On the Distribution of the Distances of Random Words

Benjamin E. Diamond
Unaffiliated
benediamond@gmail.com

Angus Gruen
Zero Latency Labs
angusgruen@gmail.com

Abstract

For each positive integer c^*, we construct an infinite sequence of Reed–Solomon codes $C \subset \mathbb{F}_q^n$, together with ball radii z, for which the proportion of $\mathbb{F}_q^n$ collectively covered by the radius-z Hamming balls decays asymptotically more slowly than $\frac{n^{c^*}}{q}$ does. To pinpoint this decay rate, we develop various new, sharp combinatorial estimates, pertaining to the volumes of balls and their intersections.

Our result proves that the *capacity conjecture* of Ben-Sasson, Carmon, Ishai, Kopparty and Saraf (J. ACM '23) is false. Our code families' *relative* rates converge to 0 and their *relative* radii converge to 1. We suggest avenues by the means of which the capacity conjecture might be resuscitated; roughly, we suggest that that conjecture be restricted to the case of families whose relative rates are bounded from below by a positive constant. Our work shows that many deployed SNARKs may be less secure than they were formerly—optimistically—assumed to be.

1 Introduction

We show that Ben-Sasson, Carmon, Ishai, Kopparty and Saraf's *capacity conjecture* [BCIKS23, Conj. 8.4] is false. Our techniques are mainly combinatorial. We introduce a new, extremely sharp estimate of the volume of the Hamming ball $\mathcal{B}_q(\mathbf{0}, z) \subset \mathbb{F}_q^n$ of radius $z \in \{1, \dots, n-1\}$ (see Theorem 3.10). For certain n, q and z, our estimate becomes sharper than the classical, entropy-based one (see Example 3.11 below). More importantly, our estimate is *controllable*, in the sense that the ratio between it and the ground truth can be driven arbitrarily close to 1, at least for suitably chosen n, q and z. We also study closely the mass covered by those points that lie in the intersections of two or more Hamming balls (see Theorem 3.16). Using these estimates, we bound from below the proportion of $\mathbb{F}_q^n$ covered by the union of *all* Hamming balls $\mathcal{B}_q(\mathbf{v}, z) \subset \mathbb{F}_q^n$, as $\mathbf{v} \in C$ varies. In fact, this quantity—that is, the probability $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z]$ that a random word is z-close to the code, where the radius $z \in \{0, \dots, n-k-1\}$ is less than C's covering radius—itself bounds from below C's proximity error $\mathsf{err}(C, z)$, as we show in Theorem 2.5.

Our result applies to all Reed–Solomon codes, over all characteristics, and all domains. In fact, we use *only* the property of Reed–Solomon codes whereby for each n, each prime power $q \geq n$, and each message length $k \in \{1, \dots, n\}$, an $[n, k, d]$-code $C \subset \mathbb{F}_q^n$ exists, and is MDS. Our result applies to all code families for which this property is true.

For each positive integer c^*, for n growing to infinity, we construct an infinite family of Reed–Solomon codes $C \subset \mathbb{F}_q^n$—or more properly, of prime powers $q \geq n$ and message lengths $k \in \{1, \dots, n\}$—such that for infinitely many n, there exists a radius $z \in \{0, \dots, n-k-1\}$ such that $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] > \frac{n^{c^*}}{q}$ (and in fact this bound necessarily holds for *each* MDS $[n, k, d]$-code $C \subset \mathbb{F}_q^n$). As we explain below, this result implies that [BCIKS23, Conj. 8.4] is false as written. In our infinite families, both our message lengths k and our differences $h \coloneqq n - k - z$ are unbounded. (We call h—that is, the difference between the covering radius and the ball radius—the *shortfall*.) That is, both k and h grow arbitrarily large as n grows. This fact shows that [BCIKS23, Conj. 8.4] can't be resuscitated *merely* by the aid of a provision that forbids families for which k and h are constant. On the other hand, our families feature both $k = o(n)$ and $h = o(n)$. That is, our *relative* rates and our *relative* shortfalls go to 0. In fact, our techniques seem to work in general only when k and h grow like $o(\sqrt{n})$, a fact we discuss further below (see e.g. Remark 4.2).

In Section 5 below, we discuss how [BCIKS23, Conj. 8.4] might be salvaged, and propose a concrete amendment. Our amendment is conservative; it insists on treating only code families whose *relative* rates and *relative* shortfalls are bounded from below by positive constants. In our suggested amendment (see Conjecture 5.1), we predict an error bound of the shape $\mathsf{err}(C, z) = K \cdot \frac{n^c}{q}$; on the other hand, our constants K and c *depend* on ρ and η, and aren't universal. We also stipulate that the conjecture's prediction kick in only for those n larger than some unspecified constant, a measure that protects our amendment against the results of this paper (see also Remark 5.2).

While this work is primarily asymptotic, it does indicate that even certain SNARK parameterizations currently in widespread practical use may be concretely less secure than they were previously thought to be. We discuss this fact, with an eye towards concrete security, in Subsection 1.5.

We reproduce [BCIKS23, Conj. 8.4] here.

Conjecture (see Conjecture 2.2)**.** There exist constants c_1 and c_2 such that for each Reed–Solomon code $C = \mathsf{RS}_{\mathbb{F}_q}[n,k]$ and each proximity parameter $z \in \{0,\dots,n-k-1\}$, C features proximity gaps with respect to z and the false witness bound

$$\mathsf{err}(C,z) \leq \frac{1}{q} \cdot \frac{n^{c_2}}{(\rho \cdot \eta)^{c_1}},$$

where we write $\rho \coloneqq \frac{k}{n}$ and $\eta \coloneqq \frac{n-k-z}{n}$.

1.1 Heuristic Strategy

We work in the absolute setting; that is, we work with absolute distances and radii, as opposed to relative ones. We write $h \coloneqq n-k-z$ for the absolute difference between the covering radius and the ball radius; we call this quantity the radius's *shortfall*. Our k and h can be viewed as absolute analogues of the above conjecture's ρ and η.

Our goal is to study the distribution of random words, and in particular the probability that a randomly chosen word $\mathbf{u}$ lies z-close to the $[n,k,d]$-code $C \subset \mathbb{F}_q^n$. It's easy to show that for each (absolute) proximity parameter $z \in \{0,\dots,n-k-1\}$, the proximity-gap error cannot be much *less* than the probability that a uniformly random word is z-close to the code. In fact, as Theorem 2.5 below shows, we have

$$\mathsf{err}(C,z) \geq \frac{q-1}{q} \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u},C) \leq z].$$

We suppress the factor of $\frac{q-1}{q}$ above—which doesn't materially affect our analysis—throughout the rest of this expository segment.

The capacity conjecture predicts precisely that the left-hand side above should be bounded from above by

$$\mathsf{err}(C,z) \leq \frac{n^{2c_1+c_2}}{q \cdot (k \cdot h)^{c_1}} \leq \frac{n^{c^*}}{q},$$

where we write $c^* \coloneqq 2c_1 + c_2$.

Our goal is to show that this is impossible. Explicitly, for each choice of c^*, we construct an infinite family of codes $C \subset \mathbb{F}_q^n$, whose block lengths n grow arbitrarily large, such that, for each code C, there exists a proximity parameter $z \in \{0,\dots,n-k-1\}$, satisfying

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u},C) \leq z] > \frac{n^{c^*}}{q}.$$

This implies that the capacity conjecture, as currently stated, must be false. It should however be noted that we will be working with rather unrealistic codes. Hence, it is entirely consistent with our results that the capacity conjecture might be preserved in some form which more heavily restricts η. In Section 5 of this paper, we discuss this possibility further (see in particular Conjecture 5.1).

1.2 An Illustrative Example

Let us begin by considering the simplest case, in which $k = 1$ (so that $\mathsf{RS}_{\mathbb{F}_q}[n,k]$ is the q-ary *repetition code* of length n) and $h = 1$ (meaning the proximity parameter $z = n-k-1 = n-2$ is one less than capacity). In this case, the probability that a random word $\mathbf{u} \leftarrow \mathbb{F}_q^n$ is z-close to the code is exactly the probability that at least two of $\mathbf{u}$'s entries are identical. This is precisely the famous *birthday paradox*, and it is well known that the probability in question looks like $\frac{n^2}{q}$.

We record a suggestive explanation of how one might obtain this estimate. As we have just seen, the probability that a birthday collision occurs is exactly the probability that a random word $\mathbf{u} \leftarrow \mathbb{F}_q^n$ lies within *some* radius-$n-2$ Hamming ball in $\mathsf{RS}_{\mathbb{F}_q}[n,1]$. In $\mathbb{F}_q^n$, the number of points in each Hamming ball of radius z is roughly equal to

$$|\mathcal{B}_q(\mathbf{0},z)| \approx e^{-\frac{n}{q}} \cdot \frac{n^{n-z}}{(n-z)!} \cdot q^z.$$

Plugging in $z = n-2$, we get:

$$|\mathcal{B}_q(\mathbf{0},n-2)| \approx e^{-\frac{n}{q}} \cdot \frac{n^2 \cdot q^{n-2}}{2}.$$

We now make the pedagogical assumption whereby the balls surrounding unequal codewords do not intersect. This assumption is of course false in general, but turns out to be realistic in some cases. Under this assumption, the probability that a random word lies in *some* (any) Hamming ball is equal to the number of codewords in the code times the probability that a random word lies in *just one* Hamming ball. For a general $[n, k, d]$-code $C \subset \mathbb{F}_q^n$, this gives us the stipulated equality

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] \doteq \frac{|C|}{q^n} \cdot |\mathcal{B}_q(\mathbf{0}, z)|.$$

We use the symbol $\doteq$ to signify an equality which is temporarily—pedagogically—assumed to be true, by fiat.

Plugging in $k = 1$ and $z = n - 2$, we get

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq n - 2] \approx e^{-\frac{n}{q}} \cdot \frac{n^2 \cdot q^{n-2}}{2} \cdot \frac{q}{q^n} = e^{-\frac{n}{q}} \cdot \frac{n^2}{2 \cdot q}.$$

For $n = 23$ and $q = 365$, this gives 0.68, which, though close, is noticeably higher than the true value, 0.507.[1]

More generally, for each k and each z, we get the estimate

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] \approx e^{-\frac{n}{q}} \cdot \frac{n^{n-z}}{(n-z)!} \cdot q^{z+k-n}. \tag{1}$$

Setting $z = n - k - 1$, we get

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq n - k - 1] \approx \frac{e^{-\frac{n}{q}}}{(k+1)!} \cdot \frac{n^{k+1}}{q}.$$

We are free to fix as $k \coloneqq c^*$ an arbitrary constant and drive n and q to infinity; doing exactly this yields our first heuristic counterexample.

In this example, the shortfall $h = 1$; in other words, the ball radius $z = n - k - 1$ is exactly one less than the code's covering radius. Below, we will see that this restriction can be relaxed.

1.3 Heuristically Handling Intersections

As mentioned, the obvious flaw of the above analysis is that it ignores intersections between balls. There turns out to be an instructive way to handle these, at least heuristically. Let's look again at our estimate for the volume of the radius-z Hamming ball:

$$|\mathcal{B}_q(\mathbf{0}, z)| \approx e^{-\frac{n}{q}} \cdot \frac{n^{n-z}}{(n-z)!} \cdot q^z.$$

For $\mathbf{u} \leftarrow \mathbb{F}_q^n$ chosen uniformly randomly and $\mathbf{v} \in C$ a codeword, we write $X_{\mathbf{v}}(\mathbf{u})$ for the random variable that's 1 if $\mathbf{u} \in \mathcal{B}_q(\mathbf{v}, z)$ and 0 otherwise. Using our above ball volume estimate, we get in this setting:

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[X_{\mathbf{0}}(\mathbf{u}) = 1] = \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[\mathbf{u} \in \mathcal{B}_q(\mathbf{0}, z)] = \frac{|\mathcal{B}_q(\mathbf{0}, z)|}{q^n} \approx \frac{e^{-\frac{n}{q}} \cdot \left(\frac{n}{q}\right)^{n-z}}{(n-z)!},$$

a quantity we call $\mathscr{P}(n, q, z)$.

By definition of $d(\mathbf{u}, C)$, we also see that

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] = \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[\exists\ \mathbf{v} \in C : X_{\mathbf{v}}(\mathbf{u}) = 1].$$

Our initial assumption above, whereby each pair of balls is disjoint, is equivalent to that whereby:

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[\exists\ \mathbf{v} \in C : X_{\mathbf{v}}(\mathbf{u}) = 1] \doteq \sum_{\mathbf{v} \in C} \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[X_{\mathbf{v}}(\mathbf{u}) = 1] = |C| \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[X_{\mathbf{0}}(\mathbf{u}) = 1] \approx |C| \cdot \mathscr{P}(n, q, z).$$

Instead of assuming disjointness, we now assume rather that the variables $X_{\mathbf{v}}$ are *independent*, an incomparable assumption. While this assumption is of course also false (it's impossible for a word to be in all balls at once, for example), it turns out to be much closer to the truth. Under exactly this assumption, we obtain:

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[\exists\ \mathbf{v} \in C : X_{\mathbf{v}}(\mathbf{u}) = 1] = 1 - \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[\forall\ \mathbf{v} \in C : X_{\mathbf{v}}(\mathbf{u}) = 0]$$

[1] This estimate improves as $\frac{n}{q}$ decreases. In the birthday example, $\frac{23}{365} \approx 0.06$ is large enough that certain higher-order terms—suppressed by our disjointness assumption—contribute material corrections to the total; we discuss these further in Section 1.3.

$$
\begin{aligned}
&\doteq 1 - \left(\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n} [X_{\mathbf{0}}(\mathbf{u}) = 0] \right)^{|C|} \\
&= 1 - \left(1 - \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n} [X_{\mathbf{0}}(\mathbf{u}) = 1] \right)^{|C|} \\
&\approx 1 - (1 - \mathscr{P}(n,q,z))^{|C|} \\
&= |C| \cdot \mathscr{P}(n,q,z) - \frac{|C|^2}{2} \cdot \mathscr{P}(n,q,z)^2 + \frac{|C|^3}{3!} \cdot \mathscr{P}(n,q,z)^3 - \cdots .
\end{aligned}
$$

The first term of this series is exactly our previous estimate; its subsequent terms yield corrections that respectively correspond to double-intersections, triple-intersections and so on. Let's try this new approximation out on the birthday problem. We again set $|C| = q = 365$, $n = 23$, and $z = n - 2 = 21$. In this way, we get:

Number of terms considered	1	2	3	4	All
Estimate	0.680	0.449	0.501	0.492	≈ 0.494

This approximation rapidly converges to something reasonably close to the true value 0.507. Interestingly, the fault in this approximation lies less in its independence assumption than in our ball estimate (upon swapping in the true ball volume, we would obtain an extremely accurate estimate).

Throughout the remainder of this paper, we prove that this second-order truncation *actually* yields a provable lower-bound for $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z]$, at least for certain, judiciously chosen n, q, z and k. In particular, using $|C| = q^k$, we get

$$
\begin{aligned}
\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n} [d(\mathbf{u}, C) \leq z] &\geq |C| \cdot \mathscr{P}(n,q,z) - \frac{|C|^2}{2} \cdot \mathscr{P}(n,q,z)^2 \\
&\geq q^k \cdot \frac{e^{-\frac{n}{q}} \cdot n^{n-z} \cdot q^{z-n}}{(n-z)!} - \frac{q^{2k}}{2} \cdot \frac{e^{-\frac{2n}{q}} \cdot n^{2n-2z} \cdot q^{2z-2n}}{(n-z)!^2},
\end{aligned}
$$

at least for those q, n and z for which our ball and intersection volume estimates become sufficiently accurate. We postulate that including higher-order terms should further improve this estimate.

1.4 The Capacity Conjecture

What does the above estimate—assuming for now that it's accurate—tell us about the capacity conjecture? That conjecture implies that, for some unspecified fixed integer $c^* \geq 1$, the following inequality holds:

$$
\frac{n^{c^*}}{q} \geq \mathsf{err}(C,z) \geq \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n} [d(\mathbf{u}, C) \leq z] \geq q^k \cdot \frac{e^{-\frac{n}{q}} \cdot n^{n-z} \cdot q^{z-n}}{(n-z)!} - \frac{q^{2k}}{2} \cdot \frac{e^{-\frac{2n}{q}} \cdot n^{2n-2z} \cdot q^{2z-2n}}{(n-z)!^2}.
$$

We want to show—for as wide a range of radii z as we can—that this can't be. We assume during this heuristic exposition that $n \ll q$; this assumption lets us ignore both exponential factors above (these will both be effectively 1). As before, we write $h = n - k - z$ for the *shortfall* attached to the radius z. We moreover fix $q = n^{c^*+1}$ for good (this quantity isn't a prime power in general; we deal with this technicality below). Plugging these choices into the bound above, we rephrase it as:

$$
\frac{1}{n} \geq \frac{n^{k+h-(c^*+1)h}}{(k+h)!} - \frac{n^{2(k+h)-2(c^*+1)h}}{2 \cdot (k+h)!^2}.
$$

Provided $0 \leq \frac{n^{k+h-(c^*+1)h}}{(k+h)!} \ll 1$, we can ignore the square term; we thus drop it for now.

To make our point here more clearly, instead of working with k and h directly, we work rather with tunable parameters that control these values (i.e., as predictable functions of n). First, we assume that k and h themselves differ multiplicatively by a fixed constant; we write A for that constant. Thus, we have $k = A \cdot h$. (Below, we deal rigorously with the matter of whether these quantities are really integers.) Since $k + h = n - z$, the sum $k + h$ equals the difference between the block length and the ball radius. We see that A represents the *relative proportions* of the complement $n - z$ occupied, respectively, by the message length and the shortfall.

Having fixed this relationship, we apply Stirling's approximation to the above inequality. We obtain in this way:

$$\frac{1}{n} \geq \frac{n^{(A-c^*)h}}{((A+1)h)!} \geq \frac{e^{(A+1)h} \cdot n^{(A-c^*)h}}{((A+1)h)^{(A+1)h} \cdot \sqrt{2\pi(A+1)h} \cdot e^{\frac{1}{12(A+1)h}}}.$$

We now further assume that $k+h = (A+1) \cdot h$ is itself a small fraction of n. That is, we fix some $\varepsilon > 0$, and assume that $(A+1) \cdot h = \varepsilon \cdot n$. Subbing this in, we get

$$\frac{1}{n} \geq \frac{e^{\varepsilon n} \cdot n^{\frac{A-c^*}{A+1} \cdot \varepsilon n}}{(\varepsilon n)^{\varepsilon n} \cdot \sqrt{2\pi \cdot \varepsilon n} \cdot e^{\frac{1}{12\varepsilon n}}}.$$

Now, we take base-n logarithms of both sides. In this way, we further reëxpress the relationship at hand as follows:

$$-1 \geq \varepsilon n \cdot \left(\log_n(e) + \frac{A-c^*}{A+1} - \log_n(\varepsilon) - 1 \right) - \frac{1 + \log_n(2\pi\varepsilon)}{2} - \frac{1}{12\varepsilon n} \cdot \log_n(e).$$

We've just shown that, for each n, c^* and ε, if we can find an A such that the above inequality becomes *false*, then we've—at least, heuristically—disproven the conjecture. But this is not hard to do. Setting the right-hand side of the inequality just above to $-\frac{1}{2}$ (in order to block out some wiggle room) and solving for A, we find that the unique solution

$$A \coloneqq \frac{c^* + E(\varepsilon, n)}{1 - E(\varepsilon, n)} \approx \frac{c^* + 1}{\log_n\left(\frac{e}{\varepsilon}\right)};$$

suffices; here, we use the auxiliary function

$$E(\varepsilon, n) \coloneqq \frac{\log_n(2\pi\varepsilon)}{2\varepsilon n} + \frac{\log_n(e)}{12(\varepsilon n)^2} - \log_n\left(\frac{e}{\varepsilon}\right) + 1.$$

We note in passing that Stirling's approximation yields an upper-bound too, which in fact differs from the lower-bound given above solely in that it *lacks* the $e^{\frac{1}{12\varepsilon n}}$ factor present in that lower-bound's denominator. This factor is extremely close to 1 for large enough n. Our step above in which we ignore the square term is thus sensible, as this Stirling-based upper-bound of the above quantity goes to 0 just as the lower-bound does.

This sketch suggests that the capacity conjecture must fail to hold whenever the above estimates are valid.

1.5 Concrete Failure Cases

While we primarily focus on—and *prove* our estimate accurate on—fairly unrealistic codes (whose block lengths n are large and their relative rates ρ small), it is worth noting that our approach hints that the capacity conjecture likely fails even in settings of real-world interest.

In particular, an illuminating question to ask is, for which values of q, n, k and z is

$$\mathsf{err}(C, z) \approx \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] \approx 1?$$

Particularly useful here is the simple estimate (1) above, which we reproduce here:

$$\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] \approx e^{-\frac{n}{q}} \cdot \frac{n^{n-z}}{(n-z)!} \cdot q^{z+k-n} = e^{-\frac{n}{q}} \cdot \frac{n^{k+h}}{(k+h)!} \cdot q^{-h};$$

as usual, we write $h \coloneqq n - k - z$.

While this estimate of course can't be strictly accurate when its right-hand side is greater than 1, it's reasonable to assume, in each such case, that its left-hand side also equals—or at least is very close to—1. Let's work in a realistic setting, for which in particular $k \gg h$ holds, and $\rho = \frac{k}{n}$ is fixed. In exactly this setup, applying Stirling's approximation again, and dropping the exponential factor, we obtain for the right-hand side above the approximation

$$\left(\frac{en}{k+h}\right)^{k+h} \cdot q^{-h} \approx \left(\frac{e}{\rho}\right)^{k+h} \cdot q^{-h}.$$

Taking base-2 logarithms and setting the right-hand side above to 0, we get:

$$0 = (k+h) \cdot \log_2\left(\frac{e}{\rho}\right) - h \cdot \log_2(q). \tag{2}$$

Upon solving for h and relativizing, we get:

$$\eta \approx \frac{\log_2(\frac{e}{\rho}) \cdot \rho}{\log_2(q)}. \tag{3}$$

Each proximity parameter whose distance away from $1 - \rho$ falls beneath this very η is liable to induce ball radii that all-but fill the space, and so cause proximity gaps to fail, as the calculation we just carried out suggests.

For each $\rho \in (0, 1)$ and each $\eta \in (0, 1 - \rho)$ for which proximity gaps are in force, to obtain b bits of security,

$$\frac{-b}{\log_2(\rho + \eta)} \tag{4}$$

queries must be made. Substituting (3) into (4), we find that, to get b bits of security, one must make

$$\frac{-b}{\log_2\left(\rho + \frac{\log_2(\frac{e}{\rho}) \cdot \rho}{\log_2(q)}\right)} \tag{5}$$

queries.

We tabulate various concrete values of (5), as well as of their respective counterparts in other regimes, in Table 1 below. For the purposes of (5) above, we assume that $\log_2(q) = 128$.

Regime	Unique Decoding	List Decoding	Random Words	Capacity
Queries Required	$\frac{-100}{\log_2(\frac{1+\rho}{2})}$	$\frac{-100}{\log_2(\sqrt{\rho})}$	$\frac{-100}{\log_2\left(\rho + \frac{\log_2(\frac{e}{\rho}) \cdot \rho}{\log_2(q)}\right)}$	$\frac{-100}{\log_2(\rho)}$
$\rho = 1/2$	240.94	200.00	102.80	100.00
$\rho = 1/4$	147.48	100.00	50.98	50.00
$\rho = 1/8$	120.47	66.67	33.89	33.33
$\rho = 1/16$	109.58	50.00	25.38	25.00
$\rho = 1/32$	104.65	40.00	20.29	20.00
$\rho = 1/64$	102.29	33.33	16.90	16.67

Table 1: The number of queries necessary to obtain 100 bits of security, in various regimes.

For example, in the case $\rho = \frac{1}{2}$, the capacity conjecture implies that, to obtain 100 bits of security without grinding, 100 queries should suffice; our results suggest that *at least* $102.80 \approx 103$ queries are necessary in this case.

Indeed, we do not intend to insinuate here that the region *between* the list decoding and random words regimes is safe to use, and highly recommend the use of proven—over conjectural—bounds in real deployments. However, we counsel those who insist on using conjectures to stay above the *random words* cutoff (3).

Acknowledgements. We would like to thank Giacomo Fenzi and Benedikt Wagner for first posing to us the problem that ultimately grew into this work. We gratefully thank Justin Drake, Will Corcoran and the Ethereum Foundation for inviting us to a delightful workshop in Cambridge, England during whose course this work began.

2 Background and Notation

We roughly follow the notation of Guruswami [Gur06].

2.1 Multinomial Coefficients

We use *multinomial coefficients* throughout. We use the notational convention

$$\binom{n}{a_1, \ldots, a_t} \coloneqq \frac{n!}{a_1! \cdot \cdots \cdot a_t! \cdot (n - a_1 - \cdots - a_t)!}.$$

That is, the quantities in the coefficient's lower part need not add to n. We automatically add the "missing part". When any among the factorials above is negative, the entire coefficient is defined to be zero. We note that our convention smoothly extends the classic binomial case (i.e., in which $t = 1$).

2.2 Linear Codes

Fix a finite field $\mathbb{F}_q$. The *disagreement set* between words $\mathbf{u}$ and $\mathbf{u}'$ in $\mathbb{F}_q^n$ is $\Delta(\mathbf{u}, \mathbf{u}') \coloneqq \{i \in \{0, \ldots, n-1\} \mid \mathbf{u}_i \neq \mathbf{u}'_i\}$. The *Hamming distance* between $\mathbf{u}$ and $\mathbf{u}'$ is $d(\mathbf{u}, \mathbf{u}') \coloneqq |\Delta(\mathbf{u}, \mathbf{u}')|$. A linear $[n, k, d]$-*code* over $\mathbb{F}_q$ is a k-dimensional linear subspace $C \subset \mathbb{F}_q^n$ such that, for each pair of distinct elements $\mathbf{v}$ and $\mathbf{v}'$ in C, $d(\mathbf{v}, \mathbf{v}') \geq d$. The code is said to be *MDS* if $d = n - k + 1$. For $\mathbf{u} \in \mathbb{F}_q^n$ arbitrary, we write $d(\mathbf{u}, C) \coloneqq \min_{\mathbf{v} \in C} d(\mathbf{u}, \mathbf{v})$ for the *distance* between $\mathbf{u}$ and the code C. For each $\mathbf{v} \in C$ and each radius z, the *Hamming ball* of radius z centered at $\mathbf{v}$ is the set $\mathcal{B}_q(\mathbf{v}, z) \coloneqq \left\{\mathbf{u} \in \mathbb{F}_q^n \;\middle|\; d(\mathbf{u}, \mathbf{v}) \leq z\right\}$.

We recall Reed–Solomon codes (see e.g. [Gur06, Def. 2.3]). Fix nonnegative *message length* and *block length* parameters k and n, as well as a subset $S \subset \mathbb{F}_q$ of size n. We write $C \subset \mathbb{F}_q^n$ for the *Reed–Solomon* code $\mathsf{RS}_{\mathbb{F}_q, S}[n, k] \coloneqq \left\{\left(P(x)\right)_{x \in S} \;\middle|\; P(X) \in \mathbb{F}_q[X]^{\prec k}\right\}$. That is, $\mathsf{RS}_{\mathbb{F}_q, S}[n, k]$ is the set of n-tuples which arise as the evaluations of some polynomial $P(X)$ of degree less than k on the set S. Reed–Solomon codes are MDS.

The *covering radius* of a code C is the minimal radius R such that for each word $\mathbf{u} \in \mathbb{F}_q^n$, $d(\mathbf{u}, C) \leq R$; in other words, $R \coloneqq \max_{\mathbf{u} \in \mathbb{F}_q^n} d(\mathbf{u}, C)$. A word $\mathbf{u} \in \mathbb{F}_q^n$ is a *deep hole* of C if $d(\mathbf{u}, C) = R$ (i.e., if $\mathbf{u}$ attains the covering radius). The covering radius of $\mathsf{RS}_{\mathbb{F}_q}[n, k]$ is $n - k$. We refer to Cohen, Honkala, Litsyn and Lobstein [CHLL97] for much more on covering codes.

2.3 Proximity Gaps

The following definition comes from Ben-Sasson, Carmon, Ishai, Kopparty and Saraf [BCIKS23], and is also implicit in previous works (see e.g. Ames, Hazay, Ishai and Venkitasubramaniam [AHIV23]). We fix a field $\mathbb{F}_q$ and an arbitrary linear $[n, k, d]$-code $C \subset \mathbb{F}_q^n$.

Definition 2.1. The code $C \subset \mathbb{F}_q^n$ features *proximity gaps* with respect to the proximity parameter $z \in \{0, \ldots, n\}$ and the false witness bound $\mathsf{err}(C, z)$ if, for each pair of words $\mathbf{u}$ and $\mathbf{u}'$ in $\mathbb{F}_q^n$, if

$$\Pr_{r \leftarrow \mathbb{F}_q}[d(\mathbf{u} + r \cdot \mathbf{u}', C) \leq z] > \mathsf{err}(C, z),$$

then there exist codewords $\mathbf{v}$ and $\mathbf{v}'$ in C and a set $D \subset \{0, \ldots, n-1\}$ of size less than or equal to z such that, for each $i \in \{0, \ldots, n-1\} \setminus D$, $\mathbf{u}_i = \mathbf{v}_i$ and $\mathbf{u}'_i = \mathbf{v}'_i$.

In case the conclusion of Definition 2.1 holds, we say that the *interleaved distance* between the pair $(\mathbf{u}, \mathbf{u}')$ and C is at most z.

Conjecture 2.2 (Ben-Sasson–Carmon–Ishai–Kopparty–Saraf [BCIKS23, Conj. 8.4])**.** *There exist constants c_1 and c_2 such that for each Reed–Solomon code $C = \mathsf{RS}_{\mathbb{F}_q}[n, k]$ and each proximity parameter $z \in \{0, \ldots, n-k-1\}$, C features proximity gaps with respect to z and the false witness bound*

$$\mathsf{err}(C, z) \leq \frac{1}{q} \cdot \frac{n^{c_2}}{(\rho \cdot \eta)^{c_1}}, \tag{6}$$

where we write $\rho \coloneqq \frac{k}{n}$ and $\eta \coloneqq \frac{n-k-z}{n}$.

Remark 2.3. Conjecture 2.2 differs nominally from [BCIKS23, Conj. 8.4] in that we ask that η be a positive integer multiple of $\frac{1}{n}$. These versions are equivalent. Indeed, [BCIKS23, Conj. 8.4] trivially implies Conjecture 2.2—the important direction for this work—since, if (6) holds for all $\eta > 0$, it certainly also holds for those which are positive integer multiples of $\frac{1}{n}$. In the reverse direction, we note that for each $z \in \{0, \ldots, n-k-1\}$, for $\eta \in \left(\frac{n-k-z-1}{n}, \frac{n-k-z}{n}\right]$, $\frac{1}{\eta^{c_1}}$ is minimized at the right-hand closed endpoint $\frac{n-k-z}{n}$. Thus, if (6) holds at $\eta = \frac{n-k-z}{n}$, then shrinking η by less than $\frac{1}{n}$ can't cause (6) to become false (doing this will only make $\frac{1}{\eta^{c_1}}$ larger).

Remark 2.4. Before we launch into our treatment proper, we point out something odd about this conjecture, which is already implicit in Subsection 1.1. Its authors seem to have included the denominator $(\rho \cdot \eta)^{c_1}$ in order to widen the margin of error for those codes for which ρ and η become too small, i.e. by increasing the error-probability threshold that those codes must surmount, if they are to induce a contravention of the conjecture's claim.

On the other hand, even in the most extreme possible case, $k = 1$ and $z = n - k - 1$ both hold. In this case, $\rho = \eta = \frac{1}{n}$, so that $\frac{1}{(\rho \cdot \eta)^{c_1}} = n^{2c_1}$. We see that Conjecture 2.2's denominator contributes, at most, a further polynomial factor to $\mathsf{err}(C, z)$, a slackening that could equally have been achieved with recourse instead to a boost to c_2. The conjecture is thus true if and only if it's true for $c_1 = 0$. While that denominator—if the conjecture were true—might have been useful for the purposes of delivering sharper *concrete* bounds in specific circumstances, it was incapable of affecting the conjecture's truth value. For further discussion of this topic, we refer to Section 5 below.

2.4 Close Word Probability

In this subsection, we note that the false witness probability $\mathsf{err}(C, z)$ above is related to the *close word probability* $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z]$. To do this, we use an averaging argument. Roughly, if the union of the radius-z Hamming balls fills too much of $\mathbb{F}_q^n$, then this mass must "hit" at least one affine line in a commensurate way.

In fact, this relationship holds for general linear codes. We fix a field $\mathbb{F}_q$ and an $[n, k, d]$-code $C \subset \mathbb{F}_q^n$, and write $R \in \{0, \dots, n\}$ for the covering radius of C.

Theorem 2.5. *If C features proximity gaps with respect to the proximity parameter $z \in \{0, \dots, R-1\}$ and the false witness bound $\mathsf{err}(C, z)$, then*

$$\mathsf{err}(C, z) \ge \frac{q-1}{q} \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z].$$

Proof. We fix $C \subset \mathbb{F}_q^n$ as in the hypothesis; that is, we assume that $C \subset \mathbb{F}_q^n$ features proximity gaps with respect to $z \in \{0, \dots, R-1\}$ and $\mathsf{err}(C, z)$. We say that a word $\mathbf{u} \in \mathbb{F}_q^n$ is *close* if $d(\mathbf{u}, C) \le z$.

Let $\mathbf{u}^* \in \mathbb{F}_q^n$ be a deep hole of C, so that $d(\mathbf{u}^*, C) = R$. Since $z < R$ by fiat, $\mathbf{u}^*$ itself is not close, and the total number of close words $\mathbf{u} \in \mathbb{F}_q^n \setminus \{\mathbf{u}^*\}$ is exactly

$$q^n \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z].$$

Moreover, as $\mathbf{u}' \in \mathbb{F}_q^n \setminus \{\mathbf{0}\}$ varies, the *distinct* punctured lines of the form $L_{\mathbf{u}'} \coloneqq \{\mathbf{u}^* + r \cdot \mathbf{u}' \mid r \in \mathbb{F}_q\} \setminus \{\mathbf{u}^*\}$ exactly partition $\mathbb{F}_q^n \setminus \{\mathbf{u}^*\}$ into $\frac{q^n-1}{q-1}$ disjoint subsets, each of size $q-1$. By the pigeonhole principle, we can thus find some $\mathbf{u}^\dagger$ such that the punctured line $L_{\mathbf{u}^\dagger}$ contains at least

$$\frac{q-1}{q^n-1} \cdot q^n \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z]$$

close words; in other words,

$$(q-1) \cdot \Pr_{\mathbf{u} \leftarrow L_{\mathbf{u}^\dagger}}[d(\mathbf{u}, C) \le z] \ge \frac{q-1}{q^n-1} \cdot q^n \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z].$$

The interleaved distance of $(\mathbf{u}^*, \mathbf{u}^\dagger)$ from C is trivially at least R, by construction of $\mathbf{u}^*$. Again since $z < R$, by our hypothesis on C, the contraposition of Definition 2.1, applied to the pair $(\mathbf{u}^*, \mathbf{u}^\dagger)$, implies that $\Pr_{r \leftarrow \mathbb{F}_q}\left[d\left(\mathbf{u}^* + r \cdot \mathbf{u}^\dagger, C\right) \le z\right] \le \mathsf{err}(C, z)$, so that

$$(q-1) \cdot \Pr_{\mathbf{u} \leftarrow L_{\mathbf{u}^\dagger}}[d(\mathbf{u}, C) \le z] \le q \cdot \Pr_{r \leftarrow \mathbb{F}_q}\left[d\left(\mathbf{u}^* + r \cdot \mathbf{u}^\dagger, C\right) \le z\right] \le q \cdot \mathsf{err}(C, z).$$

To achieve the first inequality, we note that $L_{\mathbf{u}^\dagger} = \left\{\mathbf{u}^* + r \cdot \mathbf{u}^\dagger \;\middle|\; r \in \mathbb{F}_q\right\} \setminus \{\mathbf{u}^*\}$ has at most as many close points as $\left\{\mathbf{u}^* + r \cdot \mathbf{u}^\dagger \;\middle|\; r \in \mathbb{F}_q\right\}$ does. (In fact, this inequality happens to be an equality, since $\mathbf{u}^*$ is not close.)

Combining these results, we get:

$$\mathsf{err}(C, z) \ge \frac{q-1}{q} \cdot \Pr_{\mathbf{u} \leftarrow L_{\mathbf{u}^\dagger}}[d(\mathbf{u}, C) \le z] \ge \frac{(q-1) \cdot q^n}{q \cdot (q^n-1)} \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z] \ge \frac{q-1}{q} \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z],$$

which completes the proof. The last inequality is equivalent to $\frac{q^n}{q^n-1} \ge 1$. □

Remark 2.6. The hypothesis $z \in \{0, \dots, R-1\}$ of Theorem 2.5 is necessary to make that theorem make sense. If $z \ge R$, then while that theorem's statement remains technically coherent, it becomes false for a trivial reason. Indeed, at least if $C \subset \mathbb{F}_q^n$ is Reed–Solomon, the interleaved distance of *every* pair $(\mathbf{u}, \mathbf{u}')$ from the code is at most z (in fact, at most R). In exactly this setting, $\mathsf{err}(C, z)$ becomes vacuously 0; on the other hand, $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \le z] = 1$.

3 Combinatorial Estimates

In this section, we record various sharp combinatorial estimates for the volumes of Hamming balls (see Subsection 3.3) and of their intersections (see Subsection 3.4).

3.1 Asymptotics

During our calculations below, we use a slightly bespoke notational scheme. The idea is to formalize the notion whereby the ratio between two multivariate functions can be made arbitrarily close to 1, provided that the input be taken from a region of the parameter space where a certain "controller function" becomes small. Actually, we want to say a bit more. In our notion below, the controller function's smallness induces not just a ratio close to 1, but moreover one whose absolute deviation from 1 is bounded from above by a constant times the magnitude of the controller.

For $\mathbf{x} = (x_0, \ldots, x_{n-1}) \in \mathbb{R}^n$, we write $\|\mathbf{x}\| \coloneqq \sum_{i=0}^{n-1} |x_i|$ for its L^1 norm. We fix an integer $n \geq 1$, a set $S \subset \mathbb{R}^n$, and two functions $f(\mathbf{x})$ and $g(\mathbf{x})$ on S.

Definition 3.1. We say that $f(\mathbf{x}) = g(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$, where $\mathbf{a}(\mathbf{x})$ is a function $S \to \mathbb{R}^m$, if there is a constant $K \geq 0$ and a neighborhood $U \subset \mathbb{R}^m$ containing the origin, such that, whenever $\mathbf{x} \in S$ satisfies $\mathbf{a}(\mathbf{x}) \in U$, we have $g(\mathbf{x}) \neq 0$ and

$$\left| \frac{f(\mathbf{x})}{g(\mathbf{x})} - 1 \right| \leq K \cdot \|\mathbf{a}(\mathbf{x})\|.$$

Though the "equality" and the "product" in Definition 3.1 make for convenient syntactic sugar, they aren't really real. On the other hand, Definition 3.1 behaves well, as we now argue. For one, the notion is symmetric:

Lemma 3.2. *If $f(\mathbf{x}) = g(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$, then $g(\mathbf{x}) = f(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$.*

Proof. We write $U \subset \mathbb{R}^m$ for the neighborhood attached to $\mathbf{a}(\mathbf{x})$, by hypothesis on it, and $K \geq 0$ for its associated constant. Additionally, we denote $R(\mathbf{x}) \coloneqq \frac{f(\mathbf{x})}{g(\mathbf{x})}$.

Pick an $r > 0$ for which $K \cdot r < \frac{1}{2}$, and set $U' \coloneqq \{\mathbf{x} \in U \mid \|\mathbf{x}\| < r\}$. Whenever $\mathbf{a}(\mathbf{x}) \in U'$, our hypothesis on $f(\mathbf{x})$ and $g(\mathbf{x})$ implies that $R(\mathbf{x})$ is defined, and that $|R(\mathbf{x}) - 1| \leq K \cdot \|\mathbf{a}(\mathbf{x})\| \leq K \cdot r < \frac{1}{2}$, so that $R(\mathbf{x}) > \frac{1}{2}$. In particular, $R(\mathbf{x}) > 0$, so $f(\mathbf{x}) \neq 0$. Moreover, whenever $\mathbf{a}(\mathbf{x}) \in U'$,

$$\left| \frac{g(\mathbf{x})}{f(\mathbf{x})} - 1 \right| = \left| \frac{1}{R(\mathbf{x})} - 1 \right| = \frac{|1 - R(\mathbf{x})|}{|R(\mathbf{x})|} < 2 \cdot |R(\mathbf{x}) - 1| \leq 2 \cdot K \cdot \|\mathbf{a}(\mathbf{x})\|,$$

which completes the proof. □

We also get a key transitivity property. For controller functions $\mathbf{a} : S \to \mathbb{R}^m$ and $\mathbf{b} : S \to \mathbb{R}^p$, the expression $\mathbf{a} \parallel \mathbf{b} : S \to \mathbb{R}^{m+p}$ denotes the concatenation of $\mathbf{a}$ and $\mathbf{b}$.

Lemma 3.3. *If $f(\mathbf{x}) = g(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$ and $g(\mathbf{x}) = h(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{b}(\mathbf{x})))$, then $f(\mathbf{x}) = h(\mathbf{x}) \cdot (1 + \mathcal{O}((\mathbf{a} \parallel \mathbf{b})(\mathbf{x})))$.*

Proof. Write $U_{\mathbf{a}} \subset \mathbb{R}^m$ and $U_{\mathbf{b}} \subset \mathbb{R}^p$ for the neighborhoods attached to $\mathbf{a}$ and $\mathbf{b}$, by hypothesis on them, as well as $K_{\mathbf{a}}$ and $K_{\mathbf{b}}$ for the corresponding constants. We moreover write

$$p(\mathbf{x}) \coloneqq \frac{f(\mathbf{x})}{g(\mathbf{x})} - 1 \text{ and } q(\mathbf{x}) \coloneqq \frac{g(\mathbf{x})}{h(\mathbf{x})} - 1$$

and set

$$U' \coloneqq \{(\mathbf{y_a}, \mathbf{y_b}) \in U_{\mathbf{a}} \times U_{\mathbf{b}} \mid \|\mathbf{y_a}\| + \|\mathbf{y_b}\| < 1\}.$$

In particular, $\|\mathbf{y_a}\| \cdot \|\mathbf{y_b}\| \leq \|\mathbf{y_a}\| + \|\mathbf{y_b}\|$ holds identically on U'. For each $\mathbf{x}$ for which $\mathbf{a}(\mathbf{x}) \in U_{\mathbf{a}}$ and $\mathbf{b}(\mathbf{x}) \in U_{\mathbf{b}}$, our hypothesis implies that $p(\mathbf{x})$ and $q(\mathbf{x})$ are both defined, and that $|p(\mathbf{x})| \leq K_{\mathbf{a}} \cdot \|\mathbf{a}(\mathbf{x})\|$ and $|q(\mathbf{x})| \leq K_{\mathbf{b}} \cdot \|\mathbf{b}(\mathbf{x})\|$. For each such $\mathbf{x}$ we thus get:

$$\begin{aligned}
\left| \frac{f(\mathbf{x})}{h(\mathbf{x})} - 1 \right| &= |(1 + p(\mathbf{x})) \cdot (1 + q(\mathbf{x})) - 1| && \text{(just by definition.)} \\
&\leq K_{\mathbf{a}} \cdot \|\mathbf{a}(\mathbf{x})\| + K_{\mathbf{b}} \cdot \|\mathbf{b}(\mathbf{x})\| + K_{\mathbf{a}} \cdot K_{\mathbf{b}} \cdot \|\mathbf{a}(\mathbf{x})\| \cdot \|\mathbf{b}(\mathbf{x})\| && \text{(expand and use the triangle inequality.)} \\
&\leq K_{\mathbf{a}} \cdot \|\mathbf{a}(\mathbf{x})\| + K_{\mathbf{b}} \cdot \|\mathbf{b}(\mathbf{x})\| + K_{\mathbf{a}} \cdot K_{\mathbf{b}} \cdot (\|\mathbf{a}(\mathbf{x})\| + \|\mathbf{b}(\mathbf{x})\|). && \text{(using } \|\mathbf{y_a}\| \cdot \|\mathbf{y_b}\| \leq \|\mathbf{y_a}\| + \|\mathbf{y_b}\| \text{ on } U'.) \\
&\leq (K_{\mathbf{a}} + K_{\mathbf{b}} + K_{\mathbf{a}} \cdot K_{\mathbf{b}}) \cdot (\|\mathbf{a}(\mathbf{x})\| + \|\mathbf{b}(\mathbf{x})\|),
\end{aligned}$$

which proves what we want. □

The following "multiplicativity" lemma is essentially trivial, but useful.

Lemma 3.4. *If $f(\mathbf{x}) = g(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$, then for each nowhere-zero $h(\mathbf{x})$, $f(\mathbf{x}) \cdot h(\mathbf{x}) = g(\mathbf{x}) \cdot h(\mathbf{x}) \cdot (1 + \mathcal{O}(\mathbf{a}(\mathbf{x})))$.*

Proof. There's really nothing to prove, since $\frac{f(\mathbf{x})}{g(\mathbf{x})} = \frac{f(\mathbf{x}) \cdot h(\mathbf{x})}{g(\mathbf{x}) \cdot h(\mathbf{x})}$ identically on S. □

3.2 Basic Estimates

Using the notation we've just developed, we collect a few basic facts that we'll use below.

Definition 3.5. *Stirling's approximation* for the factorial is defined as

$$S(n) = \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n.$$

Lemma 3.6. *On the set of positive integers, $n!$ is well approximated by $S(n)$ in the sense that*

$$n! = S(n) \cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

We similarly get excellent approximations for the binomials $\binom{n}{j}$ using S. Though the following statement is "trivial", we prove it pedantically, as a warmup.

Corollary 3.7. *For each $n \geq 0$ and each $j \in \{0, \ldots, n-1\}$,*

$$\binom{n}{j} = \frac{S(n)}{j! \cdot S(n-j)} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}\right)\right).$$

Proof. The statement of the corollary is exactly equivalent to that whereby $\frac{n!}{(n-j)!} = \frac{S(n)}{S(n-j)} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}\right)\right)$ (the $j!$s cancel from numerator and denominator). Combining Lemmas 3.4 and 3.6, we get that $\frac{n!}{(n-j)!} = \frac{S(n)}{(n-j)!} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right)$. On the other hand, applying Lemmas 3.2, 3.4 and 3.6 to $(n-j)!$ and $S(n-j)$, we get further that $\frac{S(n)}{(n-j)!} = \frac{S(n)}{S(n-j)} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}\right)\right)$. Finally, applying Lemma 3.3, we conclude by transitivity that $\frac{n!}{(n-j)!} = \frac{S(n)}{S(n-j)} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n}, \frac{1}{n-j}\right)\right)$.

Our goal is now to eliminate the dependence on $\frac{1}{n}$ from the big-$\mathcal{O}$ notation above. The key point is that $\frac{1}{n} \leq \frac{1}{n-j}$ as long as $j \leq n$. Unwinding the meaning of the statement above just obtained, we get that

$$\left| \frac{n!}{S(n)} \cdot \frac{S(n-j)}{(n-j)!} - 1 \right| \leq K \cdot \left(\frac{1}{n} + \frac{1}{n-j}\right) \leq 2 \cdot K \cdot \frac{1}{n-j}$$

as soon as $\left(\frac{1}{n}, \frac{1}{n-j}\right)$ lives in some sufficiently small neighborhood $U \subset \mathbb{R}^2$. Up to possibly shrinking U, we may assume that it takes the form $U = \left\{(y_0, y_1) \in \mathbb{R}^2 \,\middle|\, \max(|y_0|, |y_1|) < r\right\}$ for some $r > 0$. In this case, so long as $\frac{1}{n-j} < r$ holds, $\max\left(\frac{1}{n}, \frac{1}{n-j}\right) < r$ also does, so that $\left(\frac{1}{n}, \frac{1}{n-j}\right) \in U$ and the above expression does the job. □

Lemma 3.8. *On the set of pairs $(a, x) \in \mathbb{R}^2$ for which $x \neq 0$,*

$$\left(1 - \frac{a}{x}\right)^x = e^{-a} \cdot \left(1 + \mathcal{O}\left(\frac{a^2}{x}\right)\right).$$

Proof. The proof of this result uses a Taylor series; we skip it. □

3.3 Ball Volume Estimate

In this subsection, we record an extremely accurate estimate of the volume of the Hamming ball $\mathcal{B}_q(\mathbf{0}, z) \subset \mathbb{F}_q^n$.

The usual estimate for $|\mathcal{B}_q(\mathbf{0}, z)|$ (see e.g. Guruswami [Gur06, (3.3)]) takes the form

$$q^{n \cdot H_q(\zeta) - o(n)} \leq |\mathcal{B}_q(\mathbf{0}, z)| \leq q^{n \cdot H_q(\zeta)};$$

here, H_q is the *q-ary entropy function*. This estimate assumes that both q and the *relative radius* $\zeta \coloneqq \frac{z}{n}$ are held constant, and lets n grow to infinity. This estimate is not suitable for us, as we need finer-grained—i.e., independent—control over the quantities n, q and z. A more precise variant of this standard estimate—which evades the use of asymptotics—appears in e.g. MacWilliams and Sloane [MS74, Ch. 10 §11. Lem. 8]. We reproduce it here:

$$\frac{q^{n \cdot H_q(\zeta)}}{\sqrt{8 \cdot n \cdot \zeta \cdot (1-\zeta)}} \leq |\mathcal{B}_q(\mathbf{0}, z)| \leq q^{n \cdot H_q(\zeta)}. \tag{7}$$

In small-parameter regimes—those designed to "get a feel"—our estimate can be slightly *less* accurate than this one (see Example 3.11 below). On the other hand, it's actually sharper in the limit. The estimate (7) features lower and upper bounds that *themselves* drift apart from each other by the multiplicative ratio $\sqrt{8 \cdot n \cdot \zeta \cdot (1-\zeta)}$. For relative radius ζ held constant, this ratio grows to infinity as n grows (though if ζ simultaneously drifts towards 0 or 1, the picture becomes more complicated).

Our estimate is totally different in flavor. It's bidirectional; moreover, its *two-sided* error—i.e., the ratio between it and the ground truth—can be driven arbitrarily close to 1, provided that q, n and z are taken from suitable regions of the parameter space. As a side benefit, our estimate further manages to avoid the q-ary entropy function $H_q(\zeta)$, which is hard to make use of in proofs. We further compare our estimate to (7) in Remark 3.12 below.

We frame the matter of the ball volume $|\mathcal{B}_q(\mathbf{0}, z)|$ as a probabilistic one, concerning how many zeros we might expect a random word to contain. Indeed, a word $\mathbf{u} \in \mathbb{F}_q^n$ is in $\mathcal{B}_q(\mathbf{0}, z)$ if and only if it has at least $n - z$ zeros. The number of zeros that a *random* word is liable to have is modeled exactly by the binomial distribution $X \coloneqq \mathsf{Binom}\left(n, \frac{1}{q}\right)$. That is, we get:

$$|\mathcal{B}_q(\mathbf{0}, z)| = q^n \cdot \Pr[X \geq n - z].$$

By the definition of the binomial distribution, the probability that a random word $\mathbf{u} \leftarrow \mathbb{F}_q^n$ has exactly j zeros is

$$\Pr[X = j] = \binom{n}{j} \cdot \frac{1}{q^j} \cdot \left(1 - \frac{1}{q}\right)^{n-j}.$$

The probability that a word has at least $n - z$ zeros is thus

$$\Pr[X \geq n - z] = \sum_{j=n-z}^{n} \binom{n}{j} \cdot \frac{1}{q^j} \cdot \left(1 - \frac{1}{q}\right)^{n-j}. \tag{8}$$

In particular, we have already proved the following trivial lower-bound, which we extract and state separately.

Theorem 3.9. *For each pair of positive integers n and q, and each $z \in \{0, \ldots, n\}$, we have:*

$$|\mathcal{B}_q(\mathbf{0}, z)| \geq \binom{n}{n-z} \cdot q^z \cdot \left(1 - \frac{1}{q}\right)^z.$$

Proof. Indeed, we just need to multiply the inequality $\Pr[X \geq n - z] \geq \Pr[X = n - z]$ through by q^n. □

We now proceed to our estimate proper.

Theorem 3.10. *For each pair of positive integers $q \geq n$ and each $z \in \{1, \ldots, n-1\}$, we have the volume estimate:*

$$|\mathcal{B}_q(\mathbf{0}, z)| = \sqrt{\frac{n}{z}} \cdot \frac{n^{n-z} \cdot q^z \cdot e^{-\frac{(n-z)^2}{n} - \frac{z}{q}}}{(n-z)!} \cdot \left(1 + \mathcal{O}\left(\frac{1}{z}, \frac{z}{q^2}, \frac{z}{q \cdot (n-z)}, \frac{(n-z)^2 \cdot z}{n^2}\right)\right).$$

Proof. Our argument has two main steps. First we argue that the sum (8) is *essentially* a geometric series. That is, the ratio of each term in (8) to the one before it is bounded from above by a function that can be made arbitrarily small, for suitable inputs n, q and z. After that, we produce an estimate for the initial term $j = n - z$ in (8). For typographical convenience, we abbreviate $P_j \coloneqq \Pr[X = j]$.

We first note that for each $j \in \{1, \ldots, n-1\}$,

$$\frac{P_{j+1}}{P_j} = \frac{\binom{n}{j+1} \cdot \frac{1}{q^{j+1}} \cdot \left(1 - \frac{1}{q}\right)^{n-j-1}}{\binom{n}{j} \cdot \frac{1}{q^j} \cdot \left(1 - \frac{1}{q}\right)^{n-j}} = \frac{n-j}{(j+1) \cdot (q-1)} \leq \frac{n-j}{q \cdot j},$$

where we use $j < n \leq q$ to get the last inequality. Moreover, this quantity is decreasing in $j \in \{1, \ldots, n-1\}$. It follows that

$$\frac{P_{j+c}}{P_j} \leq \left(\frac{n-j}{q \cdot j}\right)^c$$

holds for each $j \in \{1, \ldots, n\}$ and each $c \in \{0, \ldots, n-j\}$. This implies that the sum (8) above is bounded from above by a decaying geometric series, and that we can cut it off after its initial term with only a small error. In other words,

$$\Pr[X \geq n - z] = \Pr[X = n - z] \cdot \left(1 + \mathcal{O}\left(\frac{z}{q \cdot (n-z)}\right)\right).$$

It remains to get an asymptotic estimate for $\Pr[X = j]$. For this task, we use a "chaining" technique, which owes its correctness to Lemmas 3.2, 3.3 and 3.4. Indeed, for each $j \in \{0, \ldots, n-1\}$, we have the following estimate, which roughly follows the method by which the Poisson approximation of the binomial distribution is derived.

$$
\begin{aligned}
P_j &= \binom{n}{j} \cdot \frac{\left(1 - \frac{1}{q}\right)^{n-j}}{q^j} && \text{(by definition, see above.)} \\
&= \sqrt{\frac{n}{n-j}} \cdot \frac{n^n}{j! \cdot e^j \cdot (n-j)^{n-j}} \cdot \frac{\left(1 - \frac{1}{q}\right)^{n-j}}{q^j} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}\right)\right) && \text{(use Corollary 3.7 on } \tbinom{n}{j}.) \\
&= \sqrt{\frac{n}{n-j}} \cdot \frac{n^j}{j! \cdot e^j \cdot (1 - \frac{j}{n})^{n-j}} \cdot \frac{\left(1 - \frac{n-j/q}{n-j}\right)^{n-j}}{q^j} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}\right)\right) && \text{(divide top and bottom by } n^{n-j}.) \\
&= \sqrt{\frac{n}{n-j}} \cdot \frac{n^j \cdot e^{-\frac{n-j}{q} - j}}{j! \cdot \left(1 - \frac{j(n-j)/n}{n-j}\right)^{n-j} \cdot q^j} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}, \frac{n-j}{q^2}\right)\right) && \text{(use Lemma 3.8 on } \left(1 - \tfrac{n-j/q}{n-j}\right)^{n-j}.) \\
&= \sqrt{\frac{n}{n-j}} \cdot \frac{n^j \cdot e^{\frac{j(n-j)}{n} - \frac{n-j}{q} - j}}{j! \cdot q^j} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}, \frac{n-j}{q^2}, \frac{j^2 \cdot (n-j)}{n^2}\right)\right) && \text{(use Lemma 3.8 on } \left(1 - \tfrac{j(n-j)/n}{n-j}\right)^{n-j}.) \\
&= \sqrt{\frac{n}{n-j}} \cdot \frac{n^j \cdot e^{-\frac{j^2}{n} - \frac{n-j}{q}}}{j! \cdot q^j} \cdot \left(1 + \mathcal{O}\left(\frac{1}{n-j}, \frac{n-j}{q^2}, \frac{j^2 \cdot (n-j)}{n^2}\right)\right). && \text{(simplify the numerator.)}
\end{aligned}
$$

Evaluating this expression at $j = n - z$, combining it with the first-term approximation stated above, and multiplying through by q^n to convert the resulting probability to a count, we conclude that

$$
|\mathcal{B}_q(\mathbf{0}, z)| = \sqrt{\frac{n}{z}} \cdot \frac{n^{n-z} \cdot q^z \cdot e^{-\frac{(n-z)^2}{n} - \frac{z}{q}}}{(n-z)!} \cdot \left(1 + \mathcal{O}\left(\frac{1}{z}, \frac{z}{q^2}, \frac{z}{q \cdot (n-z)}, \frac{(n-z)^2 \cdot z}{n^2}\right)\right),
$$

which is what we wanted to prove. □

Example 3.11. We tabulate the accuracy of our estimate experimentally for $q = 365$, $n = 23$, and various values of z. For each radius z, we record the true volume of the Hamming ball $\mathcal{B}_q(\mathbf{0}, z) \subset \mathbb{F}_q^n$, the entropy-based lower and upper bounds (recall (7)), and finally the estimate of Theorem 3.10 above. In our results for Theorem 3.10, we ignore that estimate's controller expressions, and just report the raw estimate. All results below are base-q logarithms.

Case $q = 365$, $n = 23$	$z = 1$	$z = 4$	$z = 7$	$z = 10$	$z = 13$	$z = 16$	$z = 19$	$z = 22$
Ground truth $\lvert\mathcal{B}_q(\mathbf{0}, z)\rvert$	1.5310	5.5387	9.1003	12.3601	15.3590	18.0969	20.5334	22.5264
Entropy lower-bound (7)	1.5243	5.5218	9.0819	12.3412	15.3398	18.0777	20.5148	22.5145
Entropy upper-bound (7)	1.6967	5.7993	9.3923	12.6642	15.6628	18.3881	20.7924	22.6870
Theorem 3.10	1.1748	4.9157	8.5155	11.9069	15.0598	17.9374	20.4766	22.5176

Our estimate gets sharper as z gets closer to n, a tendency predicted by Theorem 3.10 (we make this precise below). In fact, it's not hard to find parameters n, q and z for which our estimate beats the entropy lower-bound. In the case $n = 365$ (and again $q = 365$), our estimate beats the entropy lower-bound for the top eight high values $z \in \{357, \ldots, 364\}$. We tabulate these below.

Case $q = 365$, $n = 365$	$z = 357$	$z = 358$	$z = 359$	$z = 360$	$z = 361$	$z = 362$	$z = 363$	$z = 364$
Ground truth $\lvert\mathcal{B}_q(\mathbf{0}, z)\rvert$	363.043	363.401	363.736	364.046	364.327	364.572	364.774	364.922
Entropy lower-bound (7)	363.005	363.360	363.693	363.999	364.274	364.511	364.700	364.824
Entropy upper-bound (7)	363.355	363.700	364.019	364.310	364.567	364.779	364.934	365.000
Theorem 3.10	363.009	363.368	363.703	364.011	364.287	364.525	364.713	364.831

We're not aware of a setting in which our result overestimates the ground truth.

Remark 3.12. As an illuminating exercise, we put our estimate and the classic entropy estimate (7) on—roughly—the same footing. To make the relationship of interest easier to see, we take base-q logarithms of everything. We abbreviate $\lg \coloneqq \log_q$ for typographical reasons. In this setting, the entropy upper and lower bounds become:

$$\lg\Big(q^{n\cdot H_q(\frac{z}{n})}\Big) \approx z - z\cdot\lg\Big(\frac{z}{n}\Big) - (n-z)\cdot\lg\Big(1-\frac{z}{n}\Big), \tag{9}$$

$$\lg\left(\frac{q^{n\cdot H_q(\zeta)}}{\sqrt{8\cdot n\cdot\zeta\cdot(1-\zeta)}}\right) \approx \lg\Big(q^{n\cdot H_q(\frac{z}{n})}\Big) - \frac{1}{2}\cdot(\lg(8)+\lg(z)+\lg(n-z)-\lg(n)). \tag{10}$$

Now, as for Theorem 3.10, we must use Stirling's approximation again to handle the factorial. In log form, that approximation reads $\lg(n!) \approx n\cdot\lg(n) - n\cdot\lg(e) + \frac{1}{2}\cdot\lg(2\pi\cdot n)$. Using this, and writing $\widetilde{B}_q(\mathbf{0},z)$ for the right-hand side of Theorem 3.10 (excluding its big-$\mathcal{O}$ expression), we get:

$$\begin{aligned}\lg\Big(\widetilde{B}_q(\mathbf{0},z)\Big) &\approx \frac{1}{2}\cdot(\lg(n)-\lg(z)) + (n-z)\cdot\lg(n) + z + \left(-\frac{(n-z)^2}{n} - \frac{z}{q}\right)\cdot\lg(e)\\ &\qquad - (n-z)\cdot\lg(n-z) + (n-z)\cdot\lg(e) - \frac{1}{2}\cdot\lg(2\pi\cdot(n-z))\\ &= z - (n-z)\cdot\lg\Big(1-\frac{z}{n}\Big) - \frac{1}{2}\cdot(\lg(2\pi)+\lg(z)+\lg(n-z)-\lg(n)) + \left(n-z-\frac{(n-z)^2}{n}-\frac{z}{q}\right)\cdot\lg(e).\end{aligned}$$

As we can see, we have a term that matches part of $H_q\big(\frac{z}{n}\big)$. We also have an error term that almost exactly matches (10)'s, though with a 2π instead of an 8. The remaining unmatched parts are $-z\cdot\log_q(\frac{z}{n})$ (which both (9) and (10) have, while we don't), and $\Big((n-z)-\frac{(n-z)^2}{n}-\frac{z}{q}\Big)\cdot\log_q(e)$ (which we have, but neither (9) nor (10) does). When z is much smaller than q, $\frac{z}{q}$ disappears from the latter term, and we can write what remains of that term as $z\cdot\big(1-\frac{z}{n}\big)\cdot\log_q(e)$.

Next, we observe that $-z\cdot\log_q(\frac{z}{n})$ is the evaluation of $-z\cdot\log_q(1-x)$ at $x = 1-\frac{z}{n}$. The first-order Taylor approximation (in x) of this function, again evaluated at $x = 1-\frac{z}{n}$, is $z\cdot\big(1-\frac{z}{n}\big)\cdot\log_q(e)$, which is our term.

3.4 Intersection Estimate

In this subsection, we turn to intersections. We fix an $[n,k,d]$-code $C\subset\mathbb{F}_q^n$ and a radius $z\in\{0,\dots,n\}$. For each weight $w\in\{d,\dots,n\}$, we write $A_w\subset C$ for the set of codewords $\mathbf{v}\in C$ of weight exactly w.

For each codeword $\mathbf{v}\in C$, by translation-invariance, the respective portions of $\mathcal{B}_q(\mathbf{v},z)$ that intersect other balls $\mathcal{B}_q(\mathbf{v}',z)$ (for $\mathbf{v}'\neq\mathbf{v}$) correspond in a one-to-one manner with those portions of $\mathcal{B}_q(\mathbf{0},z)$ that intersect other balls $\mathcal{B}_q(\mathbf{v},z)$ (for $\mathbf{v}\neq\mathbf{0}$). Moreover, for each codeword $\mathbf{v}\in A_w$ of weight exactly w, again by symmetry, the cardinality $|\mathcal{B}_q(\mathbf{0},z)\cap\mathcal{B}_q(\mathbf{v},z)|$ depends only on w. Combining these facts, we get the following equality:

$$\begin{aligned}\sum_{\mathbf{v}\neq\mathbf{v}'}|\mathcal{B}_q(\mathbf{v},z)\cap\mathcal{B}_q(\mathbf{v}',z)| &= q^k\cdot\sum_{\mathbf{v}\neq\mathbf{0}}|\mathcal{B}_q(\mathbf{0},z)\cap\mathcal{B}_q(\mathbf{v},z)| && \text{(by translational symmetry.)}\\ &= q^k\cdot\sum_{w=d}^{n}|A_w|\cdot|\mathcal{B}_q(\mathbf{0},z)\cap\mathcal{B}_q(\mathbf{v},z)|. && \text{(partition sum by weight of } \mathbf{v}.)\end{aligned}$$

In the final expression above, for each sum index $w\in\{d,\dots,n\}$, we mean by $\mathbf{v}$ a (single, fixed, and arbitrary) codeword of weight exactly w. We note that we interpret $\sum_{\mathbf{v}\neq\mathbf{v}'}|\mathcal{B}_q(\mathbf{v},z)\cap\mathcal{B}_q(\mathbf{v}',z)|$ as a sum over all *ordered* pairs $(\mathbf{v},\mathbf{v}')$ for which $\mathbf{v}\neq\mathbf{v}'$. There's thus a kind of implicit double-counting in that expression; indeed, we divide by 2 when we use it below (see (12)).

This fact motivates the main result of this section, Corollary 3.17. In that corollary, we estimate the sum $\sum_{w=d}^{n}|A_w|\cdot|\mathcal{B}_q(\mathbf{0},z)\cap\mathcal{B}_q(\mathbf{v},z)|$. This sum can be viewed as a "second-order" analogue of our first-order sum estimate (see Theorem 3.10).

Lemma 3.13. *For each MDS $[n,k,d]$-code $C\subset\mathbb{F}_q^n$ and each weight $w\in\{d,\dots,n\}$, the set $A_w\subset C$ consisting of codewords $\mathbf{v}\in C$ of weight $w(\mathbf{v})=w$ has cardinality at most:*

$$|A_w| \le \binom{n}{w}\cdot q^{k-(n-w)}.$$

Proof. To each codeword $\mathbf{v} \in A_w$ of weight w, we may associate the set $S' \subset \{0, \dots, n-1\}$ of indices where $\mathbf{v}$ is 0—whose size is exactly $n-w$, by fiat—together with the values $\mathbf{v}$ takes on S'', some fixed, arbitrary $k-(n-w)$-element subset of $\{0, \dots, n-1\} \setminus S'$ (if C is MDS, then $w \geq d$ implies $k-(n-w) \geq 1$).

The union $S \coloneqq S' \cup S''$ is of size exactly k. If C is MDS, this association is injective. Indeed, each collision would witness a pair of unequal codewords whose restrictions to S were identical, a circumstance that's impossible if C is MDS (their distance would be less than or equal to $n-k < d$). □

We could replace q by $q-1$ in Lemma 3.13. That lemma's bound overcounts, as it counts codewords of weight less than w, and moreover counts those codewords multiple times.

Lemma 3.14. *For each $[n,k,d]$-code $C \subset \mathbb{F}_q^n$, each ball radius $z \in \{0, \dots, d\}$, and each weight $w \in \{d, \dots, n\}$, for each codeword $\mathbf{v} \in C$ of weight $w(\mathbf{v}) = w$, we have the intersection upper bound*

$$|\mathcal{B}_q(\mathbf{0}, z) \cap \mathcal{B}_q(\mathbf{v}, z)| \leq \sum_{s=0}^{n-w} \binom{n-w}{s} \cdot \binom{w}{w+s-z, w+s-z} \cdot q^{2z-s-w}.$$

Proof. We let nonzero $\mathbf{v} \in C$ have weight w. We write $\Delta(\mathbf{v}, \mathbf{0}) \subset \{0, \dots, n-1\}$ for the set of positions at which $\mathbf{v}$ is nonzero. By fiat, we have $|\Delta(\mathbf{v}, \mathbf{0})| = w$.

To each $\mathbf{u} \in \mathcal{B}_q(\mathbf{0}, z) \cap \mathcal{B}_q(\mathbf{v}, z)$, we may associate the number, say s, of indices $i \in \{0, \dots, n-1\} \setminus \Delta(\mathbf{v}, \mathbf{0})$ at which $\mathbf{u}_i$ is *unequal* to the common value $\mathbf{v}_i = 0$. In other words, s is the common cardinality of $\Delta(\mathbf{u}, \mathbf{0}) \setminus \Delta(\mathbf{v}, \mathbf{0}) = \Delta(\mathbf{u}, \mathbf{v}) \setminus \Delta(\mathbf{v}, \mathbf{0})$ (note that wherever $\mathbf{v}_i = 0$, the conditions $\mathbf{u}_i = 0$ and $\mathbf{u}_i = \mathbf{v}_i$ are equivalent). We may moreover associate to $\mathbf{u}$ the values it takes at these indices, of which there are at most q^s possibilities.

Since $\mathbf{u} \in \mathcal{B}_q(\mathbf{0}, z)$, the *total* number of indices $i \in \{0, \dots, n-1\}$ at which $\mathbf{u}_i = 0$ is at least $n-z$. Since exactly $n-w-s$ among these indices live in $\{0, \dots, n-1\} \setminus \Delta(\mathbf{v}, \mathbf{0})$, at least $n-z-(n-w-s) = w+s-z$ of them must live within the set $\Delta(\mathbf{v}, \mathbf{0})$. Moreover, this quantity is nonnegative, by our hypothesis on z. Thus, to $\mathbf{u}$ we may moreover associate some set $S_{\mathbf{0}} \subset \Delta(\mathbf{v}, \mathbf{0})$ say, of size exactly $w+s-z$, on which $\mathbf{u}_i = 0$ holds identically.

Similarly, since $\mathbf{u} \in \mathcal{B}_q(\mathbf{v}, z)$, we must have $\mathbf{u}_i = \mathbf{v}_i$ for at least $n-z$ indices $i \in \{0, \dots, n-1\}$. Since $n-w-s$ among these indices live in $\{0, \dots, n-1\} \setminus \Delta(\mathbf{v}, \mathbf{0})$, again at least $w+s-z$ of them must live within $\Delta(\mathbf{v}, \mathbf{0})$. We may thus pick a further set, say $S_{\mathbf{v}} \subset \Delta(\mathbf{v}, \mathbf{0})$, of size exactly $w+s-z$, such that $\mathbf{u}_i = \mathbf{v}_i$ for each $i \in S_{\mathbf{v}}$. Crucially, $S_{\mathbf{0}}$ and $S_{\mathbf{v}}$ are disjoint, since each is contained within $\Delta(\mathbf{v}, \mathbf{0})$, and no $i \in \Delta(\mathbf{v}, \mathbf{0})$ can satisfy both $\mathbf{u}_i = 0$ and $\mathbf{u}_i = \mathbf{v}_i$.

On the $w - 2 \cdot (w+s-z) = 2z-2s-w$ positions $i \in \Delta(\mathbf{v}, \mathbf{0})$ not yet accounted for (if this number is nonnegative), $\mathbf{u}_i$ can again be arbitrary.

This state of affairs is summarized in Figure 1 below.

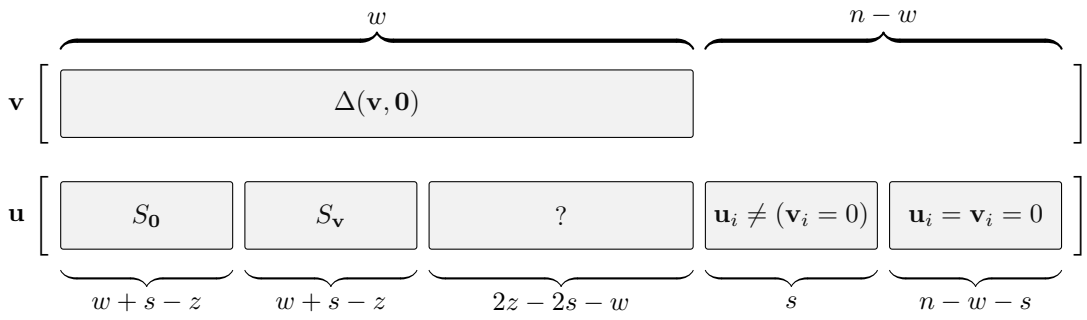


Figure 1: A depiction of the situation in Lemma 3.14.

The total number of choices we've made is thus

$$\sum_{s=0}^{n-w} \binom{n-w}{s} \cdot q^s \cdot \binom{w}{w+s-z, w+s-z} \cdot q^{2z-2s-w},$$

and these choices characterize $\mathbf{u}$ uniquely.

Simplifying the powers above, we get the expression in the statement of the lemma. □

Remark 3.15. The multinomial expressions in the statement of Lemma 3.14 are of course zero whenever any of their entries are negative. In fact, this behavior is expected. It would have been equivalent, though slightly more cumbersome, to take the sum only over $s \in \left\{\max(0, z-w), \dots, \min\left(n-w, \left\lfloor \frac{2z-w}{2} \right\rfloor\right)\right\}$; over exactly this range, both multinomials are nonzero. In particular, if $2z-w < 0$, then the entire sum in the statement of the lemma vanishes. This makes perfect sense, since if $2z < w$, the triangle inequality implies that $\mathcal{B}_q(\mathbf{0}, z) \cap \mathcal{B}_q(\mathbf{v}, z)$ must be empty.

Combining Lemmas 3.13 and 3.14, we get, for each MDS $[n,k,d]$-code $C \subset \mathbb{F}_q^n$ and each $z \in \{0,\dots,d\}$, the bound:

$$\begin{aligned}\sum_{w=d}^{n} |A_w| \cdot |\mathcal{B}_q(\mathbf{0},z) \cap \mathcal{B}_q(\mathbf{v},z)| &\le \sum_{w=d}^{n} \binom{n}{w} \cdot q^{k-(n-w)} \cdot \left(\sum_{s=0}^{n-w} \binom{n-w}{s} \cdot \binom{w}{w+s-z, w+s-z} \cdot q^{2z-s-w} \right) \\ &= \sum_{s=0}^{n-d} \sum_{w=d}^{n-s} \binom{n}{s, n-w-s, w+s-z, w+s-z} \cdot q^{2z-s+k-n} \\ &= \sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w},\end{aligned} \tag{11}$$

where, in the last step, we define

$$I_{s,w} \coloneqq q^{2z-s+k-n} \cdot \binom{n}{s, n-w-s, w+s-z, w+s-z}.$$

We similarly define

$$I'_{s,w} \coloneqq \binom{n}{s, n-w-s, w+s-z, w+s-z},$$

so that $I_{s,w} = q^{2z-s+k-n} \cdot I'_{s,w}$ for each $s \in \{0,\dots,n-d\}$ and each $w \in \{d,\dots,n-s\}$.

As long as $z \le d$, for each $w \in \{d,\dots,n\}$ and $s \in \{0,\dots,n-d\}$, $I'_{s,w}$ is nonzero precisely when $s+w \le n$ and $w \le 2z-2s$ both hold (this is essentially Remark 3.15). Moreover, in exactly this case, $2z-2s \ge w \ge d$, so that $2z-d-s \ge s$, and in particular $2z-s+k-n = 2z-d-s+1 \ge s+1 \ge 1$. In other words, the exponent $2z-s+k-n$ in $I_{s,w}$ is necessarily nonnegative (in fact positive) whenever the binomial coefficient $I'_{s,w}$ is nonvanishing.

The following theorem shows that the double-sum $\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w}$ is closely approximated by $I_{0,n}$, its lone summand for which $s=0$ and $w=n$.

Theorem 3.16. *For each MDS $[n,k,d]$-code $C \subset \mathbb{F}_q^n$ and each ball radius $z \in \left\{\lceil \frac{n}{2} \rceil, \dots, d\right\}$, we have the estimate*

$$\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w} = \binom{n}{n-z, n-z} \cdot q^{2z+k-n} \cdot \left(1 + \mathcal{O}\left(\frac{(n-z)\cdot z^2}{q}, \frac{(n-z)^2}{2z-n+1} \right)\right).$$

Proof. We prove the theorem in two steps. First, we show that $\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w}$ is close to $\sum_{w=d}^{n} I_{0,w}$ (i.e., we can restrict $s=0$). After that, we show that $\sum_{w=d}^{n} I_{0,w}$ is close to $I_{0,n}$ (we can further restrict $w=n$).

We begin with the first task. Writing

$$P_s \coloneqq \sum_{w=d}^{n-s} I_{s,w},$$

we trivially have

$$\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w} = \sum_{s=0}^{n-d} P_s.$$

We use another geometric-series-based argument. To do this, we study the ratios $\frac{P_{s+1}}{P_s}$, for $s \in \{0,\dots,n-d-1\}$, and argue that these ratios are decreasing (wherever they're defined).

We write $s^* \coloneqq \min\left(n-d, \left\lfloor \frac{2z-d}{2} \right\rfloor\right)$. If $s > s^*$, then P_s vanishes completely (this is again equivalent to Remark 3.15). Our hypothesis implies $2z \ge n \ge d$, so s^* is nonnegative. (If $2z < d$ held, this entire problem would degenerate, by the triangle inequality.) For each $s \in \{0,\dots,s^*\}$, P_s has at least one nonzero term, and is strictly positive.

We illustrate the vanishing pattern of the multinomials $I'_{s,w}$ in Figure 2, whose content should make this proof easier to follow. The lightly-shaded, dashed cells correspond to those $I'_{s,w}$ for which $s+w \le n$. Outside this region, each multinomial $I'_{s,w}$ certainly vanishes. On the other hand, $I'_{s,w}$ also vanishes unless the further condition $w \le 2z-2s$ holds. For each $s \in \{0,\dots,n-d\}$, the right-hand terminus of the s^{th} row is thus $\min(n-s, 2z-2s)$. Finally, $s^* \in \{0,\dots,n-d\}$ is the maximally-indexed row that *doesn't* consist entirely of vanishing cells. As long as $2z \ge d$, the *top-left* cell of Figure 2 is nonzero; in particular, $s^* \ge 0$ in this case. Our stronger hypothesis $2z \ge n$ implies that in fact the *top-right* cell of Figure 2 is nonzero. Whenever $s^* \ge n-d$, the entire top-left triangle of the grid becomes fully shaded. On the other hand, s^* can be less than $n-d$ in general; $s^* \ge n-d$ holds if and only if $z \ge n - \left\lfloor \frac{d}{2} \right\rfloor$ (both cases can happen). For each column index $w \in \{d,\dots,n\}$, we thus have the active range $s \in \left\{0,\dots,\min\left(n-w, \left\lfloor \frac{2z-w}{2} \right\rfloor\right)\right\}$; our hypothesis $2z \ge n$ implies that all of these ranges are nonempty.

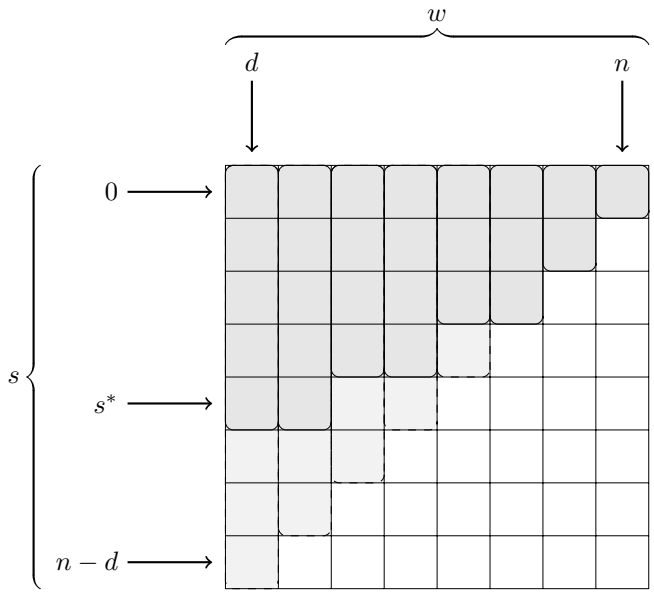


Figure 2: A depiction of the situation in Theorem 3.16.

We return to the proof. For each $s \in \{0, \ldots, s^* - 1\}$, we thus get

$$\frac{P_{s+1}}{P_s} = \frac{\sum_{w=d}^{n-s-1} q^{2z-s-1+k-n} \cdot I'_{s+1,w}}{\sum_{w=d}^{n-s} q^{2z-s+k-n} \cdot I'_{s,w}} = \frac{1}{q} \cdot \frac{\sum_{w=d}^{n-s-1} I'_{s+1,w}}{\sum_{w=d}^{n-s} I'_{s,w}}.$$

In the denominator, $I'_{s,w}$ is nonzero exactly on $w \in \{d, \ldots, \min(n-s, 2z-2s)\}$ (see again above). In the numerator, $I'_{s+1,w}$ is nonzero exactly on $w \in \{d, \ldots, \min(n-s-1, 2z-2s-2)\}$. To upper-bound the quotient-of-sums above, it thus suffices to uniformly upper-bound each *individual* quotient $\frac{I'_{s+1,w}}{I'_{s,w}}$, for $w \in \{d, \ldots, \min(n-s-1, 2z-2s-2)\}$. (This range is nonempty, in view of our assumption $s < s^*$.) In fact, over exactly this range,

$$\frac{I'_{s+1,w}}{I'_{s,w}} = \frac{\binom{n}{s+1, n-w-s-1, w+s+1-z, w+s+1-z}}{\binom{n}{s, n-w-s, w+s-z, w+s-z}} = \frac{(n-w-s) \cdot (2z-w-2s) \cdot (2z-w-2s-1)}{(s+1) \cdot (w+s+1-z)^2},$$

a quantity whose numerator and denominator are both strictly positive. Moreover, again over this range, namely $w \in \{d, \ldots, \min(n-s-1, 2z-2s-2)\}$, this quantity is decreasing in w (its numerator is decreasing in w and its denominator increasing in w). We can thus substitute $w = d$ into the above bound, and thereby get a *uniform* bound, valid for all $w \in \{d, \ldots, \min(n-s-1, 2z-2s-2)\}$. We have thus proven that, for each $s \in \{0, \ldots, s^* - 1\}$,

$$\frac{P_{s+1}}{P_s} \leq \frac{1}{q} \cdot \frac{(n-d-s) \cdot (2z-d-2s) \cdot (2z-d-2s-1)}{(s+1) \cdot (d+s+1-z)^2} = \frac{1}{q} \cdot \frac{I'_{s+1,d}}{I'_{s,d}}.$$

Our goal is now to show that this common ratio is *itself* decreasing over the range $s \in \{0, \ldots, s^* - 1\}$. We have already seen that both its numerators and denominators are identically nonvanishing for each $s \in \{0, \ldots, s^* - 1\}$. Moreover, its numerator is decreasing in s and its denominator increasing in s. We thus again substitute $s = 0$, and obtain the *uniform* bound

$$\frac{P_{s+1}}{P_s} \leq \frac{1}{q} \cdot \frac{(n-d) \cdot (2z-d) \cdot (2z-d-1)}{(d+1-z)^2} = \frac{1}{q} \cdot \frac{I'_{1,d}}{I'_{0,d}},$$

valid for each $s \in \{0, \ldots, s^* - 1\}$. In light of our hypothesis $z \leq d$, we finally get

$$\frac{P_{s+1}}{P_s} \leq \frac{1}{q} \cdot \frac{(n-z) \cdot z \cdot (z-1)}{1} \leq \frac{(n-z) \cdot z^2}{q}.$$

In view of the geometric series formula, we've thus just proven that

$$\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w} = \sum_{s=0}^{n-d} P_s = P_0 \cdot \left(1 + \mathcal{O}\left(\frac{(n-z) \cdot z^2}{q}\right)\right).$$

We turn to

$$P_0 = \sum_{w=d}^{n} I_{0,w} = \sum_{w=d}^{n} q^{2z+k-n} \cdot I'_{0,w} = q^{2z+k-n} \cdot \sum_{w=d}^{n} I'_{0,w}.$$

We show that—again for appropriate n and z—the sum $\sum_{w=d}^{n} I_{0,w}$ is dominated by $I_{0,n}$, its *last* term. To do this, we yet again study the ratios between adjacent terms, now going high-to-low.

As we've already seen, $I_{0,w}$ is nonzero on the range $w \in \{d, \dots, \min(n, 2z)\}$, and indeed for all $w \in \{d, \dots, n\}$ by our hypothesis on z. For each $w \in \{d+1, \dots, n\}$, a straightforward calculation gives:

$$\frac{I_{0,w-1}}{I_{0,w}} = \frac{I'_{0,w-1}}{I'_{0,w}} = \frac{\binom{n}{n-w+1, w-1-z, w-1-z}}{\binom{n}{n-w, w-z, w-z}} = \frac{(w-z)^2}{(n-w+1) \cdot (2z-w+1)}.$$

We call this quantity R_w. Again, R_w's numerator is increasing in w and its denominator decreasing in w; this shows that R_w is increasing over $w \in \{d+1, \dots, n\}$.

In view of the usual telescoping expression

$$\frac{I_{0,w}}{I_{0,n}} = \frac{I_{0,n-1}}{I_{0,n}} \cdot \dots \cdot \frac{I_{0,w}}{I_{0,w+1}} = R_n \cdot \dots \cdot R_{w+1},$$

valid for each $w \in \{d, \dots, n-1\}$, this increasingness implies that

$$\left| \frac{\sum_{w=d}^{n} I_{0,w}}{I_{0,n}} - 1 \right| = \left| \frac{I_{0,n-1}}{I_{0,n}} + \dots + \frac{I_{0,d}}{I_{0,n}} \right| \leq R_n + \dots + R_n^{n-d}$$

is bounded from above by a geometric series of common ratio

$$R_n = \frac{I_{0,n-1}}{I_{0,n}} = \frac{\binom{n}{1, n-1-z, n-1-z}}{\binom{n}{n-z, n-z}} = \frac{(n-z)^2}{2z-n+1}.$$

We have just shown that

$$P_0 = \sum_{w=d}^{n} I_{0,w} = I_{0,n} \cdot \left(1 + \mathcal{O}\left(\frac{(n-z)^2}{2z-n+1}\right)\right) = \binom{n}{n-z, n-z} \cdot q^{2z+k-n} \cdot \left(1 + \mathcal{O}\left(\frac{(n-z)^2}{2z-n+1}\right)\right),$$

where the second equality is the definition of $I_{0,n}$.

Applying transitivity, we obtain the statement of the theorem. □

In our applications below, we instead use the following effective variant of Theorem 3.16.

Corollary 3.17. *For each MDS $[n, k, d]$-code $C \subset \mathbb{F}_q^n$ and each ball radius $z \in \left\{\lceil \frac{n}{2} \rceil, \dots, d\right\}$, if n, q and z are such that $\frac{(n-z) \cdot z^2}{q} < 1$ and $\frac{(n-z)^2}{2z-n+1} < 1$, then*

$$\sum_{w=d}^{n} |A_w| \cdot |\mathcal{B}_q(\mathbf{0}, z) \cap \mathcal{B}_q(\mathbf{v}, z)| \leq \binom{n}{n-z, n-z} \cdot q^{2z+k-n} \cdot \frac{1}{1 - \frac{(n-z) \cdot z^2}{q}} \cdot \frac{1}{1 - \frac{(n-z)^2}{2z-n+1}}.$$

Proof. The corollary follows from (11) and the proof of Theorem 3.16 above. Indeed, by the geometric series formula, the proof of Theorem 3.16 shows that, under the hypothesis of the corollary, we actually have

$$\sum_{s=0}^{n-d} \sum_{w=d}^{n-s} I_{s,w} = \sum_{s=0}^{n-d} P_s \leq P_0 \cdot \frac{1}{1 - \frac{(n-z) \cdot z^2}{q}}$$

and

$$P_0 = \sum_{w=d}^{n} I_{0,w} \leq I_{0,n} \cdot \frac{1}{1 - \frac{(n-z)^2}{2z-n+1}}.$$

Substituting the second of these bounds into the first and using (11), we obtain the statement of the corollary on the nose. □

We write $R(n, q, z)$ for the correction term in the statement of Corollary 3.17 above.

4 Code Bounds

We are interested in $\Pr_{\mathbf{u}\leftarrow\mathbb{F}_q^n}[d(\mathbf{u},C)\leq z]$, the probability that a random word $\mathbf{u}\leftarrow\mathbb{F}_q^n$ is z-close to some MDS code $C\subset\mathbb{F}_q^n$ of dimension k, and in particular in lower-bounding this quantity. Indeed, we saw in Subsection 2.4 above that $\mathsf{err}(C,z)$ is lower-bounded by $\Pr_{\mathbf{u}\leftarrow\mathbb{F}_q^n}[d(\mathbf{u},C)\leq z]$.

To bound $\Pr_{\mathbf{u}\leftarrow\mathbb{F}_q^n}[d(\mathbf{u},C)\leq z]$ in turn from below, we relate it to an expression involving Hamming balls. Indeed,

$$\Pr_{\mathbf{u}\leftarrow\mathbb{F}_q^n}[d(\mathbf{u},C)\leq z]=\frac{\left|\bigcup_{\mathbf{v}\in C}\mathcal{B}_q(\mathbf{v},z)\right|}{q^n}.$$

Moreover, by the inclusion–exclusion principle,

$$\frac{\left|\bigcup_{\mathbf{v}\in C}\mathcal{B}_q(\mathbf{v},z)\right|}{q^n}\geq\frac{\sum_{\mathbf{v}\in C}|\mathcal{B}_q(\mathbf{v},z)|}{q^n}-\frac{1}{2}\cdot\frac{\sum_{\mathbf{v}\neq\mathbf{v}'}|\mathcal{B}_q(\mathbf{v},z)\cap\mathcal{B}_q(\mathbf{v}',z)|}{q^n}.\tag{12}$$

Theorem 3.9 above serves to lower-bound the first of the two quantities above. Corollary 3.17 serves to upper-bound the second quantity, at least for those n, q and z for which $R(n,q,z)$ is well-defined. Combining those bounds—and multiplying them through by $\frac{q^k}{q^n}$ and $\frac{1}{2}\cdot\frac{q^k}{q^n}$, respectively—we obtain the combined *lower-bound*

$$\Pr_{\mathbf{u}\leftarrow\mathbb{F}_q^n}[d(\mathbf{u},C)\leq z]\geq\binom{n}{n-z}\cdot q^{z+k-n}\cdot\left(1-\frac{1}{q}\right)^z-\frac{1}{2}\cdot\binom{n}{n-z,n-z}\cdot q^{2z+2k-2n}\cdot R(n,q,z),\tag{13}$$

where again $R(n,q,z)$ is the auxiliary correction function introduced in Corollary 3.17. We call the entire quantity on the right-hand side of (13) $\mathscr{Q}(n,q,z,k)$.

It's remarkable how clean the expression $\mathscr{Q}(n,q,z,k)$ just above really is. Up to a few nonsense factors that will disappear for large z,n and q—see in particular Lemmas 4.10 and 4.12 below—what remains is simply

$$\binom{n}{n-z}\cdot q^{z+k-n}-\frac{1}{2}\cdot\binom{n}{n-z,n-z}\cdot q^{2z+2k-2n}.$$

Moreover, upon carefully expanding these binomials in the right limiting conditions, we essentially recover the suggestive, second-order truncation we sketched in Subsection 1.3.

4.1 Finishing the Estimates

Our goal is to define an infinite family of alphabets $\mathbb{F}_q$ and MDS (for example, Reed–Solomon) $[n,k,d]$-codes $C\subset\mathbb{F}_q^n$ for which (13) becomes large infinitely many times. More specifically, in order to disprove Conjecture 2.2, we must, for c^* a prescribed, arbitrarily large integer, construct a sequence for which $\mathscr{Q}(n,q,z,k)$ grows asymptotically faster than $\frac{n^{c^*}}{q}$. We connect our analysis to Conjecture 2.2 more precisely later, in Theorem 4.14 below.

Definition 4.1. For $c^*\geq 2$ given, we set $q(n)=n^{c^*+1}$. Moreover, we define

$$f(n)\coloneqq\left\lfloor e\cdot n^{1/3}\right\rfloor,$$

set

$$z(n)\coloneqq n-f(n),$$

and finally set

$$k(n)\coloneqq\left(1-\frac{2}{3}\cdot\frac{1}{c^*+1}\right)\cdot f(n).$$

The lower bound on c^* is required for technical reasons, but is harmless. Throughout, we abuse notation by writing q, z and k for $q(n)$, $z(n)$ and $k(n)$.

We begin by recording a straightforward identity, pertaining to these specializations, which we will repeatedly have occasion to use. Indeed,

$$\log_n\left(n^{n-z}\cdot q^{k+z-n}\right)=f(n)+(c^*+1)\cdot\left(1-\frac{2}{3}\cdot\frac{1}{c^*+1}-1\right)\cdot f(n)=\frac{1}{3}\cdot f(n).\tag{14}$$

This identity lets us replace $n^{n-z}\cdot q^{k+z-n}$ with $n^{\frac{1}{3}\cdot f(n)}$ wherever the former quantity appears.

Remark 4.2. While our choice for f above has a degree of arbitrariness, we note that our bounds below critically leverage the asymptotic relationship $f(n)^2 = o(n)$. In particular, the choice $f(n) \coloneqq \lfloor e \cdot n^{1/2} \rfloor$ would not have worked.

There is a technical subtlety implicit in Definition 4.1. First, $q(n)$ might not be a prime power; secondly $k(n)$ might not be an integer. In order to assert the existence of an MDS $[n, k, d]$-code $C \subset \mathbb{F}_q^n$, we need both of these conditions to be true. We could take floors and ceilings, and set for q a prime power *at least* n^{c^*+1}. On the other hand, our estimates below are extremely sensitive, and even such minor acts as taking floors and ceilings are liable to make them false or useless. We thus opt to take a number-theoretic detour.

Lemma 4.3. *There are infinitely many primes $n \geq 1$ such that $k(n)$ is an integer.*

Proof. The statement of the lemma is equivalent to that whereby there are infinitely many primes n such that $3 \cdot (c^* + 1) \mid \lfloor e \cdot n^{1/3} \rfloor$. For each $m \geq 0$, $m = \lfloor e \cdot n^{1/3} \rfloor$ if and only if

$$n \in \left(\left(\frac{m}{e}\right)^3, \left(\frac{m+1}{e}\right)^3 \right). \tag{15}$$

It suffices to show that for each large enough m, each real interval of the above form contains a prime n (i.e., and thus *a fortiori*, those attached to integers m that moreover are multiples of $3 \cdot (c^* + 1)$ in particular do). This is thus a problem pertaining to the distribution of prime numbers.

For integral $m \geq 1$, we write $l(m) \coloneqq \left(\frac{m}{e}\right)^3$ and $u(m) \coloneqq \left(\frac{m+1}{e}\right)^3$ for the lower and upper endpoints, respectively, of the interval (15). We note that

$$u(m) - l(m) = \frac{3m^2 + 3m + 1}{e^3} \geq \frac{m^2}{e^2} = l(m)^{2/3}.$$

A classic result of Ingham [Ing37] (see also e.g. Dudek [Dud16]) establishes that, for each $\varepsilon > 0$, for each large enough x, the interval $(x, x + x^{5/8+\varepsilon})$ contains a prime[2]. Picking $\varepsilon \coloneqq \frac{1}{48}$, we get $\frac{5}{8} + \varepsilon = \frac{31}{48} < \frac{2}{3}$, which implies that

$$\Big(l(m), l(m) + l(m)^{31/48}\Big) \subset \Big(l(m), l(m) + l(m)^{2/3}\Big) \subset \Big(l(m), u(m)\Big).$$

Thus, for each large enough m, the right-hand interval too necessarily contains a prime. □

Remark 4.4. Lemma 4.3 suggests an interesting tradeoff between the exponent α in Definition 4.1—i.e., for which $f(n) = \lfloor e \cdot n^{\alpha} \rfloor$—and the ease with which primes in the intervals $I_m \coloneqq \{n \mid m = \lfloor e \cdot n^{\alpha} \rfloor\}$ might be found. These intervals' widths grow like $O\big(l(m)^{1-\alpha}\big)$ in general. Ingham [Ing37] chronicles the history of efforts to find primes in $(x, x + x^{\theta})$, for $\theta \in (0, 1)$ (of course, as θ decreases, this task becomes harder).

Our techniques seem to work for any $\alpha < \frac{1}{2}$; we chose $\alpha = \frac{1}{3}$ mainly to keep things simple. On the other hand, Lemma 4.3 implies another incidental—though fascinating—reason not to let α get too close to $\frac{1}{2}$. Indeed, if $1 - \alpha = \theta$ approached $\frac{1}{2}$ too closely from above, then the analogue of Lemma 4.3 could become very hard to prove. The strongest known, unconditional result in this direction appears to be that of Baker, Harman and Pintz [BHP01], who show that θ can be taken to be $\frac{21}{40} + \varepsilon$. As Ingham [Ing37] notes, if the *Lindelöf hypothesis* (a consequence of the Riemann hypothesis) is true, then even for $\theta = \frac{1}{2} + \varepsilon$ taken arbitrarily close to $\frac{1}{2}$, the intervals $(x, x + x^{\theta})$ eventually contain primes. The condition $\alpha < \frac{1}{2}$ thus seems—for this further, independent reason—to naturally demarcate the boundary line beyond which our techniques must cease to work.

Remark 4.5. In the statement of Lemma 4.3, we really only need n to be a prime power. It's not clear whether this relaxation stands to make Lemma 4.3 easier to prove (or makes its implicit bounds sharper).

Before proceeding to our main proof, we start with a technical lemma.

Lemma 4.6. *The quantity*

$$A(n) \coloneqq \frac{n^{\frac{1}{3} \cdot f(n)}}{f(n)!}$$

satisfies $\lim_{n \to \infty} A(n) = 0$. *Moreover, for each large enough n,* $A(n) \geq \frac{e^{-1}}{\sqrt{2\pi}} \cdot n^{-\frac{1}{6}}$.

[2] In fact, his work, together with that of his contemporaries, implies that $x^{577/925+\varepsilon}$ is enough.

Proof. First, by (a coarse form of) Robbins' Stirling variant, we have:

$$\sqrt{2\pi \cdot f(n)} \cdot \left(\frac{f(n)}{e}\right)^{f(n)} \leq f(n)! \leq \sqrt{2\pi \cdot f(n)} \cdot \left(\frac{f(n)}{e}\right)^{f(n)} \cdot e^{\frac{1}{12 \cdot f(n)}}.$$

We now bound $\left(\frac{e \cdot n^{1/3}}{f(n)}\right)^{f(n)}$ from both above and below. We note that

$$1 \leq \left(\frac{e \cdot n^{1/3}}{f(n)}\right)^{f(n)} = \left(\frac{e \cdot n^{1/3}}{\lfloor e \cdot n^{1/3} \rfloor}\right)^{f(n)} = \left(1 + \frac{e \cdot n^{1/3} - \lfloor e \cdot n^{1/3} \rfloor}{\lfloor e \cdot n^{1/3} \rfloor}\right)^{f(n)} \leq e; \tag{16}$$

in the last step, we use the inequality $(1+a)^x \leq e^{ax}$, valid whenever a and x are nonnegative.

Combining Stirling's approximation (in the denominator) with the bound (16) just stated, we bound $A(n)$ from both above and below as follows:

$$\frac{e^{-\frac{1}{12f(n)}}}{\sqrt{2\pi \cdot f(n)}} \leq \left(\frac{e \cdot n^{1/3}}{f(n)}\right)^{f(n)} \cdot \frac{e^{-\frac{1}{12f(n)}}}{\sqrt{2\pi \cdot f(n)}} \leq A(n) \leq \frac{1}{\sqrt{2\pi \cdot f(n)}} \cdot \left(\frac{e \cdot n^{1/3}}{f(n)}\right)^{f(n)} \leq \frac{e}{\sqrt{2\pi \cdot f(n)}}. \tag{17}$$

The rightmost expression clearly converges to 0, which proves the first of our two claims. As for the second, for each large enough n,

$$\begin{aligned}
A(n) &\geq \frac{e^{-\frac{1}{12f(n)}}}{\sqrt{2\pi \cdot f(n)}} && \text{(by (17).)} \\
&\geq \frac{e^{-\frac{1}{12f(n)}}}{\sqrt{2\pi \cdot e \cdot n^{1/3}}} && \text{(since } \lfloor e \cdot n^{1/3} \rfloor \leq e \cdot n^{1/3}\text{, we can remove the floor from the denominator.)} \\
&\geq \frac{e^{-1}}{\sqrt{2\pi \cdot n^{1/3}}} && \text{(since } f(n) \text{ is unbounded, } e^{-\frac{1}{12f(n)}} \geq e^{-\frac{1}{2}} \text{ for large enough } n.) \\
&= \frac{e^{-1}}{\sqrt{2\pi}} \cdot n^{-\frac{1}{6}}, &&
\end{aligned}$$

a calculation that completes the proof. □

Let us turn back now to our expression $\mathscr{Q}(n,q,z,k)$ (recall (13)). Let us write $\mathscr{B}(n,q,z,k)$ for the part of the estimate coming from the balls, and $\mathscr{I}(n,q,z,k)$ for the part coming from the intersections. In this way we get

$$\mathscr{Q}(n,q,z,k) = \mathscr{B}(n,q,z,k) - \mathscr{I}(n,q,z,k) = \mathscr{B}(n,q,z,k) \cdot \left(1 - \frac{\mathscr{I}(n,q,z,k)}{\mathscr{B}(n,q,z,k)}\right).$$

Our main goal is to prove

Proposition 4.7. *The specialization of $\mathscr{Q}(n,q,z,k)$ to the families defined in Definition 4.1 satisfies*

$$\lim_{n\to\infty} n \cdot \mathscr{Q}(n, q(n), z(n), k(n)) = \infty.$$

We attack this proposition by breaking it into two simpler propositions, both grounded in the above equality.

Proposition 4.8. *The specializations of $\mathscr{B}(n,q,z,k)$ and $\mathscr{I}(n,q,z,k)$ along the functions of Definition 4.1 jointly satisfy*

$$\lim_{n\to\infty} \frac{\mathscr{I}(n, q(n), z(n), k(n))}{\mathscr{B}(n, q(n), z(n), k(n))} = 0.$$

Proposition 4.9. *The specialization of $\mathscr{B}(n,q,z,k)$ along the functions of Definition 4.1 satisfies*

$$\lim_{n\to\infty} n \cdot \mathscr{B}(n, q(n), z(n), k(n)) = \infty.$$

By the algebraic limit theorem, Propositions 4.8 and 4.9, taken jointly, immediately imply Proposition 4.7.

Before diving into Proposition 4.8, we begin by handling a few annoying factors lurking in the expressions $\mathscr{B}(n,q,z,k)$ and $\mathscr{I}(n,q,z,k)$.

Lemma 4.10. *Upon specializing q and z along the functions of Definition 4.1, we obtain* $\lim_{n\to\infty}\left(1-\frac{1}{q}\right)^z = 1$.

Proof. We calculate

$$1 \geq \lim_{n\to\infty}\left(1-\frac{1}{q}\right)^z \geq \lim_{n\to\infty} 1-\frac{z}{q} \geq \lim_{n\to\infty} 1-n^{-c^*} = 1,$$

where the last equality is valid so long as $c^* \geq 1$. □

Lemma 4.11. *In the specialization of Definition 4.1, we further get* $\lim_{n\to\infty}\frac{(n-z)\cdot z^2}{q}=0$ *and* $\lim_{n\to\infty}\frac{(n-z)^2}{2z-n+1}=0$.

Proof. The first among this lemma's claims follows from the calculation

$$0 \leq \lim_{n\to\infty}\frac{(n-z)\cdot z^2}{q} = \lim_{n\to\infty}\frac{f(n)\cdot(n-f(n))^2}{q} \leq \lim_{n\to\infty}\frac{f(n)}{n}\cdot\frac{n^2}{n^{c^*}} = 0,$$

where the last equality holds whenever $c^* \geq 2$, since $f(n)=o(n)$. As for the second, since $f(n)=o(n)$, for each large enough n we have $2z-n+1 = n-2f(n)+1 \geq \frac{1}{2}\cdot n$, and so get

$$0 \leq \lim_{n\to\infty}\frac{(n-z)^2}{2z-n+1} \leq \lim_{n\to\infty}\frac{e^2\cdot n^{2/3}}{n-2f(n)+1} \leq \lim_{n\to\infty} 2e^2\cdot\frac{n^{2/3}}{n} = \lim_{n\to\infty} 2e^2\cdot\frac{1}{n^{1/3}} = 0,$$

which proves what we need. We note that here, $f(n)^2=o(n)$ is necessary (recall Remark 4.2). □

Lemma 4.11 in particular shows that in the setting of Definition 4.1, the hypothesis of Corollary 3.17 eventually becomes fulfilled.

Lemma 4.12. *Upon specializing q and z along the functions of Definition 4.1, we finally get* $\lim_{n\to\infty} R(n,q,z)=1$.

Proof. Recall that

$$R(n,q,z) = \frac{1}{1-\frac{(n-z)\cdot z^2}{q}}\cdot\frac{1}{1-\frac{(n-z)^2}{2z-n+1}}.$$

By the algebraic limit theorem, it suffices to argue that $R(n,q,z)$'s two factors, independently, both converge to 1. Since $\frac{1}{1-x}$ is continuous at $x=0$ and equals 1 there, it suffices in turn to prove that the respective expressions in both of those factors' denominators converge to 0. We just proved exactly this in Lemma 4.11. □

We now turn back to Proposition 4.8.

Proof of Proposition 4.8. Lemmas 4.10 and 4.12 allow us to discard the $R(n,q,z)$-factor from $\mathscr{I}(n,q,z,k)$ and the $\left(1-\frac{1}{q}\right)^z$-factor from $\mathscr{B}(n,q,z,k)$. That task done, it remains for us just to expand multinomials. More precisely:

$$\begin{aligned}
0 \leq \lim_{n\to\infty}\frac{\mathscr{I}(n,q(n),z(n),k(n))}{\mathscr{B}(n,q(n),z(n),k(n))} &= \lim_{n\to\infty}\frac{\frac{1}{2}\cdot\binom{n}{n-z,n-z}\cdot q^{2z+2k-2n}\cdot R(n,q,z)}{\binom{n}{n-z}\cdot q^{z+k-n}\cdot\left(1-\frac{1}{q}\right)^z} && \text{(this is just read directly off of (13).)}\\
&= \frac{1}{2}\cdot\lim_{n\to\infty}\frac{\binom{n}{n-z,n-z}\cdot q^{2z+2k-2n}}{\binom{n}{n-z}\cdot q^{z+k-n}} && \text{(eliminate factors, by Lemmas 4.10 and 4.12.)}\\
&= \frac{1}{2}\cdot\lim_{n\to\infty}\frac{z!}{(n-z)!\cdot(2z-n)!}\cdot q^{z+k-n} && \text{(just consolidate terms.)}\\
&\leq \frac{1}{2}\cdot\lim_{n\to\infty}\frac{n^{n-z}}{(n-z)!}\cdot q^{z+k-n} && \text{(use } \tfrac{z!}{(2z-n)!}=\tfrac{(n-f(n))!}{(n-2f(n))!}\leq(n-f(n))^{f(n)}\leq n^{f(n)}.\text{)}\\
&\leq \frac{1}{2}\cdot\lim_{n\to\infty}\frac{n^{\frac{1}{3}f(n)}}{f(n)!} && \text{(here, we again use (14).)}\\
&= 0. && \text{(this is the first statement of Lemma 4.6.)}
\end{aligned}$$

This completes the proof of the proposition. □

We now turn to Proposition 4.9. We would like to replace $\mathscr{B}(n,q,z,k)$ with the approximation of that quantity we produced in Theorem 3.10. In order to justify this replacement, we need to ensure that that theorem's controller expressions go to 0 under our specialization.

Lemma 4.13. *Upon being specialized along Definition 4.1, the big-$\mathcal{O}$ expressions in Theorem 3.10 all converge to* 0.

Proof. The vanishing of the first three controllers follows from the facts $q = n^{c^*+1}$ and $\lim_{n\to\infty} \frac{z}{n} = \lim_{n\to\infty} \frac{n-f(n)}{n} = 1$. As for the final controller, we see that:

$$0 \leq \lim_{n\to\infty} \frac{(n-z)^2 \cdot z}{n^2} \leq \lim_{n\to\infty} \frac{e^2 \cdot n^{2/3} \cdot (n - f(n))}{n^2} \leq \lim_{n\to\infty} \frac{e^2 \cdot n^{5/3}}{n^2} = \lim_{n\to\infty} \frac{e^2}{n^{1/3}} = 0,$$

which proves the lemma. Here, we again need $f(n)^2 = o(n)$ (recall again Remark 4.2). □

We're finally ready to prove Proposition 4.9.

Proof of Proposition 4.9. In light of Lemma 4.13, $\mathscr{B}(n, q(n), z(n), k(n))$ and the estimate of that quantity put forth by Theorem 3.10 are asymptotically equal. Unwinding what that statement means, we see that for each constant strictly between 0 and 1—say, $\frac{1}{\sqrt{2}}$—and each large enough n, we get the guarantee

$$\mathscr{B}(n, q(n), z(n), k(n)) \geq \frac{1}{\sqrt{2}} \cdot \sqrt{\frac{n}{z}} \cdot \frac{n^{n-z} \cdot q^{k+z-n} \cdot e^{-\frac{(n-z)^2}{n} - \frac{z}{q}}}{(n-z)!}.$$

Now clearly, $\sqrt{\frac{n}{z}} \geq 1$ for each n (and in fact this quantity converges to 1 from above). For each large enough n, we moreover have the coarse lower-bound $e^{-\frac{(n-z)^2}{n} - \frac{z}{q}} \geq \frac{1}{\sqrt{2}}$ (this quantity converges to 1 from below).

Combining these observations with (14), we get the further lower-bound:

$$\frac{1}{\sqrt{2}} \cdot \sqrt{\frac{n}{z}} \cdot \frac{n^{n-z} \cdot q^{k+z-n} \cdot e^{-\frac{(n-z)^2}{n} - \frac{z}{q}}}{(n-z)!} \geq \frac{1}{2} \cdot \frac{n^{\frac{1}{3} \cdot f(n)}}{(f(n))!}.$$

Applying Lemma 4.6, we conclude still further that, for each large enough n,

$$\mathscr{B}(n, q(n), z(n), k(n)) \geq \frac{e^{-1}}{\sqrt{8\pi}} \cdot n^{-\frac{1}{6}}.$$

Multiplying this quantity by n, we get $\frac{e^{-1}}{\sqrt{8\pi}} \cdot n^{\frac{5}{6}}$, which converges to infinity. □

In particular, our proof of Proposition 4.7 is complete.

Theorem 4.14. *For each integer $c^* \geq 2$, there is an infinite sequence of block lengths n, alphabets $q = q(n)$, message lengths $k = k(n)$, and ball radii $z = z(n)$, such that, for each n in the family, each MDS $[n, k, d]$-code $C \subset \mathbb{F}_q^n$ satisfies*

$$\mathsf{err}(C, z) > \frac{n^{c^*}}{q}.$$

Proof. By Proposition 4.7, we can find some N such that for each $n \geq N$,

$$n \cdot \mathscr{Q}(n, q, z, k) > 2.$$

For each $n \geq N$ for which n is prime and $k(n)$ is an integer—of which there are infinitely many, by Lemma 4.3—each MDS $[n, k, d]$-code (e.g., the Reed–Solomon code $C \coloneqq \mathsf{RS}_{\mathbb{F}_q}[n, k]$) satisfies

$$\begin{aligned}
\mathsf{err}(C, z) &\geq \frac{q-1}{q} \cdot \Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z] && \text{(this holds universally as long as } z < n - k\text{, by Theorem 2.5.)} \\
&\geq \frac{q-1}{q} \cdot \mathscr{Q}(n, q, z, k) && \text{(this is just (13).)} \\
&> \frac{q-1}{q} \cdot \frac{2}{n} && \text{(since } n \geq N\text{; see just above.)} \\
&\geq \frac{1}{n} && \text{(this is true as long as } q \geq 2\text{.)} \\
&= \frac{n^{c^*}}{q}. && \text{(in Definition 4.1, we defined } q \coloneqq n^{c^*+1}\text{.)}
\end{aligned}$$

This completes the proof. □

Theorem 4.14 shows that Conjecture 2.2 is false.

5 The Amended Conjecture

Here, we suggest a speculative amendment of the main Conjecture 2.2. Our main innovation is to force both the message length k and the *shortfall* $h = n - k - z$—that is, the difference between the covering radius and the proximity parameter—to grow like $\Omega(n)$ (cf. e.g. Guruswami [Gur06, § 2.1]). Our confidence that Conjecture 5.1 is true as written is rather low; we suspect that yet further restrictions on q, η, ρ and n might be necessary.

Conjecture 5.1. *For each choice of constants $\rho \in (0,1)$ and $\eta \in (0, 1-\rho)$, there exist constants $c \geq 1$ and $K > 0$ (depending on ρ and η) such that, for each large enough n and each Reed–Solomon code $C = \mathsf{RS}_{\mathbb{F}_q}[n,k]$, if $k \geq \lfloor \rho \cdot n \rfloor$ and $z \in \{0, \ldots, n - k - \lceil \eta \cdot n \rceil\}$, then C features proximity gaps with respect to z and the false witness bound*

$$\mathsf{err}(C, z) \coloneqq K \cdot \frac{n^c}{q}.$$

Ben-Sasson, Carmon, Ishai, Kopparty and Saraf's Conjecture 2.2 is essentially that which would arise from our Conjecture 5.1 if that conjecture's constants ρ and η were fixed at 0 (a parameterization which we of course forbid).

Our use of floors and ceilings in Conjecture 5.1 might look odd. Our convention has the following neat property in the UDR. For each $\rho \in (0,1)$, for $\eta \coloneqq \frac{1-\rho}{2}$ and each n, upon setting $k \coloneqq \lfloor \rho \cdot n \rfloor$, we get $\left\lfloor \frac{d-1}{2} \right\rfloor = n - k - \lceil \eta \cdot n \rceil$.

Remark 5.2. We note that our use of the constant K and our caveat that n be "large enough" are not essential, and can't change Conjecture 5.1's truth value. We incorporate those measures merely to relieve the parameter c from the burden of accommodating even tiny, trivial members of the family—those for which n remains on the order of $\frac{1}{\rho}$ or $\frac{1}{\eta}$, say. For exactly those codes, the results of this paper are liable to force c to be unrepresentatively large. As n grows, we expect the error probability to (rapidly) diminish, in the setting of Conjecture 5.1. That conjecture thus demands solely that K and c control the eventual behavior of the family, which is more likely to be tractable.

Remark 5.3. Though Conjecture 5.1's symbols ρ and η bear a superficial resemblance to the identically-named symbols in Conjecture 2.2, they actually play very different roles just above. For us, ρ and η must be fixed in advance, and our constants K and c, as well as the implicit N as of which the conjecture becomes effective, depend on them. In Conjecture 2.2, ρ and η vary as C does, and c_1 and c_2 are global and independent of them.

Example 5.4. In the case $\rho \in (0,1)$ arbitrary and $\eta \geq \frac{1-\rho}{2}$, Conjecture 5.1 is true for $c = 1$, $K = 1$ and every $n \geq 1$. Indeed, this is just proximity gaps in the unique decoding radius, a case taken care of by [BCIKS23, Thm. 4.1]. In fact, that error is essentially sharp (see Diamond and Gruen [DG25, Rem. 3]).

Moving into the list decoding regime, [BCIKS23, Thm. 1.2] further proves Conjecture 5.1 with $\rho \in (0,1)$ arbitrary and $\eta \in \left(\sqrt{\rho} - \rho, \frac{1-\rho}{2}\right)$ for parameters $c = 2$, $K = \frac{\rho^2}{(2\min(\sqrt{\rho}-\rho-\eta, \frac{\sqrt{\rho}}{20}))^7}$ and every $n \geq 1$.

We discuss a final source of informal evidence for Conjecture 5.1, connected to the condition (2). We recall that that relationship captures—heuristically—the state of affairs whereby the union of all the radius-z balls fills $\mathbb{F}_q^n$ (and so proximity gaps can't possibly hold).

More generally, upon relativizing the right-hand side of (2), we obtain the expression

$$(\rho + \eta) \cdot \log_2\left(\frac{e}{\rho}\right) - \eta \cdot \log_2(q). \tag{18}$$

The expression (18) can be understood in the following way. For $\rho \in (0,1)$ understood as a constant, as $\eta \in (0, 1-\rho)$ varies, (18)—or more properly, the minimum of that quantity and 0—heuristically captures the *base-two logarithm* of the proportion of $\mathbb{F}_q^n$ collectively covered by the radius-z Hamming balls, where $z = n - k - \lceil \eta \cdot n \rceil$. We can see that, for each fixed (and large-enough) q, this logarithm decreases linearly in η (with a slope that moreover becomes steeper as q grows). Moreover, this line crosses zero around

$$\eta \approx \frac{\rho \cdot \log_2\left(\frac{e}{\rho}\right)}{\log_2(q) - \log_2\left(\frac{e}{\rho}\right)} \tag{19}$$

(this is a slightly more accurate variant of (3)). Thus, for $\rho \in (0,1)$ constant, the proportion of $\mathbb{F}_q^n$ collectively covered by the radius-z balls drops exponentially fast in $\eta = \frac{n-k-z}{n}$, *once* that proportion becomes less than 1.

This line of reasoning suggests that the probability $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z]$ behaves like a threshold function in z. That is, as $z \in \{0, \ldots, n-k-1\}$ varies, $\Pr_{\mathbf{u} \leftarrow \mathbb{F}_q^n}[d(\mathbf{u}, C) \leq z]$ is either extraordinarily close to 0 or extraordinarily close to 1; moreover, the relative threshold at which this very probability drops is given roughly by (19).

How does (19) change as n grows? Since $q \geq n$ and everything else in (19) is constant, $\eta = O\left(\frac{1}{\log n}\right)$. We thus heuristically conclude that for $\eta = \Omega(1)$—as Conjecture 5.1 demands that it be—the proportion of $\mathbb{F}_q^n$ covered by the radius-z balls is liable to be virtually zero, and that the techniques of this work aren't likely to work.

References

[AHIV23] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. "Ligero: lightweight sublinear arguments without a trusted setup". In: *Designs, Codes and Cryptography* (2023). DOI: 10.1007/s10623-023-01222-8.

[BCIKS23] Eli Ben-Sasson, Dan Carmon, Yuval Ishai, Swastik Kopparty, and Shubhangi Saraf. "Proximity Gaps for Reed–Solomon Codes". In: *Journal of the ACM* 70.5 (Oct. 2023). DOI: 10.1145/3614423.

[BHP01] R. C. Baker, G. Harman, and J. Pintz. "The Difference Between Consecutive Primes, II". In: *Proceedings of the London Mathematical Society* 83.3 (Nov. 2001), pp. 532–562. ISSN: 0024-6115. DOI: 10.1112/plms/83.3.532.

[CHLL97] Gérard Cohen, Iiro Honkala, Simon Litsyn, and Antoine Lobstein. *Covering Codes*. Vol. 54. North-Holland Mathematical Library. North-Holland, 1997.

[DG25] Benjamin E. Diamond and Angus Gruen. "Proximity Gaps in Interleaved Codes". In: *IACR Communications in Cryptology* 1.4 (Jan. 13, 2025). ISSN: 3006-5496. DOI: 10.62056/a0ljbkrz.

[Dud16] Adrian W. Dudek. "An explicit result for primes between cubes". In: *Functiones et Approximatio Commentarii Mathematici* 55.2 (Dec. 2016), pp. 177–197. DOI: 10.7169/facm/2016.55.2.3.

[Gur06] Venkatesan Guruswami. *Algorithmic Results in List Decoding*. Vol. 2. Foundations and Trends in Theoretical Computer Science 2. now publishers, 2006. DOI: 10.1561/0400000007.

[Ing37] A. E. Ingham. "On the Difference Between Consecutive Primes". In: *The Quarterly Journal of Mathematics* os-8.1 (Jan. 1937), pp. 255–266. ISSN: 0033-5606. DOI: 10.1093/qmath/os-8.1.255.

[MS74] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. Vol. 16. North-Holland Mathematical Library. North-Holland Publishing Company, 1974.