restrict to an appropriate subsequence of n , or equivalently insert the corresponding floor or ceiling. The subscript in a term such as $o_n(1)$ indicates the variable tending to infinity. We also write

$$\Phi_t(u) := 1 - (1 - u)^t.$$

For $S \subseteq [n]$, write

$$\mathbb{F}_q^S := \{v \in \mathbb{F}_q^n : \text{supp}(v) \subseteq S\}.$$

For a linear code $D \subseteq \mathbb{F}_q^n$, its shortening to S is

$$D[S] := D \cap \mathbb{F}_q^S.$$

3 Proof of the main theorem

In this section we construct the counterexample for our main theorem, which is restated below for convenience.

Theorem 3.1. *For every prime p , every $0 < \delta < 1$, every*

$$J_{3/2}(\delta) < \rho < \Gamma(\delta),$$

and every $\eta > 0$, there are constants $A, c > 0$, an integer $D \geq 1$, and an integer $s_0 \geq 1$, such that the following holds. For every $q := p^s \geq p^{s_0}$, every block length

$$As^D \leq N < q$$

there are an $\mathbb{F}_q$ -linear code $C \subseteq \mathbb{F}_q^N$ and words $f, g \in \mathbb{F}_q^N$ with the following properties.

- *Code parameters.*

$$|\delta(C) - \delta| < \eta, \quad R(C) = 1 - \delta(C) + O_{p,\delta,\rho,\eta}\left(\frac{1}{s}\right).$$

- *Distance of the defining vectors.*

$$\delta(f, C), \delta(g, C) > \Gamma(\delta).$$

- *Exceptional coefficients.*

$$|\mathcal{Z}_\rho(f, g; C)| \geq cq.$$

3.1 The small-distance base code

We start by constructing the small distance counterexample. One should think of the distance of the code as being a small $3x > 0$, the one-and-a-half Johnson radius as x , and the distance of the two counterexample words from the code as at least $4x/3$.

Lemma 3.2. *Fix a prime p and a power $Q = p^a > 32$, with $Q \notin \{49, 121\}$. For all sufficiently large m , set $q_m := Q^{2m}$. Then there exist a linear code $C_m \subseteq \mathbb{F}_{q_m}^{n_m}$ and vectors $f_m, g_m \in \mathbb{F}_{q_m}^{n_m}$ with the following properties.*

- *Code parameters.*

$$\frac{L_m}{n_m} < \delta(C_m) < \frac{3e_m}{n_m}.$$

- *Distance of the defining vectors.*

$$\delta(f_m, C_m), \delta(g_m, C_m) > \frac{4e_m}{3n_m}.$$

- *Exceptional coefficients.*

$$|\mathcal{Z}_{e_m/n_m}(f_m, g_m; C_m)| \geq c_Q \cdot q_m.$$

where $c_Q > 0$ is a constant depending only on Q , where

$$n_m := \frac{Q^m - 1}{Q - 1}, \quad e_m := \frac{Q^{m-2} - 1}{Q - 1}, \quad L_m := 3Q^{m-3} - 3Q^{m-4} - 1.$$

Our code for [Lemma 3.2](#) is constructed in the following steps. Fix p, Q, m as in the lemma. We let X denote the set of one-dimensional subspaces, or lines, of $\mathbb{F}_Q^m$.

- For every $W \subseteq \mathbb{F}_Q^m$ of codimension two, we define its incidence vector $u_W \in \mathbb{F}_p^X$ as the vector indexed by each line $\ell \in X$ with $u_W(\ell) = 1$ if $\ell \subseteq W$ and 0 otherwise. We start by considering the ambient code

$$U_m := \text{span}_{\mathbb{F}_p} \{u_W : W \in \text{Gr}(m-2, \mathbb{F}_Q^m)\}.$$

A key fact, from [\[AD21\]](#), is that every low-weight codeword in U_m is a linear combination of at most two incidence vectors, $u_W, u_{W'}$.

- We then design two linear parity checks $\Psi_0 : \text{span}_{\mathbb{F}_{Q^{2m}}}(U_m) \rightarrow \mathbb{F}_{Q^{2m}}^2$ and define our base code as the kernel of this map. Here, note that our parity checks are defined over a larger field $\mathbb{F}_{Q^{2m}}$, so our final code is a subspace of $\mathbb{F}_{Q^{2m}}^X$. Specifically it is the kernel $\ker(\Psi_0)$. We call this code C_m .

The parameters in [Lemma 3.2](#) count the following objects: n_m is the number of lines in $\mathbb{F}_Q^m$, e_m is the number of lines in a codimension-two subspace of $\mathbb{F}_Q^m$, and L_m is a cutoff weight that we will use for low-weight codewords in U_m . The result of Adriaensen and Denaux [\[AD21\]](#) that we need is stated below.

Theorem 3.3. *Every word of U_m of weight at most L_m is an $\mathbb{F}_p$ -linear combination of at most two incidence vectors u_W .*

Proof. A codimension-two vector subspace of $\mathbb{F}_Q^m$ has projective dimension $k = m - 3$, so in the notation of [\[AD21, Section 3.1\]](#), $U_m = \mathcal{C}_{m-3}(m-1, Q)$. Because $Q > 32$ and $Q \notin \{49, 121\}$, the field size Q lies in one of the classes Q_3, Q_4, Q_5 of [\[AD21, Definition 3.2\]](#). In all three cases their cutoff $W(k, Q)$ is at least $3Q^k - 3Q^{k-1} - 1 = L_m$. The conclusion now follows from [\[AD21, Theorem 5.9\]](#). $\square$

From [Theorem 3.3](#) we get the following properties of the ambient code U_m . At a high level, these statements help us control the shortenings of the code.

Lemma 3.4. *For every sufficiently large m and every $S \subseteq X$, the following properties hold.*

- **Two-incidence-vector containment.** *If $|S| \leq L_m$, then $U_m[S]$ is contained in the span of two incidence vectors.*
- **One-incidence-vector containment.** *If $|S| \leq 4e_m/3$, then $U_m[S]$ is contained in the span of one incidence vector.*
- **Minimum distance.** *In particular,*

$$d(U_m) = e_m.$$

Proof. Any eight distinct incidence vectors, $u_{W_1}, \dots, u_{W_8}$ are linearly independent. Indeed, u_{W_1} has support of size e_m while for each $i = 2, \dots, 8$, the supports of u_{W_1} and u_{W_i} share at most $(Q^{m-3} - 1)/(Q - 1)$ indices, and

$$7 \frac{Q^{m-3} - 1}{Q - 1} < e_m.$$

By symmetry, every u_{W_i} contains at least one coordinate in its support that is not in the support of any other u_{W_j} , $j \neq i$. Hence $u_{W_1}, \dots, u_{W_8}$ are linearly independent. Now set S of size $|S| \leq L_m$. Towards the proof of the first item we first show that $\dim U_m[S] \leq 2$. To see this, suppose that $v_1, v_2, v_3 \in U_m[S]$ are linearly independent. By [Theorem 3.3](#), each v_i is a linear combination of at most two incidence vectors, all drawn from a collection $u_{W_1}, \dots, u_{W_6}$. Then we can find a $v \in U_m[S]$ such that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3,$$

and the right hand side is a nonzero linear combination of at least 3 of the $u_{W_1}, \dots, u_{W_6}$. Once again using [\[AD21\]](#) v can also be written as a linear combination of at most two incidence vectors, hence the equation above gives a non-trivial linear combination of at most 8 incidence vectors that is zero, which is a contradiction. Hence $\dim U_m[S] \leq 2$. We now consider all 3 cases for the dimension of $U_m[S]$. First, suppose that $\dim U_m[S] \leq 1$. Then the conclusion is immediate by [Theorem 3.3](#).

Now suppose $\dim U_m[S] = 2$. Choose a basis for it, v_1, v_2 say. By [Theorem 3.3](#), each v_i is a linear combination of at most two incidence vectors, say u_{W_1}, u_{W_2} for v_1 and u_{W_3}, u_{W_4} for v_2 . Every nonzero linear combination of the basis has, by [Theorem 3.3](#), a representation involving at most two incidence vectors of codimension 2 subspaces. Now identify each $v \in U_m[S]$ with the coefficients of its representation as a linear combination of $u_{W_1}, \dots, u_{W_4}$. Hence, there is a homomorphism from $U_m[S]$ to $\mathbb{F}_p^4$ such that $U_m[S]$ is identified with a two-dimensional subspace of $\mathbb{F}_p^4$. Moreover, by [Theorem 3.3](#), every nonzero vector in the image of $U_m[S]$ has Hamming weight at most two. One can check that this space is either supported on two coordinates, in which case we are done, or we are in the case $p = 2$ and, after rescaling and relabeling, its three nonzero vectors are

$$(1, 1, 0, 0), \quad (1, 0, 1, 0), \quad (0, 1, 1, 0).$$

However, this case cannot occur, as the three vectors represented by these coordinates would together have support of size at least

$$3e_m - 3 \frac{Q^{m-3} - 1}{Q - 1} = 3Q^{m-3} > L_m.$$

back in $\mathbb{F}_p^X$. This is not possible if they are all in $U_m[S]$ since $|S| \leq L_m$. This establishes the first item and we move on to the second item. Now suppose $|S| \leq 4e_m/3$. By the first item, the shortening $U_m[S]$ is in the span of at most two incidence vectors, say u_W and $u_{W'}$ for $W \subseteq \mathbb{F}_Q^m$ of codimension 2. We first have to rule out the case that its dimension is exactly two. If this were the case, $U_m[S]$ would contain nonzero multiples of both incidence vectors, would have to contain all indices in the union of their supports. Hence,

$$e_m + Q^{m-3} > \frac{4}{3}e_m,$$

a contradiction. This shows $U_m[S]$ has dimension at most one. Finally to complete the proof we show that the vector generating $U_m[S]$ is not of the form $\alpha u_W + \beta u_{W'}$ for nonzero α, β . To see this, first if $\alpha + \beta = 0$, then we would get

$$\text{wt}(\alpha u_W + \beta u_{W'}) \geq 2Q^{m-3},$$

while if $\alpha + \beta \neq 0$ we get

$$\text{wt}(\alpha u_W + \beta u_{W'}) \geq e_m + Q^{m-3}.$$

Both bounds exceed $4e_m/3$ for $Q > 32$, completing the proof of the second item. $\square$

Identify $\mathbb{F}_Q^m$ with $V := \mathbb{F}_{Q^m}$ as $\mathbb{F}_Q$ -vector spaces. We now go from the ambient code U_m to our next code C_m by constructing two parity checks given by a map $\Psi_0 : \text{span}_{\mathbb{F}_{Q^{2m}}}(U_m) \rightarrow \mathbb{F}_{Q^{2m}}^2$. For a word u , we refer to $\Psi_0(u)$ as its syndrome. Our goal with the parity checks is to have all of the distinguished low-weight words u_W for W of codimension 2 have syndromes that are linearly independent – or have distinct directions. Note that in order to accomplish this, we need to set $\mathbb{F}_{Q^{2m}}$ to be a large extension of $\mathbb{F}_{Q^m}$.

Lemma 3.5. *Let $q := q_m = Q^{2m}$. There exists an $\mathbb{F}_q$ -linear map $\Psi_0 : \text{span}_{\mathbb{F}_q}(U_m) \rightarrow \mathbb{F}_q^2$ with the following properties.*

- **Nonzero incidence syndromes.** *For every $W \in \text{Gr}(m-2, V)$,*

$$\Psi_0(u_W) \neq 0.$$

- **Independent syndrome directions.** For any distinct $W, W' \in \text{Gr}(m-2, V)$, $\Psi_0(u_W)$ and $\Psi_0(u_{W'})$ are linearly independent over $\mathbb{F}_q$.

Proof. For $W \in \text{Gr}(m-2, V)$, write its subspace polynomial as

$$L_W(X) = \prod_{w \in W} (X - w) = X^{Q^{m-2}} + a_{m-3}X^{Q^{m-3}} + a_{m-4}X^{Q^{m-4}} + \cdots + a_0X.$$

First, note that the coefficients (a_{m-3}, a_{m-4}) determine W . Intuitively, this is true because each coefficient gives a linear constraint on the elements of W and two linear constraints are needed to define a codimension-two subspace. More formally, this is clearly true for $m = 4$. For $m \geq 5$, if two subspace polynomials have the same pair of Q^{m-3} and Q^{m-4} coefficients, then their difference is a polynomial of degree at most Q^{m-5} that vanishes on the intersection of the corresponding subspaces, which has dimension at least $m - 4$. Hence the difference must be zero. Now define the three values

$$(d_0, d_1, d_2) := (Q^{m-2} - 1, 2Q^{m-2} - Q^{m-3} - 1, 2Q^{m-2} - Q^{m-4} - 1).$$

Taking the logarithmic derivative of L_W , we get

$$\sum_{w \in W} w^{d_0} = a_0, \quad \sum_{w \in W} w^{d_1} = -a_0 a_{m-3}, \quad \sum_{w \in W} w^{d_2} = -a_0 a_{m-4}.$$

Indeed, $L'_W(X) = a_0$, and

$$\frac{a_0}{L_W(X)} = \sum_{w \in W} \frac{1}{X - w} = \sum_{r \geq 0} \left(\sum_{w \in W} w^r \right) X^{-r-1}.$$

Expanding the left-hand side as a Laurent series at infinity and comparing the relevant coefficients gives the three identities above. Choose $\beta \in \mathbb{F}_q \setminus \mathbb{F}_{Q^m}$, and choose a nonzero representative $x_\lambda \in \lambda$ for each $\lambda \in X = \text{Gr}(1, V)$. Since each d_i is divisible by $Q - 1$, the value $x_\lambda^{d_i}$ is independent of the choice of representative. Define

$$\Psi_0(v) := \left(\sum_{\lambda \in X} v_\lambda x_\lambda^{d_0}, \sum_{\lambda \in X} v_\lambda x_\lambda^{d_1} + \beta \sum_{\lambda \in X} v_\lambda x_\lambda^{d_2} \right).$$

Then Ψ_0 is $\mathbb{F}_q$ -linear and for every $W \in \text{Gr}(m-2, V)$,

$$\Psi_0(u_W) = \frac{a_0}{Q-1} (1, -a_{m-3} - \beta a_{m-4}).$$

Since $a_0 \neq 0$, the syndrome $\Psi_0(u_W)$ is nonzero. It remains to show that these syndromes are pairwise linearly independent. Suppose $W, W' \in \text{Gr}(m-2, V)$ satisfy

$$\text{span}_{\mathbb{F}_q}(\Psi_0(u_W)) = \text{span}_{\mathbb{F}_q}(\Psi_0(u_{W'})).$$

Since the first coordinate of each syndrome is nonzero, normalizing that coordinate to 1 gives

$$a_{m-3} + \beta a_{m-4} = a'_{m-3} + \beta a'_{m-4}.$$

Because 1 and β are linearly independent over $\mathbb{F}_{Q^m}$, it follows that

$$a_{m-3} = a'_{m-3}, \quad a_{m-4} = a'_{m-4}.$$

Hence $W = W'$. Therefore the incidence vectors u_W have pairwise linearly independent syndromes. $\square$

Fix U_m and Ψ_0 as in [Lemma 3.5](#), and set $D_m := \text{span}_{\mathbb{F}_q}(U_m)$ and $C_m := \ker(\Psi_0)$. We are ready to prove [Lemma 3.2](#).

Proof of Lemma 3.2. The desired code is C_m . We first bound the distance of C_m . For the lower bound, suppose that $0 \neq c \in C_m$ has support S with $|S| \leq L_m$. By this identity and [Lemma 3.4](#), c lies in the $\mathbb{F}_q$ -span of at most two incidence vectors, say $u_W, u_{W'}$. By [Lemma 3.5](#), the syndromes $\Psi_0(u_W), \Psi_0(u_{W'})$ are linearly independent, so no nonzero linear combination of $u_W, u_{W'}$ is in C_m , contradicting $c \in C_m$. Therefore $\delta(C_m) > \frac{L_m}{n_m}$. For the upper bound, choose three codimension-two subspaces W_1, W_2, W_3 inside a common $(m-1)$ -dimensional subspace such that their pairwise intersections have dimension $m-3$ and their triple intersection has dimension $m-4$. Since

$$\Psi_0(u_{W_1}), \Psi_0(u_{W_2}), \Psi_0(u_{W_3}) \in \mathbb{F}_q^2,$$

these three syndrome vectors are linearly dependent. Hence there is a nontrivial linear combination $h = c_1 u_{W_1} + c_2 u_{W_2} + c_3 u_{W_3} \in C_m$. The incidence vectors are linearly independent, so $h \neq 0$. Moreover,

$$\text{supp}(h) \subseteq \text{supp}(u_{W_1}) \cup \text{supp}(u_{W_2}) \cup \text{supp}(u_{W_3}).$$

By inclusion-exclusion,

$$|\text{supp}(u_{W_1}) \cup \text{supp}(u_{W_2}) \cup \text{supp}(u_{W_3})| = 3e_m - 3 \frac{Q^{m-3} - 1}{Q - 1} + \frac{Q^{m-4} - 1}{Q - 1} = 3Q^{m-3} + \frac{Q^{m-4} - 1}{Q - 1} < 3e_m.$$

Therefore $\text{wt}(h) < 3e_m$, which proves the upper bound on $\delta(C_m)$. We next find the counterexample words f_m and g_m . By the displayed identity and [Lemma 3.4](#), every word in D_m of weight at most $4e_m/3$ lies in the $\mathbb{F}_q$ -span of a single incidence vector. Hence the syndrome of every such word lies on one of the incidence syndrome directions. There are $q+1$ one-dimensional subspaces of $\mathbb{F}_q^2$, while

$$|\text{Gr}(m-2, \mathbb{F}_Q^m)| = \left(\frac{1}{(Q^2-1)(Q^2-Q)} + o_m(1) \right) q.$$

Thus, for all sufficiently large m , there are at least two syndrome directions that are not represented by incidence vectors. After a linear transformation, we can assume without loss of generality that these directions are $(1, 0)$ and $(0, 1)$. We take $f_m, g_m \in D_m$ to be the words with

$$\Psi_0(f_m) = (1, 0), \quad \Psi_0(g_m) = (0, 1).$$

Now we establish that each word is at Hamming distance greater than $4e_m/3$ from C_m . If f_m were at Hamming distance at most $4e_m/3$ from C_m , then for some $c \in C_m$, the word $f_m - c$ would have weight at most $4e_m/3$ and syndrome $(1, 0)$. By [Lemma 3.4](#), every word of this weight is a scalar multiple of an incidence vector u_W for $W \in \text{Gr}(m-2, \mathbb{F}_Q^m)$, but by construction this word cannot have syndrome in $\text{span}((1, 0))$. Hence, we have a contradiction. The same argument applies to g_m , so we get

$$\delta(f_m, C_m), \delta(g_m, C_m) > \frac{4e_m}{3n_m}.$$

Finally, we show that f_m, g_m have many exceptional coefficients. For each $W \in \text{Gr}(m-2, \mathbb{F}_Q^m)$, we know that $\Psi_0(u_W)$ is in neither $\text{span}((1, 0))$ nor $\text{span}((0, 1))$. We may therefore choose a scalar $\alpha_W \in \mathbb{F}_q^\times$ such that

$$\Psi_0(\alpha_W u_W) = (1, \beta_W)$$

for some $\beta_W \in \mathbb{F}_q$. The values β_W , as W ranges over $\text{Gr}(m-2, \mathbb{F}_Q^m)$, are distinct because the incidence vectors have pairwise linearly independent syndromes. Moreover, $\Psi_0(f_m + \beta_W g_m - \alpha_W u_W) = 0$,

and therefore $f_m + \beta_W g_m - \alpha_W u_W \in C_m$. Since $\text{wt}(u_W) = e_m$, we obtain $\delta(f_m + \beta_W g_m, C_m) \leq \frac{e_m}{n_m}$. Thus every β_W lies in $\mathcal{Z}_{e_m/n_m}(f_m, g_m; C_m)$. Since $|\text{Gr}(m-2, \mathbb{F}_Q^m)| = \Theta_Q(q)$, there is a constant $c_Q > 0$ such that

$$|\mathcal{Z}_{e_m/n_m}(f_m, g_m; C_m)| \geq c_Q q,$$

as desired. $\square$

3.2 Amplifying to all distances

[Lemma 3.2](#) gives a counterexample at small distance, and in this section we show how to amplify it to obtain counterexamples at all distances. In the process, we also reduce the alphabet size. At a high level, our strategy is to first view the code from [Lemma 3.2](#) as a code over a smaller field by simply viewing its alphabet symbols as vectors over a subfield – this makes the distance even smaller. We then amplify the distance by taking random linear combinations of subsets of coordinates, again over the subfield. It will be helpful to define and recall

$$\Phi_t(u) = 1 - (1 - u)^t, \quad \Gamma(\delta) = 1 - (1 - \delta)^{4/9}$$

Intuitively, we will amplify all of the distances in our small distance counterexample [Lemma 3.2](#) according to $u \mapsto \Phi_t(u)$. For small u and large t , observe that the code's distance, the proximity-gaps radius, and the distance of the counterexample words will approach their desired targets $\delta, J_{3/2}(\delta), \Gamma(\delta)$ respectively.

After amplifying, we are not guaranteed to get near-MDS parameters, but we resolve this in a straightforward manner by randomly enlarging the code.

Proof of [Theorem 3.1](#). Set

$$\eta_0 := \min \left\{ \eta, \frac{\delta - \Gamma(\delta)}{4} \right\}.$$

Choose $a \geq 1$ so that $Q := p^a > 32$, $Q \notin \{49, 121\}$, and Q is sufficiently large in terms of δ, ρ, η . Choose an integer $t \geq 1$ and $\xi > 0$ such that, for all sufficiently large m ,

$$\Phi_t\left(\frac{e_m}{n_m}\right) + 3\xi < \rho, \quad \Phi_t\left(\frac{4e_m}{3n_m}\right) - 3\xi > \Gamma(\delta),$$

and

$$|\Phi_t(u) - \delta| < \frac{\eta_0}{2} \quad \forall u \in \left[\frac{L_m}{n_m}, \frac{3e_m}{n_m} \right].$$

This maps all of the relevant distances we consider (the distance of f, g from the code, the distance of the exceptional combinations to the code, and the distance of the code itself) sufficiently close to their targets. Now fix s , set $q := p^s$, and let $d := \gcd(s, 2a)$, $m := s/d$, and $r := 2a/d$. Then $r \leq 2a$ and $Q^{2m} = q^r$. Thus, for all sufficiently large s , [Lemma 3.2](#) gives a code $C_m \subseteq \mathbb{F}_{q^r}^{n_m}$ and vectors f_m, g_m satisfying

$$\frac{L_m}{n_m} < \delta(C_m) < \frac{3e_m}{n_m}, \quad \delta(f_m, C_m), \delta(g_m, C_m) > \frac{4e_m}{3n_m},$$

with at least $c_Q q^r$ exceptional coefficients at radius e_m/n_m . Write $E := \mathcal{Z}_{e_m/n_m}(f_m, g_m; C_m)$ for simplicity. Note that $0 \notin E$. We partition the nonzero elements of $\mathbb{F}_{q^r}$ into $(q^r - 1)/(q - 1)$ one-dimensional $\mathbb{F}_q$ -subspaces, that is, identifying each element with a line in $\mathbb{F}_q^r$. Since $|E| \geq c_Q q^r$, one of these subspaces contains at least

$$c_Q q^r \frac{q - 1}{q^r - 1} \geq \frac{c_Q}{2} q$$

elements of E . Write this subspace as $b\mathbb{F}_q$ for some $b \in \mathbb{F}_{q^r}^\times$. Thus there is a set $T \subseteq \mathbb{F}_q$ of size at least $(c_Q/2)q$ such that $bt \in E$ for every $t \in T$. Replacing g_m by bg_m and continuing to denote the result by g_m , we have

$$\delta(f_m + tg_m, C_m) \leq \frac{e_m}{n_m}$$

for every $t \in T$. Since C_m is $\mathbb{F}_{q^r}$ -linear, this replacement preserves $\delta(g_m, C_m)$. Hence, this gives us cq exceptional coefficients in $\mathbb{F}_q$, where $c := c_Q/2$, for f_m, g_m . We can now view the code as living over $\mathbb{F}_q$ while preserving a constant fraction of exceptional coefficients. We now amplify and reduce the alphabet in one step. For each $j \in [N]$, independently sample $i_{j,1}, \dots, i_{j,t} \in [n_m]$ uniformly with replacement and a uniformly random $\mathbb{F}_q$ -linear functional $\ell_j : (\mathbb{F}_{q^r})^t \rightarrow \mathbb{F}_q$, and define

$$\mathcal{A}(v)_j := \ell_j(v_{i_{j,1}}, \dots, v_{i_{j,t}}).$$

If v has relative weight u , then

$$\Pr[\mathcal{A}(v)_j \neq 0] = \left(1 - \frac{1}{q}\right) \Phi_t(u).$$

Hence, by a Chernoff bound, the relative weight of $\mathcal{A}(v)$ is within ξ of this quantity except with probability $e^{-\Omega_\xi(N)}$. It remains to make this concentration hold simultaneously for all vectors that occur in the argument. All of these vectors lie in

$$D_m := \text{span}_{\mathbb{F}_{q^r}}(U_m) :$$

this includes the code C_m , the vectors f_m, g_m , all differences $f_m - c$ and $g_m - c$, and all of the exceptional witnesses, so it suffices to show that all target distances are as desired for every vector in D_m . To count how many vectors we need to union bound over, we bound the dimension of D_m . Let $k_m := \dim_{\mathbb{F}_p} U_m$. View $\mathbb{F}_Q^m$ as an am -dimensional vector space over $\mathbb{F}_p$. Every codimension-two $\mathbb{F}_Q$ -subspace W then has codimension $2a$ over $\mathbb{F}_p$, so there are $\mathbb{F}_p$ -linear forms $\lambda_1, \dots, \lambda_{2a}$ whose common zero set is W . The polynomial

$$P_W(x) := \prod_{j=1}^{2a} (1 - \lambda_j(x)^{p-1})$$

is the indicator of W and has degree at most $2a(p-1)$. Choosing one nonzero representative x_ℓ for each $\ell \in X$, we have

$$u_W = (P_W(x_\ell))_{\ell \in X}.$$

Hence U_m is contained in the image of the evaluation map from polynomials in am variables of degree at most $2a(p-1)$ to $\mathbb{F}_p^X$. Therefore

$$k_m \leq \binom{am + 2a(p-1)}{2a(p-1)} = O_{p,Q}(m^{2a(p-1)}).$$

Then $\dim_{\mathbb{F}_{q^r}} D_m = k_m$ and $|D_m| = q^{rk_m}$. For any fixed $v \in D_m$, the preceding Chernoff bound fails with probability at most $2e^{-2\xi^2 N}$. Thus the probability that it fails for some $v \in D_m$ is at most $2q^{rk_m} e^{-2\xi^2 N}$. Now $r = O_{p,Q}(1)$, $m = O(s)$, and $\log q = \Theta_p(s)$, while $k_m = O_{p,Q}(m^{2a(p-1)})$. Consequently

$$rk_m \log q = O_{p,Q}(s^{2a(p-1)+1}).$$

It follows that for suitable constants $A > 0$ and $D = 2a(p-1) + 1$, whenever $N \geq As^D$ the last probability is less than one. We may therefore fix a map $\mathcal{A}$ such that simultaneously for every $v \in D_m$,

$$\frac{\text{wt}(\mathcal{A}(v))}{N} = \left(1 - \frac{1}{q}\right) \Phi_t\left(\frac{\text{wt}(v)}{n_m}\right) \pm \xi.$$

Set

$$C_0 := \mathcal{A}(C_m), \quad f := \mathcal{A}(f_m), \quad g := \mathcal{A}(g_m).$$

For sufficiently large s , we have $1/q < \xi$. Thus the displayed estimate differs from $\Phi_t\left(\frac{\text{wt}(v)}{n_m}\right)$ by at most 2ξ , uniformly over $v \in D_m$. Applying this first to the nonzero words of C_m , and using monotonicity of Φ_t , gives

$$|\delta(C_0) - \Phi_t(\delta(C_m))| \leq 2\xi.$$

Since

$$\frac{L_m}{n_m} < \delta(C_m) < \frac{3e_m}{n_m},$$

the choice of t and ξ gives

$$|\delta(C_0) - \delta| < \eta_0 \leq \eta,$$

as ξ is sufficiently small. Similarly, for every $c \in C_m$,

$$\frac{\text{wt}(\mathcal{A}(f_m - c))}{N} \geq \Phi_t(\delta(f_m, C_m)) - 2\xi,$$

and hence

$$\delta(f, C_0) \geq \Phi_t\left(\frac{4e_m}{3n_m}\right) - 2\xi > \Gamma(\delta).$$

The same argument gives $\delta(g, C_0) > \Gamma(\delta)$. Finally, for every $z \in T$ there is some $c_z \in C_m$ such that

$$\frac{\text{wt}(f_m + zg_m - c_z)}{n_m} \leq \frac{e_m}{n_m}.$$

Therefore

$$\delta(f + zg, C_0) \leq \Phi_t\left(\frac{e_m}{n_m}\right) + 2\xi < \rho.$$

Thus every $z \in T$ remains exceptional after amplification. Since $|T| \geq cq$, we conclude that

$$|\mathcal{Z}_\rho(f, g; C_0)| \geq cq.$$

It remains only to make the code near-MDS. Let $d_0 := d(C_0)$ and $b_0 := \lfloor \Gamma(\delta)N \rfloor$, and set

$$h := \left\lceil \log_q(8 \cdot 2^N) \right\rceil.$$

Since $\delta(C_0) > \delta - \eta_0$ and $\eta_0 \leq (\delta - \Gamma(\delta))/4$, we have $d_0 - b_0 - 1 > 0$ for all sufficiently large s . If C_0 already has Singleton defect at most h , set $C := C_0$. Otherwise, put $K := N - d_0 + 1 - h$ and choose a uniformly random K -dimensional supercode $C \supseteq C_0$. For any fixed $x \notin C_0$,

$$\Pr[x \in C] = \frac{q^{K - \dim C_0} - 1}{q^{N - \dim C_0} - 1} \leq 2q^{K-N} = 2q^{-d_0+1-h}.$$

There are at most $2^N q^{d_0-1}$ vectors of weight at most $d_0 - 1$. Since none of these nonzero vectors belongs to C_0 , the probability that C has minimum distance below d_0 is at most $2^{N+1} q^{-h} \leq \frac{1}{4}$.

Similarly, if either f or g is within Hamming distance b_0 of C , then for some $y \in \{f, g\}$ and some word u of weight at most b_0 , the vector $y - u$ lies in C . None of these vectors lies in C_0 , by the distance bounds already proved. There are at most $2^{N+1}q^{b_0}$ such vectors, so the probability of this event is at most

$$2^{N+2}q^{b_0-d_0+1-h} \leq \frac{1}{2}.$$

Thus with positive probability both properties hold simultaneously. Fix such a supercode C . Then $d(C) = d_0$ and

$$\delta(f, C), \delta(g, C) > \Gamma(\delta).$$

Moreover, C has Singleton defect at most h . Since $q = p^s$ and $N < q$,

$$\frac{h}{N} = O_p\left(\frac{1}{s}\right),$$

and hence

$$R(C) = 1 - \delta(C) + O_{p,\delta,\rho,\eta}\left(\frac{1}{s}\right).$$

Finally, $C_0 \subseteq C$, so passing to C can only decrease the distance of $f + zg$ to the code. Therefore every coefficient exceptional for C_0 remains exceptional for C , and

$$|\mathcal{Z}_\rho(f, g; C)| \geq cq.$$

Together with $\delta(C) = \delta(C_0)$, this gives $|\delta(C) - \delta| < \eta$ and proves [Theorem 3.1](#). $\square$

4 A counterexample of concrete size

We conclude by giving a concrete counterexample. The purpose is to demonstrate that our counterexamples can exist at rates, block lengths, and field sizes that are within practical ranges for SNARK constructions. To obtain useful finite parameters, we use the same construction as our main theorem but give a significantly tighter analysis of the amplification step so as to maintain a reasonable block length. This requires replacing the generic union bound by a finer-grained analysis that separates vectors by weight. The counterexample below and its analysis were found by using AI, specifically GPT-5.5, to specialize our main construction to reasonable concrete parameters. The proof below is written by the same AI as our primary motivation for this section is as a proof-of-concept.

Theorem 4.1. *There are an $\mathbb{F}_{2^{128}}$ -linear code $C \subseteq \mathbb{F}_{2^{128}}^{2^{36}}$ and vectors $f, g \in \mathbb{F}_{2^{128}}^{2^{36}}$ with the following properties.*

- *Code parameters.*

$$0.4419 < \delta(C) < 0.4511, \quad 0.5410875 < R(C) < 0.5502875.$$

- *Distance of the defining vectors.*

$$\delta(f, C), \delta(g, C) > 0.22945.$$

- *Exceptional coefficients.* At radius $\rho := 0.1816 < \delta(C)/2$, we have

$$|\mathcal{Z}_\rho(f, g; C)| > 2^{54}$$

Proof. We specialize the Grassmannian incidence construction to $Q = 512 = 2^9$ and $m = 5$, and take the syndrome field to be $q = 2^{128}$. The raw block length and incidence weight are

$$n := n_5 = \frac{512^5 - 1}{511} = 68,853,957,121, \quad e_5 := \frac{512^3 - 1}{511} = 262,657.$$

The lower-distance cutoff is $L_5 = 3 \cdot 512^2 - 3 \cdot 512 - 1 = 784,895$; set $d_0 := 784,896$ and $d_1 := 3 \cdot 512^2 + 1 = 786,433$. The number of codimension-two subspaces of $\mathbb{F}_{512}^5$ is

$$B := \binom{5}{2}_{512} = \frac{(512^5 - 1)(512^4 - 1)}{(512^2 - 1)(512 - 1)} = 18,049,720,589,484,545 > 2^{54}.$$

Since $90 \nmid 128$, the explicit syndrome field used in [Lemma 3.5](#) is not a subfield of $\mathbb{F}_q$. We instead choose a uniformly random $\mathbb{F}_q$ -linear map

$$\Psi_0 : \text{span}_{\mathbb{F}_q}(U_5) \longrightarrow \mathbb{F}_q^2.$$

For distinct codimension-two subspaces W, W' , the vectors $u_W, u_{W'}$ are linearly independent, so their images under a uniformly random Ψ_0 are independent uniform vectors in $\mathbb{F}_q^2$. The probability that these two images are linearly dependent is

$$q^{-1} + q^{-2} - q^{-3}.$$

Consequently, a union bound gives

$$\binom{B}{2} (q^{-1} + q^{-2} - q^{-3}) < 4.8 \cdot 10^{-7} < 1.$$

Thus there is a map for which all incidence-vector syndromes are pairwise linearly independent. Fix such a map and set

$$C_5 := \ker(\Psi_0).$$

The map Ψ_0 is surjective, and hence

$$\dim_{\mathbb{F}_q} C_5 = \dim_{\mathbb{F}_2} U_5 - 2.$$

Since $B < q - 1$, there are at least two syndrome directions that are not represented by incidence vectors. By surjectivity, we can choose $f_5, g_5 \in \text{span}_{\mathbb{F}_q}(U_5)$ with syndromes in two such directions, exactly as in the proof of [Lemma 3.2](#). The same proof shows that C_5 has minimum distance at least d_0 , has a nonzero codeword of weight at most d_1 , and satisfies

$$n \delta(f_5, C_5), n \delta(g_5, C_5) > \frac{4e_5}{3}.$$

Thus, by integrality,

$$n \delta(f_5, C_5), n \delta(g_5, C_5) \geq 350,210.$$

Moreover, every codimension-two subspace gives a distinct exceptional coefficient z , and there is a word v_z of weight e_5 such that

$$f_5 + zg_5 - v_z \in C_5.$$

We amplify to block length $N := 2^{36}$ using locality $t := 52,500$. For each $j \in [N]$, independently sample $i_{j,1}, \dots, i_{j,t} \in [n]$ uniformly with replacement and choose a uniformly random $\mathbb{F}_q$ -linear functional $\ell_j : \mathbb{F}_q^t \rightarrow \mathbb{F}_q$. Define

$$A(v)_j := \ell_j(v_{i_{j,1}}, \dots, v_{i_{j,t}}).$$

If $\text{wt}(v) = w$, then $\text{wt}(A(v))$ is binomial with N trials and success probability

$$p_w := \left(1 - \frac{1}{q}\right) \left(1 - \left(1 - \frac{w}{n}\right)^t\right).$$

At the four relevant weights,

w	p_w
262,657	0.1814919007...
350,210	0.2343498938...
784,896	0.4503500124...
786,433	0.4509937976...

We now choose a single map A for which all relevant vectors concentrate simultaneously. For $0 < a, b < 1$, write

$$D(a\|b) := a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}.$$

We use the Chernoff estimate

$$\Pr[\text{Bin}(N, b) \leq aN] \leq e^{-ND(a\|b)} \quad (a < b),$$

and the analogous upper-tail estimate when $a > b$. We first record the dimension of the base code. By Hamada's rank formula [Ham68, Theorem 1], specialized to points and planes of $\text{PG}(4, 2^9)$, the binary rank is $1 + \text{tr}(D^9)$, where, indexing the rows and columns by $(4, 3)$,

$$D_{s,s'} = [x^{2s'-s}](1+x)^5, \quad D = \begin{pmatrix} \binom{5}{4} & \binom{5}{2} \\ \binom{5}{5} & \binom{5}{3} \end{pmatrix} = \begin{pmatrix} 5 & 10 \\ 1 & 10 \end{pmatrix}.$$

Direct exponentiation therefore gives

$$\dim_{\mathbb{F}_2} U_5 = 1 + \text{tr}(D^9) = 3,604,584,376.$$

The two independent syndrome checks therefore give

$$k := \dim_{\mathbb{F}_q} C_5 = 3,604,584,374.$$

For a set $S \subseteq [n]$ of size w , the shortening $C_5[S]$ has minimum distance at least d_0 . Singleton therefore gives

$$\dim C_5[S] \leq \min\{k, \max\{0, w - d_0 + 1\}\}.$$

Consequently, the number of codewords of weight w is at most

$$\binom{n}{w} q^{\min\{k, \max\{0, w - d_0 + 1\}\}}.$$

The same bound applies to the intersection of either affine coset $f_5 + C_5$ or $g_5 + C_5$ with a fixed weight layer, since for a fixed support its intersection with the coset is either empty or an affine translate of a shortening of C_5 . Set $W := 12,550,000$. For the derivative calculations below, regard w as a real variable and put

$$r_w := \left(1 - \frac{w}{n}\right)^t.$$

For fixed $a < p_w$, we have

$$\frac{\partial}{\partial b} D(a||b) = \frac{b-a}{b(1-b)}, \quad \frac{p'_w}{1-p_w} = \frac{t}{n-w} \frac{(1-q^{-1})r_w}{q^{-1} + (1-q^{-1})r_w}.$$

For $w \leq W$, we have $r_w \geq r_W > 6.97 \cdot 10^{-5}$, and hence

$$N \frac{p_w - a}{p_w(1-p_w)} p'_w > 0.999 \frac{Nt}{n-w} \left(1 - \frac{a}{p_w}\right). \quad (1)$$

We first control the minimum distance of the image code. Set $\alpha := 0.4419$. For $d_0 \leq w \leq W$, using $\binom{n}{w} \leq (en/w)^w$, the logarithm of the union-bound contribution at weight w is at most

$$F(w) := w \log \frac{en}{w} + (w - d_0 + 1) \log q - ND(\alpha||p_w).$$

Here $W - d_0 + 1 = 11,765,105 < k$, so $w - d_0 + 1$ is the active shortening exponent throughout this interval. Moreover,

$$F'(w) = \log \frac{n}{w} + \log q - N \frac{p_w - \alpha}{p_w(1-p_w)} p'_w.$$

Since $p_w \geq p_{d_0} > 0.45035$, Equation (1) gives

$$N \frac{p_w - \alpha}{p_w(1-p_w)} p'_w > 982, \quad \log \frac{n}{w} + \log q \leq \log \frac{n}{d_0} + 128 \log 2 < 101.$$

Consequently, $F'(w) < -881$ throughout the interval. At $w = d_0$,

$$d_0 \log \frac{en}{d_0} + \log q < 9,718,630 < 9,922,939 < ND(0.4419||p_{d_0}).$$

Thus the total contribution of these weight layers is exponentially small. For $w \geq W$, we instead use the global bound $|C_5| = q^k$. Since p_w is increasing,

$$k \log q < 319,808,959,479 < 319,867,786,434 < ND(0.4419||p_W).$$

Thus, except with negligible probability,

$$\text{wt}(A(c)) > 0.4419N \quad \text{for every nonzero } c \in C_5.$$

We next control the two affine cosets. Set $\beta := 0.22945$ and $w_* := 350,210$. There are no words of $f_5 + C_5$ or $g_5 + C_5$ of weight below w_* . For weights below d_0 , each support contains at most one word from either coset; above d_0 we again use the shortening bound. The logarithmic exponents of the union-bound contributions are

$$G_0(w) := w \log \frac{en}{w} - ND(\beta||p_w) \quad (w_* \leq w < d_0)$$

and

$$G_1(w) := w \log \frac{en}{w} + (w - d_0 + 1) \log q - ND(\beta||p_w) \quad (d_0 \leq w \leq W).$$

At the smallest possible weight,

$$w_* \log \frac{en}{w_*} < 4,618,908 < 4,620,069 < ND(\beta||p_{w_*}).$$

Since $p_w \geq p_{w_*} > 0.2343$ on the first interval and $p_w \geq p_{d_0} > 0.45035$ on the second, Equation (1) gives

$$G'_0(w) = \log \frac{n}{w} - N \frac{p_w - \beta}{p_w(1 - p_w)} p'_w, \quad G'_1(w) = G'_0(w) + \log q,$$

and therefore

$$G'_0(w) < 13 - 1083 < -1070, \quad G'_1(w) < 101 - 25,675 < -25,574,$$

respectively. At the transition between the intervals, using $\log q < 89$, we have

$$G_1(d_0) = G_0(d_0) + \log q < G_0(w_*) - 1070(d_0 - w_*) + 89 < G_0(w_*).$$

Thus the largest exponent over all weights below W occurs at $w = w_*$. For $w \geq W$,

$$k \log q < 319,808,959,479 < 469,738,761,478 < ND(0.22945 \| p_W).$$

Therefore, simultaneously for every $c \in C_5$,

$$\text{wt}(A(f_5 - c)) > 0.22945N, \quad \text{wt}(A(g_5 - c)) > 0.22945N$$

except with negligible probability. We now control the exceptional witnesses. Each v_z has weight exactly e_5 , and $ND(0.1816 \| p_{e_5}) > 2702$. Since $\log B < 38$, a union bound shows that, except with probability less than e^{-2664} ,

$$\text{wt}(A(v_z)) \leq 0.1816N$$

simultaneously for all B distinguished coefficients. Finally, let $h_0 \in C_5$ be the nonzero word of weight at most d_1 given by the three-incidence construction. Since $ND(0.4511 \| p_{d_1}) > 1565$, we have

$$\text{wt}(A(h_0)) < 0.4511N$$

except with probability less than e^{-1565} . The sum of all failure probabilities is less than one. Hence there is a choice of A for which all of the preceding properties hold simultaneously. Fix such a map and set

$$C_0 := A(C_5), \quad f_0 := A(f_5), \quad g_0 := A(g_5).$$

Since every nonzero codeword maps to a word of positive weight, A is injective on C_5 , so $\dim C_0 = k$. We have

$$0.4419 < \delta(C_0) < 0.4511, \quad \delta(f_0, C_0), \delta(g_0, C_0) > 0.22945.$$

Furthermore, for every distinguished coefficient z ,

$$f_0 + zg_0 - A(f_5 + zg_5 - v_z) = A(v_z),$$

with $A(f_5 + zg_5 - v_z) \in C_0$. Hence $\delta(f_0 + zg_0, C_0) \leq 0.1816$, and therefore

$$|\mathcal{Z}_{0.1816}(f_0, g_0; C_0)| \geq B.$$

It remains to complete C_0 to a near-MDS code. Put $d := d(C_0)$, $b := \lfloor 0.22945N \rfloor$, and

$$h := \left\lceil \log_q(4 \cdot 2^N) \right\rceil = \left\lceil \frac{N+2}{128} \right\rceil = 536,870,913.$$

Since $d < 0.4511N$ and $k = 3,604,584,374$, the initial Singleton defect satisfies

$$N - k - d + 1 > 0.5489N - k > h.$$

We therefore put $K := N - d + 1 - h$ and choose a uniformly random K -dimensional supercode $C \supseteq C_0$. For every fixed $x \notin C_0$,

$$\Pr[x \in C] \leq 2q^{K-N} = 2q^{-d+1-h}.$$

There are at most $2^N q^{d-1}$ vectors of weight at most $d-1$, so

$$\Pr[d(C) < d] \leq 2^{N+1} q^{-h} \leq \frac{1}{2}.$$

Similarly, if either f_0 or g_0 is within Hamming distance b of C , then one of at most $2^{N+1} q^b$ corresponding vectors lies in C . Hence the probability of this event is at most

$$2^{N+2} q^{b-d+1-h} \leq q^{-1},$$

where we used

$$d - b - 1 > (0.4419 - 0.22945)N - 1 > 0.212N.$$

The two failure probabilities sum to less than one, so there exists such a supercode C satisfying

$$d(C) = d(C_0), \quad \delta(f_0, C), \delta(g_0, C) > 0.22945.$$

Moreover, C has dimension K and hence Singleton defect exactly h . Since $N = 2^{36}$,

$$h = \frac{N}{128} + 1, \quad R(C) = 1 - \delta(C) - \frac{1}{128}.$$

The distance bounds therefore imply

$$0.5410875 < R(C) < 0.5502875.$$

Since $C_0 \subseteq C$, every coefficient exceptional for C_0 remains exceptional for C . Relabeling f_0, g_0 as f, g , we have

$$|\mathcal{Z}_{0.1816}(f, g; C)| \geq B = 18,049,720,589,484,545 > 2^{54}.$$

Finally,

$$J_{3/2}(\delta(C)) < J_{3/2}(0.4511) < 0.1816 < 0.22095 = \frac{0.4419}{2} < \frac{\delta(C)}{2}.$$

Thus

$$\frac{\delta(C)}{2} - 0.1816 > 0.03935.$$

Since $B > 2^{54}$ and $N = 2^{36}$, we also have

$$|\mathcal{Z}_{0.1816}(f, g; C)| > 2^{18}N.$$

□

Acknowledgments

The first author was hosted by the Technological Innovation, Entrepreneurship, and Strategic Management (TIES) Group at the MIT Sloan School of Management while this research was conducted and is grateful for their hospitality. The third author was at a16z crypto as a research intern while this research was conducted.

References

- [ABF26] Gal Arnon, Dan Boneh, and Giacomo Fenzi. Open problems in list decoding and correlated agreement. Cryptology ePrint Archive, Paper 2026/680, 2026.
- [ACFY24] Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. STIR: Reed-Solomon proximity testing with fewer queries. In Leonid Reyzin and Douglas Stebila, editors, *Advances in Cryptology - CRYPTO 2024 - 44th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2024, Proceedings, Part X*, volume 14929 of *Lecture Notes in Computer Science*, pages 380–413. Springer, 2024.
- [ACFY25] Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. WHIR: Reed-Solomon proximity testing with super-fast verification. In Serge Fehr and Pierre-Alain Fouque, editors, *Advances in Cryptology - EUROCRYPT 2025 - 44th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Madrid, Spain, May 4-8, 2025, Proceedings, Part IV*, volume 15604 of *Lecture Notes in Computer Science*, pages 214–243. Springer, 2025.
- [AD21] S. Adriaensen and L. Denaux. Small weight codewords of projective geometric codes. *Journal of Combinatorial Theory, Series A*, 180:105395, 2021.
- [Ano23] Anonymous. Anonymous reference, 2023.
- [Ano26] Anonymous. Anonymous reference, 2026.
- [BBHR18] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast Reed-Solomon interactive oracle proofs of proximity. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, Prague, Czech Republic, July 9-13, 2018*, volume 107 of *LIPIcs*, pages 14:1–14:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.
- [BCF⁺25] Martijn Brehm, Binyi Chen, Ben Fisch, Nicolas Resch, Ron D. Rothblum, and Hadas Zeilberger. Blaze: Fast snarks from interleaved RAA codes. In Serge Fehr and Pierre-Alain Fouque, editors, *Advances in Cryptology - EUROCRYPT 2025 - 44th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Madrid, Spain, May 4-8, 2025, Proceedings, Part IV*, volume 15604 of *Lecture Notes in Computer Science*, pages 123–152. Springer, 2025.
- [BCH⁺25] E. Ben-Sasson, D. Carmon, U. Haböck, S. Kopparty, and S. Saraf. On proximity gaps for Reed–Solomon codes. Cryptology ePrint Archive, Paper 2025/2055, 2025.
- [BCI⁺23] E. Ben-Sasson, D. Carmon, Y. Ishai, S. Kopparty, and S. Saraf. Proximity gaps for Reed–Solomon codes. *Journal of the ACM*, 70(5):31:1–31:57, 2023.
- [BGKS20] E. Ben-Sasson, L. Goldberg, S. Kopparty, and S. Saraf. DEEP-FRI: Sampling outside the box improves soundness. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *LIPIcs*, pages 5:1–5:32, 2020.
- [DG24] B. E. Diamond and A. Gruen. Proximity gaps in interleaved codes. *IACR Communications in Cryptology*, 1(4):8, 2024.

- [DG25] B. E. Diamond and A. Gruen. On the distribution of the distances of random words. Cryptology ePrint Archive, Paper 2025/2010, 2025.
- [GG25] R. Goyal and V. Guruswami. Optimal proximity gaps for subspace-design codes and (random) Reed–Solomon codes. ECCC TR25-166, 2025.
- [GGSW26] R. Goyal, V. Guruswami, Y. Sun, and M. Wootters. Locality of curve-decoding and improved proximity gaps. arXiv:2607.08516, 2026.
- [GKL24] Y. Gao, H. Kan, and Y. Li. Linear proximity gap for linear codes within the 1.5 Johnson bound. Cryptology ePrint Archive, Paper 2024/1810, 2024.
- [Ham68] Noboru Hamada. The rank of the incidence matrix of points and d -flats in finite geometries. *Journal of Science of the Hiroshima University, Series A-I (Mathematics)*, 32(2):381–396, 1968.
- [Kam26] A. Kambiré. Proximity gaps conjecture fails near capacity over prime fields. arXiv:2604.09724, 2026.
- [KKH26] Dmitry Krachun, Stepan Kazanin, and Ulrich Haböck. Failure of proximity gaps close to capacity. *IACR Cryptol. ePrint Arch.*, 2026:782, 2026.
- [XZS22] T. Xie, Y. Zhang, and D. Song. Orion: Zero knowledge proof with linear prover time. In *Advances in Cryptology—CRYPTO 2022, Part IV*, pages 299–328, 2022.
- [ZCF24] H. Zeilberger, B. Chen, and B. Fisch. Basefold: Efficient field-agnostic polynomial commitment schemes from foldable codes. In *Advances in Cryptology—CRYPTO 2024, Part X*, pages 138–169, 2024.
- [Zei24] H. Zeilberger. Khatam: Reducing the communication complexity of code-based SNARKs. Cryptology ePrint Archive, Paper 2024/1843, 2024.