This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

List Decoding Random Euclidean Codes and Infinite Constellations

Yihan Zhang*, Shashank Vatedka†
*Institute of Science and Technology Austria
†Department of Electrical Engineering, Indian Institute of Technology Hyderabad

***Abstract*—We study the list decodability of different ensembles of codes over the real alphabet under the assumption of an omniscient adversary. It is a well-known result that when the source and the adversary have power constraints P and N respectively, the list decoding capacity is equal to $\frac{1}{2}\log\frac{P}{N}$. Random spherical codes achieve constant list sizes, and the goal of the present paper is to obtain a better understanding of the smallest achievable list size as a function of the gap to capacity. We show a reduction from arbitrary codes to spherical codes, and derive a lower bound on the list size of typical random spherical codes. We also give an upper bound on the list size achievable using nested Construction-A lattices and infinite Construction-A lattices. We then define and study a class of infinite constellations that generalize Construction-A lattices and prove upper and lower bounds for the same. Other goodness properties such as packing goodness and AWGN goodness of infinite constellations are proved along the way. Finally, we consider random lattices sampled from the Haar distribution and show that if a certain conjecture that originates in analytic number theory is true, then the list size grows as a polynomial function of the gap-to-capacity.**

I. Introduction

In this paper, we study a problem of communication in the presence of a power-constrained adversary. Consider a point-to-point communication setup where a sender wants to communicate a message $m \in \{0,1\}^{nR}$ of nR bits to a receiver through a real-valued channel corrupted by a malicious *omniscient* adversary. The transmitter sends a signal[1] $\underline{x} \in \mathbb{R}^n$ in n channel uses. The adversary can observe the transmitted signal and corrupt it by adding a noise vector $\underline{s} \in \mathbb{R}^n$, which is allowed to be any noncausal function of $\underline{x}$ and the transmission protocol. The sender and the adversary have power constraints of P and N respectively, i.e., we impose the restriction that $\|\underline{x}\| \leqslant \sqrt{nP}$ and $\|\underline{s}\| \leqslant \sqrt{nN}$.[2] The goal is to design a transmission scheme that provides a high data rate R while ensuring a zero probability of error of decoding the transmited message from $\underline{y} = \underline{x}+\underline{s}$ at the receiver. This problem turns out to be equivalent to the classical sphere packing problem or *one of designing codes with a minimum-distance constraint*, where we want the maximum R such that there is a set of 2^{nR} points within a ball of radius $\sqrt{nP}$ with the property that every pair of points is spaced $2\sqrt{nN}$ apart. Finding the capacity of this channel remains an open problem, but nonmatching upper and lower bounds are known [2], [3].

Adversarial channels in the presence of omniscient adversaries can be viewed as an information-theoretic interpretation of the classical problem of designing error-correcting codes with minimum-distance constraints. Alternatively, they can be viewed as problems of communication with zero error in the presence of a power-constrained jammer with very strong capabilities. Besides the encoder/decoder and codebook, the adversary/jammer is also allowed to design his noise/jamming signal based on the full knowledge of the transmitted signal. Codes resilient to such errors are of interest in applications where there is a threat of potentially strong adversarial attacks. These codes can also be used over additive noise channels where the power of the noise signal is known, but the exact noise distribution could be unknown, and potentially even be correlated with the transmitted signal. Such problems are also studied under the broad framework of arbitrarily varying channels (AVCs) in the literature [4]. This is in contrast to the Discrete Memoryless Channel (DMC) model in classical Shannon theory, where the noise is independent of the transmitted signal, obeys a fixed (known) law, and a small but nonzero probability of decoding error is allowed.

We study a slight variant of this problem, where instead of uniquely decoding the transmitted message m, the receiver attempts to recover a list of L codewords with the guarantee that the transmitted codeword lies in this list. This is called the *list decoding* problem, also known as *multiple packing*, the latter asks for the maximum R such that there is a set of 2^{nR} points in a ball of radius $\sqrt{nP}$ such that there is no point in $\mathbb{R}^n$ to which there are more than L points from the set within distance $\sqrt{nN}$. This problem is well studied at least in the context of binary adversarial channels [5]. In this paper, we attempt to systematically study upper and lower bounds on achievable list sizes for various ensembles of random codes for the real channel.

List decoding for adversarial channels is an interesting problem in its own right, but can also be a very useful tool in several other problems. For instance, Langberg [6] showed that if there exists a coding scheme that achieves a list size that is at most polynomial in the blocklength n, then even a

[1]This work was done in part when Shashank Vatedka was at the Chinese University of Hong Kong, where he was supported by CUHK Direct Grants 4055039 and 4055077. He would like to acknowledge funding from a seed grant offered by IIT Hyderabad and the Start-up Research Grant (SRG/2020/000910) from the Science and Engineering Research Board (SERB), India. Yihan Zhang has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 682203-ERC-[Inf-Speed-Tradeoff].

This paper was presented in part at 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France [1].

[1]We use underline to denote vectors of length n.

[2]Without further specification, $\|\cdot\|$ denotes the L^2-norm.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

small amount of shared secret key (just about $\Theta(\log n)$ bits kept secret from the adversary) between the sender-receiver pair suffices to ensure that the true message can be uniquely decoded by the receiver. List decoding can also serve as a useful proof technique for obtaining bounds on the capacities of other adversarial channels [7], [8], [9].

For an adversarial channel with L^2 power constraints, it is known (see e.g. [8, Appendix D]) that if the transmission rate R is greater than $\frac{1}{2}\log\frac{P}{N}$, then no coding scheme can achieve subexponential (in n) list sizes. On the other hand, it is also known ([8, Appendix D]) that random spherical codes of rate $R < \frac{1}{2}\log\frac{P}{N}$ can achieve constant (in n) list sizes. We can therefore call $\frac{1}{2}\log\frac{P}{N}$ to be the *list decoding capacity* of this channel. Once this is established, it is of interest to find the least possible list sizes that are achievable as a function of $\delta := \frac{1}{2}\log\frac{P}{N} - R$. We will show that in order to find the order-optimal list sizes as a function of the rate gap to capacity, it suffices to only study spherical codes, where all codewords $\underline{x}$ have norm $\|\underline{x}\| = \sqrt{nP}$. It is known that random spherical codes have list sizes upper bounded by $\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})$. We show that "typical" random spherical codes have list sizes which grow as $\Omega(1/\delta)$. In an attempt to devise more "practical" coding schemes that achieve the list decoding capacity, we look for structured codes that can guarantee small list sizes. Specifically, we investigate a class of nested lattice codes and find lower bounds on the list size. We show that random nested Construction-A lattices achieve list sizes $2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})}$. To the best of our knowledge, this is the first such result which shows that lattice codes can achieve constant list sizes. However, the list sizes are exponentially worse than the list sizes for random spherical codes. We conjecture that there exist lattice codes that achieve list sizes of $\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})$ and provide some heuristic calculations to support this.

We then relax the power constraint of the transmitter and study the list decodability of *infinite constellations* (ICs). Infinite constellations generalize lattices, and to the best of our knowledge, were first studied systematically in the context of channel coding by Poltyrev [10]. Poltyrev showed that there exist ICs that are good codes for the additive white Gaussian noise (AWGN) channel. In this paper, we introduce an ensemble of periodic infinite constellations and study upper and lower bounds on the list size of typical ICs. A list decodable code for the power-constrained (for both the transmitter and the adversary) adversarial channel can be obtained by taking the intersection of the IC with a ball of radius $\sqrt{nP}$. We show that the code obtained by taking this intersection achieves list size $\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})$.

Note. This paper is a substantially extended version of the 5-page conference paper [1] which appeared in the proceedings of ISIT 2019. In [1], results were stated without formal proofs or with only proof sketches. The current paper includes full proofs that were omitted in the conference version and novel results including the list decoding capacity theorem for ICs (Theorem 24), covering goodness of ICs (Proposition 43), and list size upper bound for Haar lattices (Lemma 39) under a Poisson heuristic (Heuristic 38).

A. Overview of our results

Let us now formally describe the problem. The sender encodes a message $m \in \{0,1\}^{nR}$ into a codeword $\underline{x}$ in $\mathbb{R}^n$ which is intended for the receiver. The sender has a transmit power constraint, which is modeled by demanding that the L^2 norm $\|\underline{x}\|$ must be no larger than $\sqrt{nP}$ for some $P > 0$. The transmission is observed noncausally by an adversary who corrupts the transmitted vector by adding a noise vector $\underline{s}$ to $\underline{x}$. The adversary has a power constraint of N, which means that $\|\underline{s}\| \leqslant \sqrt{nN}$ for some $N > 0$. However, $\underline{s}$ is allowed to otherwise be any function of $\underline{x}$ and the codebook. The receiver obtains $\underline{y} = \underline{x} + \underline{s}$. The list decoder takes $\underline{y}$ as input and outputs a list of L messages[3], and an error is said to have occurred if the true message m is not in this list.

Definition 1 (List decodability over $\mathbb{R}$)**.** Let $P, N \in \mathbb{R}_{>0}$ and $L \in \mathbb{Z}_{>0}$. We say that a code $\mathcal{C} \subset \mathbb{R}^n$ is (P, N, L)-*list decodable* if

- The code satisfies a maximum power constraint of P, i.e., we have $\|\underline{x}\|_2^2 \leqslant nP$ for all $\underline{x} \in \mathcal{C}$.
- An omniscient adversary with power N cannot enforce a list size greater than L, i.e., for all $\underline{x} \in \mathcal{C}$ and all $\underline{s} \in \mathcal{B}(0, \sqrt{nN})$,[4] we have $|\mathcal{C} \cap \mathcal{B}(\underline{x} + \underline{s}, \sqrt{nN})| \leqslant L$.

The *rate* of $\mathcal{C}$ is defined as $R(\mathcal{C}) := \frac{1}{n}\log|\mathcal{C}|$.[5]

A rate $R \in \mathbb{R}$ is said to be *achievable* for (P, N, L)-list decoding if for infinitely many n, there exist codes $\mathcal{C} \subset \mathbb{R}^n$ having rate $R(\mathcal{C}) \geqslant R$ that are (P, N, L)-list decodable.

Remark 1. Another version of list decodability requires $|\mathcal{C} \cap \mathcal{B}(\underline{y}, \sqrt{nN})| \leqslant L$ *for every* $\underline{y} \in \mathbb{R}^n$, not just for $\underline{y} = \underline{x} + \underline{s}$ for some $\underline{x} \in \mathcal{C}$ and $\underline{s} \in \mathcal{B}(0, \sqrt{nN})$. However, these two definitions are equivalent. This is because for $\underline{y}$ that does not result from any codeword via a noise vector of length at most $\sqrt{nN}$, clearly $\mathcal{C} \cap \mathcal{B}(\underline{y}, \sqrt{nN}) = \varnothing$ and the condition $|\mathcal{C} \cap \mathcal{B}(\underline{y}, \sqrt{nN})| \leqslant L$ trivially holds. Therefore, it does not make a difference in the definition to include $\underline{y}$ that is not a feasible channel output.

In the definition above, we do not prohibit L from being a function of n. In many applications, it suffices to have list sizes that grow as $\mathcal{O}(n^\gamma)$ for a suitably small γ. However, in this paper, we aim for constant list sizes.

Definition 2 (List decoding capacity)**.** Fix any $P, N > 0$. We say that $C(P, N)$ is the *list decoding capacity* if for every $\delta > 0$, there exists a $\gamma > 0$ such that $C(P, N) - \delta$ is achievable for $(P, N, \mathcal{O}(n^\gamma))$-list decoding, and for every $\delta > 0$, there exist no codes of rate $C(P, N) + \delta$ which are $(P, N, 2^{o(n)})$-list decodable.

The following result is folklore, and a proof can be found in [8].

[3]We sometimes abuse terminology and interchangeably talk about lists of messages and lists of codewords. This does not introduce confusion since in this paper we are only concerned with codes whose encoder is a one-to-one map, i.e., a deterministic encoder.

[4]Here $\mathcal{B}(\underline{c}, r)$ denotes an L^2 ball in $\mathbb{R}^n$ of radius r centered at $\underline{c}$. We at times also write $\mathcal{B}^n(\underline{c}, r)$ to emphasize the ambient dimension.

[5]Here log is short for $\log_2$.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

Theorem 1 (Folklore, [8])**.** *For any $P, N > 0$,*

$$C(P,N) = \left[\frac{1}{2}\log\frac{P}{N}\right]^+.$$ [6]

Again, we are in search of structured ensembles of codes achieving ideally the same list decoding performance as random codes. This problem is not as extensively studied as in the finite field case.

The class of problems that we are interested in is the following:

- Suppose that we desire a target rate $R = C(P,N) - \delta$, for some small $\delta > 0$. Then what is the smallest list size L that we can achieve? Specifically, we are interested in the dependence of L on δ.
- What are the fundamental lower bounds on the list size for a fixed δ?
- If we restrict ourselves to structured codes, e.g., nested lattice codes [11], then what list sizes are achievable?

It was shown in [8] that $\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)$ list sizes are achievable using random spherical codes. If we define $\mathcal{S}^{n-1}(0,\sqrt{nP}) := \{\underline{x} \in \mathbb{R}^n : \|\underline{x}\| = \sqrt{nP}\}$ to be the $(n-1)$-dimensional sphere of radius $\sqrt{nP}$, then

Lemma 2 ([8])**.** *Let $P > N > 0$. Fix any $\delta > 0$, and let $R := C(P,N) - \delta$. Let $\mathcal{C}$ be a random codebook of rate R obtained by choosing the codewords independently and uniformly over $\mathcal{S}^{n-1}(0,\sqrt{nP})$, then*

$$\Pr\left[\mathcal{C} \text{ is not } \left(P, N, \mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)\right)\text{-list decodable}\right] \leqslant 2^{-\Omega(n)}.$$

Our contributions for (P,N,L)-list decoding are summarized as follows.

- We derive lower bounds on the list size of random spherical codes. We show that if $R = C(P,N) - \delta$, then L grows as $\Omega(1/\delta)$ with high probability (whp).
- We then investigate the achievable list sizes for random nested lattice codes, and show that if $R = C(P,N) - \delta$, then $L = 2^{\mathcal{O}\left(\frac{1}{\delta}\log^2\frac{1}{\delta}\right)}$ is achievable using Construction-A lattices.
- Conditioned on a conjecture (Conjecture 37) for random lattices, we provide heuristic calculations which suggest that lattice codes might achieve list sizes that grow as $\mathcal{O}(\mathrm{poly}(1/\delta))$.

We then perform a systematic study of the problem of list decoding infinite constellations in $\mathbb{R}^n$. An *infinite constellation* is defined as a countable subset of $\mathbb{R}^n$.

Definition 3. An infinite constellation $\mathcal{C} \subset \mathbb{R}^n$ is said to be (N,L)-*list decodable* if for every $\underline{y} \in \mathbb{R}^n$, we have

$$|\mathcal{C} \cap \mathcal{B}(\underline{y}, \sqrt{nN})| \leqslant L.$$

The *density* of the constellation is defined as

$$\Delta(\mathcal{C}) := \limsup_{a\to\infty} \frac{|\mathcal{C} \cap [-a/2, a/2]^n|}{a^n}.$$

[6]Here $[a]^+$ is defined as $\max\{a, 0\}$.

The *normalized logarithmic density (NLD)*, defined as

$$R(\mathcal{C}) := \frac{\log \Delta(\mathcal{C})}{n},$$

is a measure of the "rate" of an infinite constellation. The *effective volume* of $\mathcal{C}$ is defined as $V(\mathcal{C}) = 1/\Delta(\mathcal{C})$ and the *effective radius* $r_{\text{eff}}(\mathcal{C})$ is defined as the radius of a ball having volume equal to $V(\mathcal{C})$.

Remark 2. In general, one cannot replace $\limsup$ with $\lim$ in the definition of $\Delta(\mathcal{C})$. However, in this paper, we are only concerned with periodic ICs, i.e., ICs that are unions of translations of a finite set. For such ICs, the limit does exist. See [12] and [13] for more discussions on the definition of density.

Remark 3. For a periodic IC $\mathcal{C}$, the NLD of $\mathcal{C}$ remains the same if we replace the cube $[-a/2, a/2]^n$ with any centrally symmetric connected compact set $a\mathcal{B} \subset \mathbb{R}^n$ with nonempty interior and compute the NLD in the following way

$$\Delta(\mathcal{C}) = \limsup_{a\to\infty} \frac{|\mathcal{C} \cap a\mathcal{B}|}{a^n \,\mathrm{Vol}(\mathcal{B})}.$$

Definition 4 (List decoding capacity)**.** Fix any $N > 0$. We say that $C(N)$ is the *list decoding capacity* if for every $\delta > 0$, there exists a $\gamma > 0$ such that $C(N) - \delta$ is achievable for $(N, \mathcal{O}(n^\gamma))$-list decoding, and for every $\delta > 0$, there exist no codes of rate $C(N) + \delta$ which are $(N, 2^{o(n)})$-list decodable.

We show that the list decoding capacity for infinite constellations is $C(N) = \frac{1}{2}\log\frac{1}{2\pi eN}$.

Clearly, every lattice is an infinite constellation. We show that if Λ is a random Construction-A lattice with $r_{\text{eff}}(\Lambda) \geqslant \sqrt{nN}2^\delta$ (or $R(\Lambda) \leqslant C(N) - \delta$), then Λ is $(N, 2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})})$-list decodable. We also introduce a class of random periodic infinite constellations $\mathcal{C}$ with $r_{\text{eff}}(\mathcal{C}) = \sqrt{nN}2^\delta$ (or $R(\mathcal{C}) = C(N) - \delta$) which have list sizes that grow as $\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})$. Additionally, we show a matching lower bound on the list size for these random infinite constellations.

Remark 4. A code satisfying the requirements in Definition 1 can be list decoded when used on a channel governed by an *omniscient* adversary with a maximum probability of error constraint. This is because the decoder will output a small list for every transmitted codeword $\underline{x} \in \mathcal{C}$ and every attack vector $\underline{s} \in \mathcal{B}(0, \sqrt{nN})$. For $L = 1$, our problem reduces to the problem of packing nonintersecting balls of radius $\sqrt{nN}$ such that their centers lie within $\mathcal{B}(0, \sqrt{nP})$.

We could relax the problem by assuming that the decoder uses an average probability of error criterion (where the average is over messages picked uniformly at random) and the adversary knows only the codebook but not the transmitted codeword. This models an *oblivious* adversary and the problem was studied by Hosseinigoki and Kosut [14] (see also [15]). They showed that the list decoding capacity for this problem is $\frac{1}{2}\log(1 + P/N)\mathbb{1}\{L > N/P\}$. Variants of list decoding with other types of adversaries have been studied in the literature: see, e.g., [16], [17].

An intermediate model that lies between the omniscient and the oblivious models is the *myopic* model. In this model, we

assume that a myopic adversary sees a noncausal noisy version of the transmitted codeword. This problem was studied in [8].

Bounds in this paper and prior work on list sizes of ensembles of codes/ICs with gap-to-capacity δ are summarized in Table I.

B. Organization of the paper

In Sec. II, we survey the literature on list decoding over finite fields. Notation and prerequisite facts and lemmas are listed in Sec. III and Sec. III-B, respectively. A table of frequently used notation can be found in Appendix A. A lower bound on list sizes of random spherical codes is provided in Sec. IV while some of the calculations are deferred to Appendix B. In Sec. V, we turn to study list decodability of random nested Construction-A lattice codes. For the benefit of the readers who are not familiar with lattices, a quick primer is provided in Sec. V-B. We define infinite constellations in Sec. VI, and give matching upper and lower bounds on list sizes of an ensemble of regular infinite constellations in Sec. VII. Results on other goodness properties of ICs are presented in Appendix C. Finally we give some heuristic results on the list sizes achieved by lattice codes. We recall the Haar distribution on the space of lattices in Sec. VIII, then introduce two important integration formulas by Siegel and Rogers in Sec. IX-A and their improvements in Sec. IX-B. We prove a list size upper bound conditioned on a conjecture in Sec. X-A. We conclude the paper in Sec. XI with several open problems.

II. Prior Work

A. Prior work on list decoding over finite fields

Given a prime power q and $R \in (0,1)$, how can one construct a subset $\mathcal{C}$ of $\mathbb{F}_q^n$ of size q^{nR} such that the points in $\mathcal{C}$ are as far apart (in Hamming distance) as possible? This question is motivated by communication through noisy channels and is studied under different notions of "far apart". Consider a transmitter who wishes to convey an arbitrary q-ary message of length nR to a receiver through a noisy channel. To protect the information against noise, the transmitter can add some redundancy and send a coded version of the message through the channel. Classical coding theory is devoted to the study of the situation where the codeword is a length-n vector over $\mathbb{F}_q$ and the adversary who has access to the transmitted codeword is allowed to change any np symbols where $0 < p < 1$ is a constant. The receiver is then required to figure out the original message given a maliciously corrupted word. For fixed q and p, the goal is to design an as-large-as-possible set of codewords so as to get as-much-as-possible information through in n uses of the channel, while ensuring that the receiver can decode the message correctly under any legitimate attack by the adversary. We are interested in the asymptotic behaviour of the throughput as the blocklength n grows.

It is not hard to see that the above question is equivalent to the question of determining the optimal density of packing Hamming balls of radii np in $\mathbb{F}_q^n$. The best possible density R is widely open. People thus consider relaxed versions of this problem. Instead of asking the receiver to output a unique correct message, we allow him to output a list of L messages which is guaranteed to contain the correct one. For fixed q, there is now a tradeoff among three quantities: R, p and L. The question is nontrivial only when L is required to be small, otherwise outputting all q^{nR} messages is always a valid scheme. This problem is called *list decoding*. It is sometimes also referred to as *multiple packing* since it can be alternatively thought of as packing balls such that they can overlap but there are no more than L balls on top of any point in the space. Let d_{H} denote the Hamming distance and let $\mathcal{B}_{\mathrm{H}}^n(\underline{y}, np) \coloneqq \{\underline{x} \in \mathbb{F}_q^n : d_{\mathrm{H}}(\underline{y}, \underline{x}) \leqslant np\}$ denote the Hamming ball of radius np centered at $\underline{y}$.

Definition 5 (List decodability over $\mathbb{F}_q$)**.** A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is said to be (p, L)*-list decodable* if for any $\underline{y} \in \mathbb{F}_q^n$, $\left|\mathcal{C} \cap \mathcal{B}_{\mathrm{H}}^n(\underline{y}, np)\right| \leqslant L$.

It turns out that such a relaxation indeed makes the problem easier. In this case, we entirely understand the information-theoretic limit of list decoding. Specifically, Zyablov and Pinsker [18] showed:

Theorem 3 (List decoding capacity over $\mathbb{F}_q$, [18])**.** *For any constant $\delta > 0$, any $0 < p < 1 - 1/q$ and any n large enough, there exists a $(p, 1/\delta)$-list decodable code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate $1 - H_q(p) - \delta$; on the other hand, any code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate $1 - H_q(p) + \delta$ is not (p, L)-list decodable unless $L = q^{\Omega(n\delta)}$.*

The sharp threshold $1 - H_q(p)$ around which the list size exhibits a phase transition is known as the list decoding capacity, denoted by C. The stunning point of the above theorem is that the list size can be made a constant, independent of n, while a rate close to the list decoding capacity is still achieved.

Throughout the paper, we use δ to denote the gap between the rate that the code is operating at and the list decoding capacity, i.e., $R = C - \delta$.

Let τ denote the gap between the adversary's power and the list decoding radius $1 - 1/q$, i.e., $p = 1 - 1/q - \tau$.

First note that expressing the rate as a function of the list size is equivalent to expressing the list size as a function of the gap to capacity. Lower (resp. upper) bounds on the rate naturally translate to upper (resp. lower) bounds on the list size and vice versa. Indeed, for any increasing function f, claiming that a rate $R \geqslant C - \frac{1}{f(L)}$ can be achieved by a (p, L)-list decodable code is equivalent to claiming the existence of a rate $R = C - \delta$ code whose list size is at most $L \leqslant f^{-1}(1/\delta)$ under an adversary with power budget p. We will state prior results and our results only in the second form.

The aforementioned list decoding capacity theorem (Theorem 3) is obtained via standard random coding argument. Indeed, it is well-known and easy to show that the list size of a random code (of which each codeword is sampled uniformly and independently) of rate $1 - H_q(p) - \delta$ against a power-p adversary is at most $1/\delta$ with high probability (whp). It also turns out [19], [20] that $1/\delta$ is the correct scaling for the list size of a random code. Namely, there is an essentially

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

TABLE I: Upper and lower bounds on list sizes of codes and infinite constellations. All bounds hold for the corresponding ensembles with high probability. The parameter δ denotes the gap-to-capacity, where the capacity equals $\frac{1}{2}\log\frac{P}{N}$ in the power-constrained case and $\frac{1}{2}\log\frac{1}{2\pi eN}$ in the power-unconstrained case.

Code/IC	List size	Reference
Uniform spherical code	$L=\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)$	Folklore [8, Appendix D]
	$L=\Omega\left(\frac{1}{\delta}\right)$	Proposition 18
Nested Construction-A lattice code	$L=2^{\mathcal{O}\left(\frac{1}{\delta}\log^2\frac{1}{\delta}\right)}$	Theorem 21
Infinite nested Construction-A lattice	$L=2^{\mathcal{O}\left(\frac{1}{\delta}\log^2\frac{1}{\delta}\right)}$	Theorem 27
(Λ_0,q,M) IC	$L=\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)$	Proposition 28
	$L=\Omega\left(\frac{1}{\delta}\right)$	Lemma 29
Haar lattice code	$L=\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)$	Lemma 39 under Heuristic 38
	$L=\mathcal{O}\left(\left(\frac{1}{\delta}\right)^{1+1/c}\right)$	Lemma 40 under Conjecture 37

matching[7] lower bound $1/\delta$ via second moment method.

Note that $1-H_q(p)$ is also equal to the Shannon's channel capacity of a Binary Symmetric Channel with crossover probability p (BSC(p)). However, as we will elaborate in subsequent sections, this is not the case over the reals.

The above list decoding capacity theorem pinpointed the information-theoretic limit of list decoding which is attained by random codes. In computer science, researchers are generally interested in finding structured or even explicit ensembles of objects with the same asymptotic behavior as the uniformly random ensemble. In the setting of list decoding, given the threshold up to which constant-in-n list size is possible, the ultimate goal is to construct explicit[8] codes with the same list decoding performance as random codes. As an intermediate step which is also interesting in its own right, people aim to reduce the amount of randomness used in the construction and shoot for more "structured" ensembles of codes. A natural candidate is linear codes. However, sadly, even if we restrict our attention to linear codes, its list decodability is still not completely understood. Specifically, a random linear code over $\mathbb{F}_q$ of rate R refers to a random subspace of $\mathbb{F}_q^n$ uniformly sampled from all subspaces of a fixed dimension nR.

Conjecture 4. *For any $\delta>0$, prime power q and $0<p<1-1/q$, a random linear code of rate $1-H_q(p)-\delta$ over $\mathbb{F}_q$ is $(p,1/\delta)$-list decodable whp.*

The conjecture is known to be true over $\mathbb{F}_2$ [20].[9] However, it is open in other cases if we insist the universal constant in the list size to be one. In particular, the conjecture becomes more challenging when we work in the high-noise low-rate regime and in large fields. For instance, consider the following scenario, the adversary's power p is so large that close to $1-1/q$, say $p=1-1/q-\tau$ for a very small $\tau>0$ which can even scale with δ. Then by the continuity of the entropy function, the capacity is very low and in particular vanishes as $\tau\to 0$. In this case, many existing list decodability results for random linear codes degenerate in the sense that the list size blows up when $\tau\to 0$. Another extreme case which is tricky to handle is when the field size q is very large and is potentially an increasing function of $\delta\to 0$ and/or $n\to\infty$. In this case, many techniques in the literature also fail.

From now on, when we talk about large q (i.e., the large field size regime), we refer to q which can scale with $1/\delta$ or n; when we talk about large p or small rate (i.e., the high-noise low-rate regime), we refer to τ which can be a function of δ.

Now we survey a (potentially non-exhaustive) list of work regarding the combinatorial list decoding performance of random linear codes.

1) A classical work by Zyablov and Pinsker [18] showed that a random linear code of rate $1-H_q(p)-\delta$ is $(p,q^{\mathcal{O}(1/\delta)})$-list decodable whp. (See also, e.g., [21] for a proof sketch.)
2) Guruswami, Håstad, Sudan and Zuckerman [22] showed the *existence* of a *binary* linear code of rate $1-H(p)-\delta$ which is $(p,1/\delta)$-list decodable. To this end, they defined a potential function as a witness of non-list decodability and analyzed its evolving dynamics during the process of sampling a basis of the random linear code.
3) Guruswami, Håstad and Kopparty [21] showed that a random linear code of rate $1-H_q(p)-\delta$ is $(p,C_{p,q}/\delta)$-list decodable whp. However $C_{p,q}$ blows up when p gets close to $1-1/q$ or q is large. They used Ramsey-theoretic tools to control low-rank lists. As for high-rank lists, naive bounds suffice.
4) Cheraghchi, Guruswami and Velingker [23] showed that a random linear code of rate $\Omega\left(\frac{\tau^2}{\log^3(q/\tau)\log q}\right)$ is $((1-1/q)(1-\tau),\mathcal{O}(1/\tau^2))$-list decodable with *constant* probability. These parameters are optimal in the low-rate regime *up to polylog factors in $1/\tau$ and q*. In their paper, ideas such as average-radius relaxation, connections to restricted isometry property (RIP) and chaining method were brought into view. These techniques were later extensively explored and significantly developed.
5) Wootters [24] showed that a random linear code of rate $\Omega(\tau^2/\log q)$ is $((1-1/q)(1-\tau),\mathcal{O}(1/\tau^2))$-list decodable whp. This is an improvement on [23] via similar techniques and also fills in the gap in [21] for large p.
6) Rudra and Wootters [25], [26], [27] employed the heavy machinery of generic chaining to provide improved

[7] Actually, Li and Wootters [20] showed that for any constant $\gamma>0$, the list size of a random code is bounded from *below* by $\frac{1-\gamma}{\delta}$ whp.

[8] Rigorously, there are two commonly used definitions of explicitness in the literature. To give an explicit linear code, it suffices to

1) either construct its generator matrix in $\mathrm{poly}(n)$ time deterministically;
2) or compute each entry of its generator matrix in $\mathrm{poly}\log(n)$ time deterministically.

[9] In fact, [20] proved that the list size is only $H(p)/\delta$ (roughly, ignoring some technicalities concerning integrality).

bounds when the field size is very large, say, scaling with $1/\delta$ and even n. The parameter regimes become complicated in this situation and we do not copy their results here.

7) Li and Wootters [20] showed that a random *binary* linear code of rate $1-H(p)-\delta$ is $(p, H(p)/\delta+2)$-list decodable whp for *any* $p \in (0, 1/2)$ and $\delta > 0$. They did so by lifting the existential result in [22] to a high-probability one.
8) Most recently, there is an exciting line of work [28], [29], [30] which makes significant progress on understanding the list sizes of random linear codes. In particular, the authors of these papers characterized the threshold rate (which is difficult to evaluate) of list decodable random linear codes. For *binary* random linear codes, they showed that the list size is (essentially) exactly $H(p)/\delta$.

One can find a summary of aforementioned results in Table II.

Despite a long line of research regarding list decoding, we are far from a complete understanding. Besides attempts towards Conjecture 4, on the negative side, it turns out [19] that there is a matching $\Omega(1/\delta)$ lower bound on the list size of random linear codes. Namely, if we sample a linear code uniformly at random, its list size is $\Omega(1/\delta)$ *whp*. Nonetheless, in general, the best lower bound on list size for an *arbitrary* code is still $\Omega(\log(1/\delta))$ [31], [32], [33], [34], [19]. There is an exponential gap between the best upper and lower bounds even over $\mathbb{F}_2$. Closing this gap is also a long standing open problem. For arbitrarily list decodable codes with list size L, Blinovsky's bound was improved by Ashikhmin, Barg and Litsyn [35] for the $L = 2$ case and by Polyanskiy [36] for the case where $L \geqslant 3$ is odd. For general omniscient adversarial channels beyond bitflip and erasure channels, the critical L^* at which the list-L capacity vanishes has recently been determined by Zhang, Budkuley and Jaggi [37].

B. Prior work on erasure list decoding over finite fields

Similar questions were also posed and studied under the erasure model. In this case, the adversary is allowed to replace any np coordinates of the codeword with question marks.

Definition 6 (List decodability under erasures over $\mathbb{F}_q$)**.** A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is said to be (p, L)-erasure list decodable if for any $\mathcal{T} \subseteq [n]$ of cardinality $(1-p)n$ and any $\underline{y} \in \mathbb{F}_q^{(1-p)n}$, $\left|\{\underline{x} \in \mathcal{C} : \underline{x}|_{\mathcal{T}} = \underline{y}\}\right| \leqslant L$.

It is known that the erasure list decoding capacity is $1-p$, coinciding with the capacity of a Binary Erasure Channel with erasure probability p (BEC(p)).

Theorem 5 (List decoding capacity under erasures over $\mathbb{F}_q$, [5], Theorem 10.3, 10.8)**.** *For any small constant $\delta > 0$, any $0 < p < 1$ and any n large enough, there exists a $(p, \mathcal{O}(1/\delta))$-erasure list decodable code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate $1-p-\delta$; on the other hand, any code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate $1-p+\delta$ is not (p, L)-erasure list decodable unless $L = q^{\Omega(n\delta)}$.*

The achievability part again follows from a random coding argument and the scaling $\Theta(1/\delta)$ of the list size of a random code is tight whp via a second moment computation [19]. For general codes, it can be shown that the list size is at least $\Omega(\log(1/\delta))$ using an erasure version of the punctured Plotkin-type bound (see, e.g., [5], Theorem 10.8).

On the other hand, if we restrict our attention to *linear* codes, the situation seems a little worse. The list size of a random linear code turns out to be $2^{\Theta(1/\delta)}$ whp (see, e.g., [5], Theorem 10.6 for an upper bound and [19] for a matching lower bound). Intuitively, for a linear code $\mathcal{C}$, the list $\{\underline{x} \in \mathcal{C} : \underline{x}|_{\mathcal{T}} = \underline{y}|_{\mathcal{T}}\}$ corresponding to a received word $\underline{y} \in (\mathbb{F}_q \cup \{?\})^n$ with $(1-p)n$ unerased locations in $\mathcal{T} \subseteq [n]$ forms an affine subspace. The list size is therefore exponential in the rank of the list. For general linear codes, it can be shown that the list size is at least $\Omega(1/\delta)$ using a connection to generalized Hamming weights [38]. Although we do not have a provable separation working uniformly in all parameter regimes, it is believed that the list size of linear codes is larger than that of nonlinear codes under erasure list decoding. Narrowing down the exponential gap for either general codes or linear codes seems to be a particularly tricky task.

Upper and lower bounds on list sizes of ensembles of random codes and arbitrary codes are listed in Table III for comparison.

It is worth mentioning that recently there are several breakthroughs towards explicit constructions of "good" codes in the high-noise low-rate regime using tools from pseudorandomness. Ta-Shma [39] constructed an *explicit* δ-balanced[10] binary linear code of rate $\Omega(\delta^{2+\gamma})$ where $\gamma = \Theta\left(\left(\frac{\log\log\frac{1}{\delta}}{\log\frac{1}{\delta}}\right)^{1/3}\right) = o(1)$, *almost* matching the Gilbert–Varshamov bound in this regime. Ta-Shma's beautiful and ingenious construction is done by casting the problem in the language of δ-balanced sets and analyzing a *long* random walk on a carefully constructed expander graph. His result is concerned with explicit codes with good distances. A more relevant result to our topic is an explicit construction of a *near-optimal* erasure list decodable code [40]. The authors viewed the problem as constructing explicit dispersers and managed to construct an explicit binary *nonlinear* $(1-\tau, \mathrm{poly}\log\frac{1}{\tau})$-erasure list decodable code of rate $\tau^{1+\gamma}$ (where τ is inverse-polynomial in n and $\gamma > 0$ is a small constant), borrowing tools from the theory of extractors. The list size and the rate they got are both near-optimal.

Going beyond combinatorial bounds and constructions, there is also research regarding efficient list decoding algorithms. For instance, recently Dinur et al. [41] showed how *double samplers* give rise to a generic way of amplifying distance so as to enable efficient list decoding algorithms. Followup works by Alev et al. [42], Jeronimo et al. [43] and Jeronimo et al. [44] equipped Ta-Shma's codes (and their variants) with efficient list decoding and unique decoding algorithms using connections to high-dimensional expanders and the Sum-of-Square hierarchy.

As we saw, the list size problem is not well understood under the adversarial model. However, it is completely characterized if we are willing to further relax the problem by limiting the adversary to be *oblivious*. Specifically, we call

[10] A binary linear code is said to be δ-*balanced* if the weight of each codeword is between $\frac{1-\delta}{2}n$ and $\frac{1+\delta}{2}n$.

TABLE II: A non-exhaustive list of results on list decodability of random linear codes.

Field size	Noise level	Rate	List size	whp / with constant probability / existential	Reference
$q \geqslant 2$	$p \in (0, 1-1/q)$	$R = 1 - H_q(p) - \delta$	$L = q^{\mathcal{O}(1/\delta)}$	whp	[18]
$q = 2$	$p \in (0, 1/2)$	$R = 1 - H(p) - \delta$	$L \leqslant 1/\delta$	existential	[22]
$q \geqslant 2$	$p \in (0, 1-1/q)$	$R = 1 - H_q(p) - \delta$	$L \leqslant C_{p,q}/\delta$	whp	[21]
$q \geqslant 2$	$p = (1-1/q)(1-\tau)$	$R = \Omega\left(\frac{\tau^2}{\log^3(q/\tau)\log q}\right)$	$L = \mathcal{O}(1/\tau^2)$	with constant probability	[23]
$q \geqslant 2$	$p = (1-1/q)(1-\tau)$	$R = \Omega(\tau^2/\log q)$	$L = \mathcal{O}(1/\tau^2)$	whp	[24]
$q = 2$	$p \in (0, 1/2)$	$R = 1 - H(p) - \delta$	$L \leqslant H(p)/\delta + 2$	whp	[20]
$q = 2$	$p \in (0, 1/2)$	$R = 1 - H(p) - \delta$	$L \leqslant \lfloor H(p)/\delta \rfloor + 1$	whp	[29]

TABLE III: Upper and lower bounds on list sizes of random codes and arbitrary codes.

Channel model	Code	List size	Reference
Error	Random codes	$L \leqslant 1/\delta$ whp	Folklore
	Random binary codes	$L \geqslant \frac{1-2^{-\Omega_p(1/\delta)}}{\delta}$ whp	[20]
	Random binary linear codes	$L \leqslant H(p)/\delta + 2$ whp	[20]
	Random binary linear codes	$L \leqslant \lfloor H(p)/\delta \rfloor + 1$ whp	[29]
	Random linear codes	$L = \mathcal{O}_{p,q}(1/\delta)$ whp	[21]
	Random linear codes	$L \geqslant \lfloor H_q(p)/\delta + 0.99 \rfloor - 1$ whp	[29]
	Arbitrary codes	$L = \Omega_{p,q}(\log(1/\delta))$	[31], [32], [33], [34], [19]
Erasure	Random binary codes	$L \leqslant \frac{1-p+H(p)}{\delta} - 1$ whp	[5], Theorem 10.9
	Random codes	$L \geqslant \frac{1-p}{2\delta}$ whp	[19]
	Arbitrary binary codes	$L \geqslant \log(1 + p/\delta)$	[5], Theorem 10.14
	Random binary linear codes	$L \leqslant 2^{H(p)/\delta} - 1$ whp	[5], Theorem 10.11
	Random linear codes	$L \geqslant \frac{1}{q} 2^{\frac{p(1-p)}{16\delta}}$ whp	[19]
	Arbitrary binary linear codes	$L \geqslant 1 + p/\delta$	[38]

the adversary *omniscient* if the error pattern is a function of the transmitted codeword, i.e., the adversary sees the codeword before he designs the attack vector. Otherwise, an adversary is said to be *oblivious* if the error pattern is independent of the transmitted codewords, i.e., the adversary knows nothing more than the codebook and has to design the attack vector before the codeword is transmitted. The list size-vs.-rate tradeoff is known to a fairly precise extent for *general* discrete memoryless oblivious adversarial channels.

For a general oblivious discrete memoryless Arbitrarily Varying Channel (AVC) $W(y|x,s)$ *without* constraints, Hughes [45] completely characterized its list decoding capacity under *any* L. Specifically, the list-L capacity $C(L)$ equals

$$C(L) = \max_{P_{\mathbf{x}}} \min_{P_{\mathbf{s}}} I(\mathbf{x}; \mathbf{y}) \tag{1}$$

if $L > L*$, where

$$L^* := \max \Bigg\{ \ell \in \mathbb{Z}_{\geqslant 0} : \begin{array}{l} \exists U(s|x_1, \cdots, x_\ell), \forall x_0, x_1, \cdots, x_\ell, \forall \pi \in S_{\ell+1}, \\ \sum_s U(s|x_1, \cdots, x_\ell) W(y|x_0, s) = \\ \quad \sum_s U(s|x_{\pi(1)}, \cdots, x_{\pi(\ell)}) W(y|x_{\pi(0)}, s) \end{array} \Bigg\} \tag{2}$$

and $C(L) = 0$ otherwise. For oblivious AVCs *under* constraints, the critical list size L^* is known [46] though the exact capacity $C(L)$ is open.

In this paper, we will focus on combinatorial/information-theoretic limits of list decoding various ensembles of random codes over $\mathbb{R}$ against omniscient adversaries.

C. Prior work on list decoding over the reals

In this section, we briefly recall what is known about list decoding over the reals. As we will see, it is much less studied and understood, which motivates this work.

1) Shlosman and Tsfasman [47] studied the sphere packing density of a) a random lattice sampled from the Haar distribution; b) a Poisson point process (PPP).
2) Blinvosky [48] later generalized their results on PPP to list-L packing for any constant $L \in \mathbb{Z}_{\geqslant} 1$. However, the proof therein is problematic and is recently corrected in [49] without affecting the result. PPPs are a natural family of ICs. The heuristic results for Haar lattices in Sec. X-A2 of this paper can be cast as (rigorous) list size bounds for PPPs. The bounds in [48], [49] provide finite list size bounds for PPPs which subsume our bounds in Sec. X-A2. However, our proof for $\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)$ list sizes is much simpler.
3) Some bounds [50], [51] on the list-L capacity of arbitrary spherical codes are also known. Note that the proof in [50] is problematic and is recently corrected in [49] without affecting the result.
4) In computer science, there have been results [52], [53] on efficient list decoding algorithms for explicit lattices, e.g., Barnes–Wall lattices, Barnes–Sloane lattices, etc. However, these lattices have rate/NLD way below the capacity.
5) Recently, the authors of the current paper systematically revisited the list decoding problem over $\mathbb{R}$ with *constant* list sizes. Various upper and lower bounds [49] were derived for this problem.

III. Notation and Preliminaries

A. Notation

General notation. We use standard Bachmann-Landau (Big-Oh) notation for asymptotic functions.

For any $q \in \mathbb{R}_{>0}$, we write $\log_q(\cdot)$ for the logarithm to the base q. In particular, let $\log(\cdot)$ denote the logarithm to the base two and let $\ln(\cdot)$ denote the logarithm to the base e.

Sets. For any two sets $\mathcal{A}$ and $\mathcal{B}$ with additive and multiplicative structures, let $\mathcal{A}+\mathcal{B}$ and $\mathcal{A}\cdot\mathcal{B}$ denote the Minkowski sum and Minkowski product of them which are defined as

$$\mathcal{A}+\mathcal{B} \coloneqq \{a+b\colon a\in\mathcal{A}, b\in\mathcal{A}\}, \quad \mathcal{A}\cdot\mathcal{B} \coloneqq \{a\cdot b\colon a\in\mathcal{A}, b\in\mathcal{B}\}.$$

If $\mathcal{A}=\{x\}$ is a singleton set, we write $x+\mathcal{B}$ and $x\mathcal{B}$ for $\{x\}+\mathcal{B}$ and $\{x\}\cdot\mathcal{B}$. For any finite set $\mathcal{X}$ and any integer $0 \leqslant k \leqslant |\mathcal{X}|$, we use $\binom{\mathcal{X}}{k}$ to denote the collection of all subsets of $\mathcal{X}$ of size k, i.e.,

$$\binom{\mathcal{X}}{k} \coloneqq \{\mathcal{Y}\subseteq\mathcal{X}\colon |\mathcal{Y}|=k\}.$$

For $M\in\mathbb{Z}_{>0}$, we let $[M]$ denote the set of the first M positive integers $\{1,2,\cdots,M\}$.

For a subset $\mathcal{T}=\{i_1,\cdots,i_t\}\subseteq[n]$ of t coordinates and a vector $\underline{x}\in\mathcal{X}^n$ over some alphabet $\mathcal{X}$, we use $\underline{x}|_{\mathcal{T}}$ to denote the vector obtained by restricting $\underline{x}$ to the coordinates in $\mathcal{T}$, i.e.,

$$\underline{x}|_{\mathcal{T}} \coloneqq [\underline{x}_{i_1},\cdots,\underline{x}_{i_t}]^\top.$$

Similar notation can be defined for a subset $\mathcal{A}$ of $\mathcal{X}^n$

$$\mathcal{A}|_{\mathcal{T}} \coloneqq \{\underline{x}|_{\mathcal{T}}\colon \underline{x}\in\mathcal{A}\}.$$

For any $\mathcal{A}\subseteq\mathcal{X}$, the indicator function of $\mathcal{A}$ is defined as, for any $x\in\mathcal{X}$,

$$\mathbb{1}_{\mathcal{A}}(x) = \begin{cases} 1, & x\in\mathcal{A} \\ 0, & x\notin\mathcal{A} \end{cases}.$$

At times, we will slightly abuse notation by saying that $\mathbb{1}_{\mathsf{A}}$ is 1 when event A happens and zero otherwise.

Let $\|\cdot\|_2$ denote the Euclidean/L^2-norm. Specifically, for any $\underline{x}\in\mathbb{R}^n$,

$$\|\underline{x}\|_2 \coloneqq \left(\sum_{i=1}^n \underline{x}_i^2\right)^{1/2}.$$

For brevity, we also write $\|\cdot\|$ for the L^2-norm.

Let $\mathrm{Vol}_n(\cdot)$ denote the n-dimensional Lebesgue volume of an Euclidean body. Specifically, for any Euclidean body $\mathcal{A}\subseteq\mathbb{R}^n$,

$$\mathrm{Vol}_n(\mathcal{A}) = \int_{\mathcal{A}} \mathrm{d}\underline{x} = \int_{\mathbb{R}^n} \mathbb{1}_{\mathcal{A}}(\underline{x})\mathrm{d}\underline{x},$$

where $\mathrm{d}\underline{x}$ denotes the differential of $\underline{x}$ with respect to (wrt) the Lebesgue measure on $\mathbb{R}^n$. When the dimension n is obvious from the context, we will also use the shorthand notation $|\cdot|$ for $\mathrm{Vol}_n(\cdot)$. If $\mathcal{A}\subseteq\mathbb{R}^n$ is an n-dimensional body with nonempty interior, we write $\mathrm{Vol}(\mathcal{A})$ for $\mathrm{Vol}_n(\mathcal{A})$; if $\mathcal{A}\subseteq\mathbb{R}^n$ is an $(n-1)$-dimensional hypersurface, we write $\mathrm{Area}(\mathcal{A})$ for $\mathrm{Vol}_{n-1}(\mathcal{A})$.

Sets are denoted by capital letters in calligraphic typeface, e.g., $\mathcal{C},\mathcal{I}$, etc. In particular, let $\mathcal{S}_2^{n-1}$ denote the $(n-1)$-dimensional unit Euclidean sphere wrt L^2-norm, i.e.,

$$\mathcal{S}_2^{n-1} \coloneqq \left\{\underline{y}\in\mathbb{R}^n\colon \|\underline{y}\|_2=1\right\}.$$

The area of this sphere is denoted A_{n-1}.

Let $\mathcal{B}_2^n$ denote the n-dimensional unit Euclidean ball wrt L^2-norm, i.e.,

$$\mathcal{B}_2^n \coloneqq \left\{\underline{y}\in\mathbb{R}^n\colon \|\underline{y}\|_2\leqslant 1\right\}.$$

We denote the volume of this ball, i.e., $\mathrm{Vol}(\mathcal{B}_2^n)$, by V_n.

An $(n-1)$-dimensional Euclidean sphere centered at $\underline{x}$ of radius r is denoted by

$$\mathcal{S}_2^{n-1}(\underline{x},r) = \underline{x}+r\mathcal{S}_2^{n-1} = \{\underline{y}\in\mathbb{R}^n : \|\underline{y}\|_2=r\}.$$

An n-dimensional Euclidean ball centered at $\underline{x}$ of radius r is denoted by

$$\mathcal{B}_2^n(\underline{x},r) = x+r\mathcal{B}_2^n = \{\underline{y}\in\mathbb{R}^n : \|\underline{y}\|_2\leqslant r\}.$$

For any $\underline{x}\in\mathbb{F}_q^n$, let $wt_{\mathrm{H}}(\underline{x})$ denote the Hamming weight of $\underline{x}$, i.e., the number of nonzero entries of $\underline{x}$:

$$wt_{\mathrm{H}}(\underline{x}) \coloneqq \{i\in[n]\colon \underline{x}_i\neq 0\}.$$

For any $\underline{x},\underline{y}\in\mathbb{F}_q^n$, let $d_{\mathrm{H}}(\underline{x},\underline{y})$ denote the Hamming distance between $\underline{x}$ and $\underline{y}$, i.e., the number of locations where they differ:

$$d_{\mathrm{H}}(\underline{x},\underline{y}) \coloneqq wt_{\mathrm{H}}(\underline{x}-\underline{y}) = \left\{i\in[n]\colon \underline{x}_i\neq\underline{y}_i\right\}.$$

We can define balls and spheres in $\mathbb{F}_q^n$ centered around some point of certain radii wrt Hamming metric as well:

$$\mathcal{S}_{\mathrm{H}}^n(\underline{x},r) \coloneqq \left\{\underline{y}\in\mathbb{F}_q^n\colon d_{\mathrm{H}}(\underline{x},\underline{y})=r\right\},$$

$$\mathcal{B}_{\mathrm{H}}^n(\underline{x},r) \coloneqq \left\{\underline{y}\in\mathbb{F}_q^n\colon d_{\mathrm{H}}(\underline{x},\underline{y})\leqslant r\right\}.$$

We will drop the subscript and superscript for the associated metric and dimension when they are clear from the context.

Probability. Random variables are denoted by lower case letters in boldface or capital letters in plain typeface, e.g., $\mathbf{m},\mathbf{x},\mathbf{s},U,W$, etc. Their realizations are denoted by corresponding lower case letters in plain typeface, e.g., m,x,s,u,w, etc. Vectors of length n, where n is the block-length, are denoted by lower case letters with underlines, e.g., $\underline{\mathbf{x}},\underline{\mathbf{s}},\underline{x},\underline{s}$, etc. The i-th entry of a vector is denoted by a subscript i, e.g., $\underline{\mathbf{x}}_i,\underline{\mathbf{s}}_i,\underline{x}_i,\underline{s}_i$, etc. Matrices are denoted by capital letters in boldface, e.g., $\mathbf{I},\mathbf{\Sigma}$, etc. The (i,j)-th entry of a matrix $\mathbf{M}$ is denoted by $\mathbf{M}_{ij}$.

The probability mass function (pmf) of a discrete random variable $\mathbf{x}$ or a random vector $\underline{\mathbf{x}}$ is denoted by $p_{\mathbf{x}}$ or $p_{\underline{\mathbf{x}}}$ respectively. Here with a slight abuse of notation, we use the same to denote the probability density function (pdf) of $\mathbf{x}$ or $\underline{\mathbf{x}}$ if they are continuous. If every entry of $\underline{\mathbf{x}}$ is independent and identically distributed (iid) according to $p_{\mathbf{x}}$, then we write $\underline{\mathbf{x}}\sim p_{\mathbf{x}}^{\otimes n}$. In other words,

$$p_{\underline{\mathbf{x}}}(\underline{x}) = p_{\mathbf{x}}^{\otimes n}(\underline{x}) \coloneqq \prod_{i=1}^n p_{\mathbf{x}}(\underline{x}_i).$$

Let $\mathcal{U}(\Omega)$ denote the uniform distribution over some probability space Ω. Let $\mathcal{N}(\underline{\mu}, \mathbf{\Sigma})$ denote the n-dimensional Gaussian distribution with mean vector $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$.

We use $H(\cdot)$ to denote interchangeably Shannon entropy and differential entropy; the exact meaning will usually be clear from context. In particular, if $p_{\underline{\mathbf{x}}}\colon \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ is a pdf of a random vector $\underline{\mathbf{x}}$ in $\mathbb{R}^n$, $H(\underline{\mathbf{x}})$ denotes the differential entropy of $\underline{\mathbf{x}} \sim p_{\underline{\mathbf{x}}}$,

$$H(\underline{\mathbf{x}}) = -\int_{\mathbb{R}^n} p_{\underline{\mathbf{x}}}(\underline{x}) \log p_{\underline{\mathbf{x}}}(\underline{x}) \mathrm{d}\underline{x}.$$

For any $p \in [0,1]$, $H(p)$ denotes the binary entropy

$$H(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

For any $q \in \mathbb{Z}_{\geqslant 2}$ and any $p \in [0,1]$, $H_q(p)$ denotes the q-ary entropy

$$H_q(p) = p \log_q (q-1) + p \log_q \frac{1}{p} + (1-p) \log_q \frac{1}{1-p}.$$

Algebra. For any field F, we use $\mathrm{SL}(n,F)$ and $\mathrm{GL}(n,F)$ to denote the special linear group and the general linear group over F of degree n, i.e.,

$$\mathrm{SL}(n,F) \coloneqq \left\{\mathbf{M} \in F^{n\times n}\colon \det(\mathbf{M}) = 1\right\},$$
$$\mathrm{GL}(n,F) \coloneqq \left\{\mathbf{M} \in F^{n\times n}\colon \det(\mathbf{M}) \neq 0\right\}.$$

For any vector space V of dimension n and any integer $0 \leqslant k \leqslant n$, the Grassmannian $\mathbf{Gr}(k,V)$ is the collection of all k-dimensional subspaces of V, i.e.,

$$\mathbf{Gr}(k,V) \coloneqq \{U \leqslant V \text{ subspace}\colon \dim_F U = k\}.$$

B. Preliminaries

Probability. We will need the following form of Chernoff bound.

Lemma 6. *Let $X_1, \cdots, X_N$ be independent Bernoulli random variables and let $X \coloneqq \sum_{i=1}^N X_i$. Then for any $0 \leqslant \delta \leqslant 1$, we have*

$$\Pr\left[X \geqslant (1+\delta)\mathbb{E}\left[X\right]\right] \leqslant \exp\left(-\frac{\delta^2}{3}\mathbb{E}\left[X\right]\right),$$
$$\Pr\left[X \leqslant (1-\delta)\mathbb{E}\left[X\right]\right] \leqslant \exp\left(-\frac{\delta^2}{2}\mathbb{E}\left[X\right]\right).$$

The following lemma is an easy corollary of Chebyshev inequality.

Lemma 7. *For any nonnegative random variable X, $\Pr\left[X=0\right] \leqslant \mathrm{Var}\left[X\right]/\mathbb{E}\left[X\right]^2$.*

Recall two facts about the moments of Gaussian and Poisson random variables.

Fact 8. *Let $\mathbf{g} \sim \mathcal{N}(0,1)$, then*

$$\mathbb{E}\left[\mathbf{g}^k\right] = \begin{cases} 0, & k \text{ odd} \\ (k-1)!!, & k \text{ even} \end{cases},$$

where $\ell!! \coloneqq \ell(\ell-2)(\ell-4)\cdots 3\cdot 1$ denotes the double factorial of ℓ odd.

Fact 9. *Let $\mathbf{p} \sim \mathrm{Pois}(\lambda)$, then*

$$\mathbb{E}\left[\mathbf{p}^k\right] = e^{-\lambda} \sum_{i=0}^{\infty} \frac{i^k}{i!} \lambda^i.$$

Poisson random variables are additive.

Fact 10. *If $\mathbf{p}_1 \sim \mathrm{Pois}(\lambda_1)$ and $\mathbf{p}_2 \sim \mathrm{Pois}(\lambda_2)$ are independent, then $\mathbf{p}_1 + \mathbf{p}_2 \sim \mathrm{Pois}(\lambda_1+\lambda_2)$.*

We know the following tail bounds for Poisson random variables.

Lemma 11. *Let $\mathbf{p} \sim \mathrm{Pois}(\lambda)$ and $\ell > \lambda, m < \lambda$, then*

$$\Pr\left[\mathbf{p} > \ell\right] < \frac{e^{-\lambda}(e\lambda)^\ell}{\ell^\ell}, \quad \Pr\left[\mathbf{p} < m\right] < \frac{e^{-\lambda}(e\lambda)^m}{m^m}.$$

Lemma 12 ([54])**.** *Let $\mathbf{p} \sim \mathrm{Pois}(\lambda)$ and $\Delta > 0$, then*

$$\Pr\left[\mathbf{p} - \lambda \geqslant \Delta\right] \leqslant e^{\frac{\Delta^2}{2(\lambda+\Delta)}},$$
$$\Pr\left[\mathbf{p} - \lambda \leqslant -\Delta\right] \leqslant e^{\frac{\Delta^2}{2(\lambda+\Delta)}},$$
$$\Pr\left[|\mathbf{p} - \lambda| \geqslant \Delta\right] \leqslant 2e^{\frac{\Delta^2}{2(\lambda+\Delta)}}.$$

Geometry. It is well-known that Stirling's approximation gives an asymptotic expression for factorials.

Lemma 13. *For any $n \in \mathbb{Z}_{>0}$, $n! = \sqrt{2\pi n}(n/e)^n(1+o(1))$.*

We can use the above lemma to obtain the asymptotic behaviour of binomial coefficients. At times, we also resort to the following cheap yet convenient bounds.

Lemma 14. *For any $n \in \mathbb{Z}_{>0}$ and $0 \leqslant k \leqslant n$, $(n/k)^k \leqslant \binom{n}{k} \leqslant (en/k)^k$.*

Recall the formulas and asymptotics of the volume of a unit Euclidean ball and the area of a unit Euclidean sphere.

Fact 15. $V_n \coloneqq \mathrm{Vol}(\mathcal{B}_2^n) = \frac{\pi^{\frac{n}{2}}}{\Gamma(n/2+1)} = \frac{1}{\sqrt{\pi n}}\left(\frac{2\pi e}{n}\right)^{n/2}(1+o(1))$.

Fact 16. $A_{n-1} \coloneqq \mathrm{Area}(\mathcal{S}_2^{n-1}) = \frac{n\pi^{\frac{n}{2}}}{\Gamma(n/2+1)} = \sqrt{\frac{n}{\pi}}\left(\frac{2\pi e}{n}\right)^{n/2}(1+o(1))$.

IV. List Decodability of Spherical Codes

We now investigate lower bounds on the list size L for codes that operate at rate $R = C(P,N) - \delta$.

A. A reduction from an arbitrary code to a spherical code

We first show that it suffices to prove a lower bound on list size for spherical codes.

Lemma 17. *Suppose there exists a (P,N,L)-list decodable code $\mathcal{C} \subset \mathcal{B}_2^n(0,\sqrt{nP})$ of rate R. Then, there exists a $\left(P, N, \frac{P}{4N}L\right)$-list decodable code $\mathcal{C}' \subset \mathcal{S}_2^{n-1}(0,\sqrt{nP})$ of asymptotically the same rate.*

Proof. Given any (P,N,L)-list decodable (ball) code $\mathcal{C}$ in $\mathcal{B}_2^n(0,\sqrt{nP})$, we can construct a $\left(P, N, \frac{P}{4N}L\right)$-list decodable spherical code $\mathcal{C}'$ on $\mathcal{S}_2^{n-1}(0,\sqrt{nP})$. Indeed, we just project

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

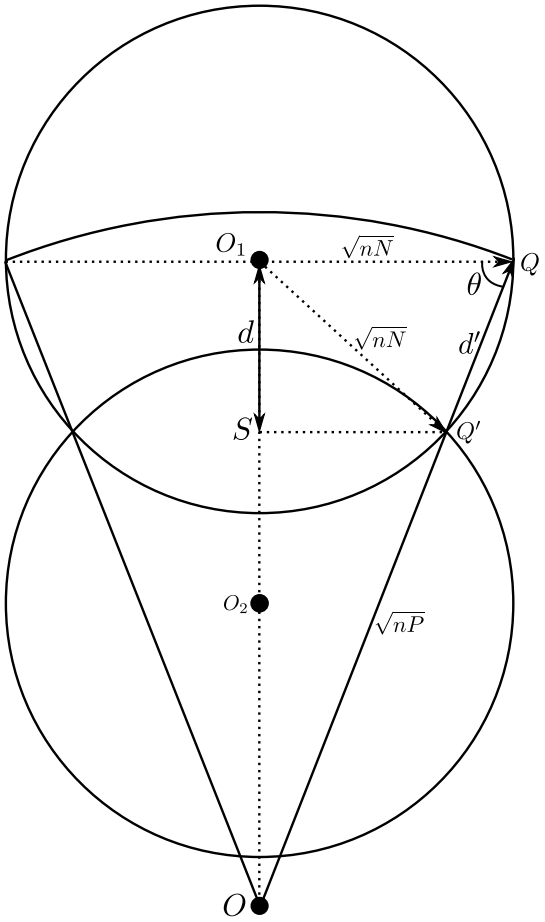


Fig. 1: A covering of the cone using balls. This figure is useful for understanding the reduction (Lemma 17) from the list-decodability of ball codes to that of spherical codes. The ball centered at O has radius $\sqrt{nP}$. We project radially all codewords inside the ball onto the sphere. The number of codewords (after reduction) on the spherical cap centered at O_1 is the same as that (before reduction) in the cone with angular radius $\angle O_1OQ$. By list-decodability of the ball code, there are at most L codewords in any ball of radius $\sqrt{nN}$. Therefore, to bound the number of codewords in the cone, it suffices to cover the cone using balls of radius $\sqrt{nN}$. By elementary geometry, we need at most $|OO_1|/(2d)$ where $|OO_1| = \sqrt{n(P-N)}$ and $2d$ is the distance between the centers O_1 and O_2 of two consecutive balls in the covering.

all codewords radially onto $\mathcal{S}^{n-1}\left(0, \sqrt{nP}\right)$. Then we know that for any direction $\underline{\theta} \in \mathcal{S}^{n-1}$,

$$\left|\mathcal{C}' \cap \mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN})\right| \leqslant |\mathcal{N}|\, L, \tag{3}$$

where $\mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN})$ is a cap of radius $\sqrt{nN}$ on the sphere $\mathcal{S}^{n-1}(0, \sqrt{nP})$ along direction $\underline{\theta}$,

$$\mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN}) := \left\{\underline{x} \in \mathcal{S}^{n-1}(0, \sqrt{nP})\colon \langle \underline{x}, \underline{\theta}\rangle \geqslant \sqrt{n(P-N)}\right\}; \tag{4}$$

and $\mathcal{N}$ is a $\sqrt{nN}$-covering[11] of the cone

$$\mathcal{K}(\underline{\theta}) := \left\{\lambda \underline{x}\colon \underline{x} \in \mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN}), \lambda \in [0,1]\right\}$$

induced by the cap. Note that according to the definition (Eqn. (4)), the center of the base of $\mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN})$ is O_1 and the radius of the base is $\sqrt{nN}$. Or equivalently, the angular radius of $\mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN})$ is $\angle O_1OQ$.

To see why Eqn. (3) implies $(P, N, |\mathcal{N}|L)$-list decodability of $\mathcal{C}'$, i.e., $|\mathcal{C}' \cap \mathcal{B}(\underline{y}, \sqrt{nN})| \leqslant |\mathcal{N}|L$ *for any* $\underline{y} \in \mathbb{R}^n$, one should note that (i) $\mathcal{C}ap^{n-1}(\underline{\theta}, \sqrt{nN}) = \mathcal{S}_2^{n-1}(0, \sqrt{nP}) \cap \mathcal{B}_2^n(\underline{y}, \sqrt{nN})$ for some $\underline{y}$ such that $\|\underline{y}\| = \sqrt{n(P-N)}$; (ii) $\underline{y} \in \mathcal{S}_2^{n-1}(0, \sqrt{n(P-N)})$ maximizes the surface area of the intersection between $\mathcal{B}_2^n(\underline{y}, \sqrt{nN})$ and $\mathcal{S}_2^{n-1}(0, \sqrt{nP})$. Therefore, if Eqn. (3) holds for any $\underline{y} = \underline{\theta}\sqrt{n(P-N)}$ (for some $\underline{\theta} \in \mathcal{S}^{n-1}$), then the same must hold for any $\underline{y} \in \mathbb{R}^n$.

We can upper bound $|\mathcal{N}|$ by

$$|\mathcal{N}| \leqslant \frac{\sqrt{P-N}}{2d}, \tag{5}$$

where d is shown in Fig. 1 and will be computed momentarily. This can be seen by staring at the geometry of a covering as shown in Figure 1. One way to cover the cone $\mathcal{K}(\underline{\theta})$ is to align the centers of the balls $\mathcal{B}(\cdot, \sqrt{nN})$ on the ray rooted at the center O of $\mathcal{S}^{n-1}(0, \sqrt{nP})$ in the direction $\underline{\theta}$. We enumerate such ball in ascending order from the surface to the center of the sphere $\mathcal{S}^{n-1}(0, \sqrt{nP})$. That is, the ball whose center is closest to the surface of $\mathcal{S}^{n-1}(0, \sqrt{nP})$ is the 1-st one and the ball whose center is closest to the center of $\mathcal{S}^{n-1}(0, \sqrt{nP})$ is the $|\mathcal{N}|$-th one. Let $2d_i$ denote the distance between the centers of the i-th and the $(i+1)$-st balls. Since all centers are on the segment OO_1 of length $\sqrt{n(P-N)}$ and we apparently have $d_1 < d_2 < \cdots < d_{|\mathcal{N}|-1}$, we can upper bound $|\mathcal{N}|$ by $\frac{\sqrt{P-N}}{2d}$ where $d := d_1$.

We now compute d. By symmetry, the distance between the centers O_1 and O_2 of the first two balls is equal to $2d$ where $d := SO_1/\sqrt{n} = SO_2/\sqrt{n}$. Since the triangles $\Delta OSQ'$ and ΔOO_1Q are similar, d is given by the following equation

$$\frac{|OS|}{|OO_1|} = \frac{|OQ'|}{|OQ|} \iff \frac{\sqrt{P-N}-d}{\sqrt{P-N}} = \frac{\sqrt{P}-d'}{\sqrt{P}}, \tag{6}$$

where $d' := Q'Q/\sqrt{n}$. On the other hand, the triangle $\Delta O_1Q'Q$ is isosceles with side length $O_1Q' = O_1Q = \sqrt{nN}$. Let $\theta := \angle O_1QQ'$. It is immediate that $d' = 2\sqrt{N}\cos\theta = 2N/\sqrt{P}$ since $\cos\theta = QO_1/QO = \sqrt{N/P}$ in ΔQO_1O. Plugging it into Eqn. (6) and solving d, we have $d = 2\frac{N}{P}\sqrt{P-N}$. Hence by Eqn. (5), $|\mathcal{N}| \leqslant \frac{P}{4N}$. Substituting it to Eqn. (3) finishes the proof.

□

Remark 5. In fact, since we are covering a cone rather than a cylinder, the most economical way of covering is not to align the balls with consecutive distance $2d$. Indeed, the optimal covering $\mathcal{N}^*$ has strictly increasing distances $2d = d_1 < d_2 < \cdots < d_{|\mathcal{N}^*|-1}$, where d_i is the half distance between the

[11] A Δ-*covering* (a.k.a. a Δ-net) $\mathcal{N}$ of a metric space $(\mathcal{X}, d)$ is a subset $\mathcal{N} \subset \mathcal{X}$ satisfying that for any $x \in \mathcal{X}$, there exists an $x' \in \mathcal{N}$ such that $d(x, x') \leqslant \Delta$.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

centers O_i and O_{i+1} of the i-th and the $(i+1)$-st balls. One can compute each d_i explicitly. Although our bound is crude, it is still a valid and simple upper bound and is tight for covering a cylinder.

B. List size lower bound for uniformly random spherical codes

Although we are not able to obtain a lower bound for arbitrary spherical codes as in [50], [51], we can obtain a lower bound for uniformly random spherical codes.

Proposition 18. *Fix $P > N > 0$, and let $C = \frac{1}{2}\log\frac{P}{N}$. For every $\delta > 0$, if $\mathcal{C}$ is a random spherical code on $\mathcal{S}^{n-1}(0,\sqrt{nP})$ of rate $C-\delta$, then*

$$\Pr\left[\mathcal{C} \text{ is } \left(P, N, \frac{c'}{\delta}-1\right)\text{-list decodable}\right] \leqslant 2^{-\Theta(n)},$$

for every $c'<C$.

Proof. The proof follows a second-moment method as in Guruswami and Narayanan [19] for binary codes.

Choose a $\sqrt{n\varepsilon}$-net $\mathcal{Y}$ for $\mathcal{S}^{n-1}\left(0,\sqrt{n\left(P-N\right)}\right)$ for some constant $\varepsilon>0$. In other words, $\mathcal{Y}\subset\mathcal{S}^{n-1}\left(0,\sqrt{n\left(P-N\right)}\right)$ and for all $\underline{y}\in\mathcal{S}^{n-1}\left(0,\sqrt{n\left(P-N\right)}\right)$, we have $\min_{\underline{u}\in\mathcal{Y}}\|\underline{y}-\underline{u}\|\leqslant\sqrt{n\epsilon}$.

For any spherical code $\mathcal{C}$, define

$$W := \sum_{\underline{y}\in\mathcal{Y}}\sum_{\{m_1,\cdots,m_L\}\in\binom{\mathcal{M}}{L}} \mathbb{1}\left\{\psi(m_1),\cdots,\psi(m_L)\in\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right\} \tag{7,8}$$

where $\mathcal{M} := \{0,1,\cdots,2^{nR}-1\}$ is the set of messages and $\psi(m)$ denotes the codeword corresponding to m. Let $M := |\mathcal{M}| = 2^{nR}$. Clearly, $W = 0$ if and only if (iff) $\mathcal{C}$ is $(P,N,L-1)$-list decodable. Letting $\mathcal{A} := \mathcal{B}^n\left(0,\sqrt{nP}+\sqrt{nN}\right)\setminus\mathcal{B}^n\left(0,\sqrt{nP}-\sqrt{nN}\right)$,

$$\begin{aligned}
&\Pr\left[\mathcal{C} \text{ is } (P,N,L-1)\text{-list decodable}\right]\\
&= \Pr\left[\bigcap_{\underline{y}\in\mathcal{A}}\left\{\left|\mathcal{C}\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right|<L\right\}\right]\\
&\leqslant \Pr\left[\bigcap_{\underline{y}\in\mathcal{Y}}\left\{\left|\mathcal{C}\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right|<L\right\}\right]\\
&= \Pr\left[W=0\right]\\
&\leqslant \mathrm{Var}\left[W\right]/\mathbb{E}\left[W\right]^2, \qquad (9)
\end{aligned}$$

where the last inequality (9) follows from Lemma 7. Let

$$\mu := \frac{\mathrm{Area}\left(\mathcal{C}ap^{n-1}\left(\sqrt{nN}\right)\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}\left(\sqrt{nP}\right)\right)},$$

$$\nu := \frac{\left|\mathcal{Y}\cap\mathcal{C}ap^{n-1}(\sqrt{nN(P-N)/P})\right|}{|\mathcal{Y}|}.$$

Then, we show that

$$\mathbb{E}[W] \geqslant (M/L)^L\,|\mathcal{Y}|\mu^L, \tag{10}$$

and

$$\mathrm{Var}\left[W\right] \leqslant |\mathcal{Y}|^2 L\nu^2 M^L\mu^{L-1}. \tag{11}$$

See Appendix B-B and B-C for the details. Plugging these in Eqn. (9), we get

$$\Pr\left[\mathcal{C} \text{ is } (P,N,L-1)\text{-list decodable}\right] \leqslant L^{2L+1}\nu^2\mu^{-L-1}M^{-L}. \tag{12}$$

We need an upper bound on ν which is given by Eqn. (73). Let $c_2 := \frac{3\sqrt{P}}{2(\sqrt{N(P-N)/P}+3\sqrt{\varepsilon}/2)}$. Eqn. (73) implies

$$\begin{aligned}
\nu &\leqslant c_2\left(\frac{\sqrt{N(P-N)/P}+3\sqrt{\varepsilon}/2}{\sqrt{P-N}}\right)^n\\
&= c_2 2^{n\log\left(\sqrt{N/P}+\frac{3\sqrt{\varepsilon}}{2\sqrt{P-N}}\right)}\\
&= c_2 2^{-\frac{n}{2}\log\frac{P}{N}+n\log\left(1+\frac{3\sqrt{\varepsilon}\sqrt{P/N}}{2\sqrt{P-N}}\right)}\\
&\leqslant c_2 2^{-n\left(\frac{1}{2}\log\frac{P}{N}-\varepsilon'\right)},
\end{aligned}$$

where the last inequality follows since $\log(1+x)\leqslant 2x$ for $x\geqslant 0$ and $\varepsilon' := 3\sqrt{\frac{P\varepsilon}{N(P-N)}}$.

We also need a lower bound on μ:

$$\begin{aligned}
\mu &\geqslant \frac{\mathrm{Vol}\left(\mathcal{B}^{n-1}\left(0,\sqrt{nN}\right)\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}\left(0,\sqrt{nP}\right)\right)}\\
&= c_3 2^{-n\left(\frac{1}{2}\log\frac{P}{N}+o(1)\right)}, \qquad (13)
\end{aligned}$$

for some constant $c_3>0$. The probability (12) we want to upper bound is at most

$$\begin{aligned}
&\Pr\left[\mathcal{C} \text{ is } (P,N,L-1)\text{-list decodable}\right]\\
&\leqslant L^{2L+1}c_2^2c_3^{-L-1}2^{-2n\left(\frac{1}{2}\log\frac{P}{N}-\varepsilon'\right)}2^{n(L+1)\left(\frac{1}{2}\log\frac{P}{N}+o(1)\right)-nRL}\\
&= L^{2L+1}c_2^2c_3^{-L-1}2^{n\left(\delta L-\frac{1}{2}\log\frac{P}{N}+2\varepsilon'+o(1)\right)}\\
&= L^{2L+1}c_2^2c_3^{-L-1}2^{n\left(\delta L-\frac{1}{2}\log\frac{P}{N}+\delta+o(1)\right)}.
\end{aligned}$$

In the last equation, we set $\varepsilon'=\delta/2$, i.e., $\varepsilon=\frac{N(P-N)}{P}\left(\frac{\delta}{6}\right)^2$. The constant-in-$n$ terms downstairs and the $o(n)$ term in the exponent are not important. The probability that $\mathcal{C}$ is list decodable vanishes in n when $L<\frac{\frac{1}{2}\log\frac{P}{N}}{\delta}-1$. That is to say, for a uniformly random spherical code to be $(P,N,L-1)$-list decodable with high probability, L has to be at least $C/\delta-1$, where $C=\frac{1}{2}\log\frac{P}{N}$. □

We would like to emphasize that the above result only implies that a typical random code is not $(P,N,c'/\delta-1)$-list decodable *with high probability*. This does not claim the *non-existence* of $(P,N,c'/\delta-1)$-list decodable codes of rate $C-\delta$.

V. List Decodability of Nested Construction-A Lattice Codes

A. Nested lattice codes

Recall that a lattice Λ is a discrete subgroup of $\mathbb{R}^n$, and can be written as $\mathbf{G}\mathbb{Z}^n$ where $\mathbf{G}$ is called a generator matrix of Λ. The following subsection summarizes some of the important concepts on lattices and lattice codes. The concepts we use

are quite standard in the literature on lattices [11], [55]. The reader who is familiar with this line of work may skip the following subsection.

B. A primer on lattices and nested lattice codes

For a tutorial introduction to lattices and their applications, see the books by Zamir [55], Conway and Sloane [56], or the notes by Barvinok [57]. Here, we summarize the essential concepts required to develop our results.

If $\underline{v}_1, \ldots, \underline{v}_\kappa$ are linearly independent vectors in $\mathbb{R}^n$, then the set of all integer linear combinations of $\underline{v}_1, \ldots, \underline{v}_\kappa$ is called the lattice generated by the vectors $\underline{v}_1, \ldots, \underline{v}_\kappa$, i.e.,

$$\Lambda := \left\{ \sum_{i=1}^{\kappa} a_i \underline{v}_i : a_i \in \mathbb{Z} \right\}.$$

If $\mathbf{G} = [\underline{v}_1 \cdots \underline{v}_\kappa]$, then we can write $\Lambda = \mathbf{G}\mathbb{Z}^\kappa$. The matrix $\mathbf{G}$ is called a generator matrix for Λ. The generator matrix of a lattice is not unique. The integer κ is invariant for a lattice and is called the rank of Λ. In this paper, we only consider lattices in $\mathbb{R}^n$ having rank n. It is obvious that Λ is a discrete subgroup of $\mathbb{R}^n$ under vector addition. It is also a fact that every discrete subgroup of $\mathbb{R}^n$ is a lattice [57].

For any lattice Λ, it is natural to define the quantizer Q_Λ which maps every point in $\mathbb{R}^n$ to the closest lattice point, i.e., for every $\underline{x} \in \mathbb{R}^n$,

$$Q_\Lambda(\underline{x}) := \operatorname*{argmin}_{\underline{y} \in \Lambda} \|\underline{y} - \underline{x}\|, \tag{14}$$

where we assume that ties (in computing the closest lattice point) are resolved according to some arbitrary but fixed rule. Associated with the quantizer is the quantization error

$$[\underline{x}] \bmod \Lambda := \underline{x} - Q_\Lambda(\underline{x}).$$

For every lattice Λ, we define the following parameters:

- The set
$$\mathcal{P}(\Lambda) := \{\mathbf{G}\underline{x} : \underline{x} \in [0,1)^n\},$$
where $\mathbf{G}$ is a generator matrix of Λ, is called the fundamental parallelepiped of Λ.
- The fundamental Voronoi region $\mathcal{V}(\Lambda)$ is the set of all points in $\mathbb{R}^n$ which are closest to the zero lattice point. In other words,
$$\mathcal{V}(\Lambda) := \{\underline{x} \in \mathbb{R}^n : Q_\Lambda(\underline{x}) = 0\}.$$
Any set $\mathcal{S} \subset \mathbb{R}^n$ such that the set of translates of $\mathcal{S}$ by lattice points, i.e., $\{\mathcal{S} + \underline{x} : \underline{x} \in \Lambda\}$ form a partition of $\mathbb{R}^n$, is called a fundamental region of Λ. It is a fact that every fundamental region of Λ has the same volume equal to $\det \Lambda := |\det(\mathbf{G})|$, where $\mathbf{G}$ is any generator matrix of Λ. The quantity $\det \Lambda$ is called the determinant or covolume of Λ.
- The covering radius $r_{\mathrm{cov}}(\Lambda)$ is the radius of the smallest closed ball in $\mathbb{R}^n$ which contains $\mathcal{V}(\Lambda)$. It is also equal to the length of the largest vector within $\mathcal{V}(\Lambda)$.
- The packing radius $r_{\mathrm{pack}}(\Lambda)$ is the radius of the largest open ball which is contained within $\mathcal{V}(\Lambda)$. Equivalently, it is half the minimum distance between two lattice points.
- The effective radius $r_{\mathrm{eff}}(\Lambda)$ is equal to the radius of a ball having volume equal to $\mathrm{Vol}(\mathcal{V}(\Lambda))$.

Clearly, we have $r_{\mathrm{pack}}(\Lambda) \leqslant r_{\mathrm{eff}}(\Lambda) \leqslant r_{\mathrm{cov}}(\Lambda)$.

In the context of power-constrained communication over Gaussian channels, a nested lattice code is typically the set of all lattice points within a convex compact subset of $\mathbb{R}^n$, i.e., $\mathcal{C} = \Lambda \cap \mathcal{B}$ for some set $\mathcal{B} \subset \mathbb{R}^n$. Usually $\mathcal{B}$ is taken to be $\mathcal{B}(0, \sqrt{nP})$ or $\mathcal{V}(\Lambda_0)$ for some lattice Λ_0 constructed so as to satisfy the power constraint.

If Λ_0, Λ are two lattices in $\mathbb{R}^n$ with the property that $\Lambda_0 \subsetneq \Lambda$, then Λ_0 is said to be nested within (or, a sublattice of) Λ. We call Λ the fine lattice, and Λ_0 the coarse lattice. A nested lattice code with a fine lattice Λ and coarse lattice $\Lambda_0 \subsetneq \Lambda$ is the finite set $\Lambda \cap \mathcal{V}(\Lambda_0)$.

Lattices have been extensively used for problems of packing, covering and communication over Gaussian channels. For many problems of interest, we want to construct high-dimensional lattices Λ such that $r_{\mathrm{pack}}(\Lambda)/r_{\mathrm{eff}}(\Lambda)$ is as large as possible, and $r_{\mathrm{cov}}(\Lambda)/r_{\mathrm{eff}}(\Lambda)$ is as small as possible. A class of lattices that has these properties is the class of Construction-A lattices, which we describe next.

Let q be a prime number, and $\mathcal{C}_{\mathrm{lin}}$ be an (n, κ) linear code over $\mathbb{F}_q$. The Construction-A lattice obtained from $\mathcal{C}_{\mathrm{lin}}$ is defined to be

$$\Lambda(\mathcal{C}_{\mathrm{lin}}) := \{\underline{v} \in \mathbb{Z}^n : [\underline{v}] \bmod (q\mathbb{Z}^n) \in \Phi(\mathcal{C})\},$$

where Φ denotes the natural embedding of $\mathbb{F}_q^n$ in $\mathbb{R}^n$.[12] An equivalent definition is that $\Lambda(\mathcal{C}_{\mathrm{lin}}) = \Phi(\mathcal{C}_{\mathrm{lin}}) + q\mathbb{Z}^n$. We make use of the following result to choose our coarse lattices:

Theorem 19 ([11]). *For every $\delta > 0$, there exist sequences of prime numbers q_n and positive integers κ_n such that if $\mathcal{C}_{\mathrm{lin}}$ is a randomly chosen linear code[13] over $\mathbb{F}_{q_n}$, then*

$$\Pr\left[\frac{r_{\mathrm{pack}}(\Lambda(\mathcal{C}_{\mathrm{lin}}))}{r_{\mathrm{eff}}(\Lambda(\mathcal{C}_{\mathrm{lin}}))} < \frac{1}{2} - \delta \text{ or } \frac{r_{\mathrm{cov}}(\Lambda(\mathcal{C}_{\mathrm{lin}}))}{r_{\mathrm{eff}}(\Lambda(\mathcal{C}_{\mathrm{lin}}))} > 1 + \delta \right] = o(1).$$

C. List decodability of nested lattice code

Our goal is to construct good nested lattice pairs (Λ, Λ_0) with $\Lambda_0 \subset \Lambda$, and our nested lattice code will be defined as $\mathcal{C} := \Lambda \cap \mathcal{V}(\Lambda_0)$. The nested lattice code satisfies the power constraint if $r_{\mathrm{cov}}(\Lambda_0) \leqslant \sqrt{nP}$.

We now prove an upper bound on the list size for nested lattice codes. Our goal is to show the following:

Theorem 20. *Let $0 < \delta < 0.9$ and $P > N$. There exist nested lattice codebooks of rate $\frac{1}{2}\log_2 \frac{P}{N} - \delta$ that are $(P, N, 2^{\mathcal{O}(\frac{1}{\delta}\log_2^2 \frac{1}{\delta})})$-list decodable.*

D. List size upper bound for nested Construction-A lattice codes

We start with a (full rank) coarse lattice Λ_0 that satisfies

$$\frac{r_{\mathrm{cov}}(\Lambda_0)}{r_{\mathrm{eff}}(\Lambda_0)} \leqslant 2^{\delta/8} \tag{15}$$

[12] Since q is a prime, the natural embedding simply maps $a \in \mathbb{F}_q$ to $a \in \mathbb{R}$.

[13] The (n, κ_n) random code is obtained by choosing an $n \times \kappa_n$ generator matrix uniformly at random over $\mathbb{F}_q$.

and

$$\frac{r_{\text{pack}}(\Lambda_0)}{r_{\text{eff}}(\Lambda_0)} > \frac{1}{4}. \tag{16}$$

Such lattices are guaranteed to exist (for sufficiently large n) by [11] (See Section V-B). The lattice is suitably scaled so that $r_{\text{cov}}(\Lambda_0) = \sqrt{nP}$ and this will ensure that the codebook satisfies the power constraint. Note that scaling the lattice by a constant factor scales $r_{\text{pack}}, r_{\text{eff}}$ and r_{cov} by the same amount, and the ratios in Eqn. (15) and (16) remain unchanged. Let $\mathbf{G}_{\Lambda_0}$ be a generator matrix for Λ_0, and q be the smallest prime number that satisfies

$$1 + \frac{\sqrt{P}}{q\sqrt{N}} \leqslant 2^{\delta/8}. \tag{17}$$

Note that q is independent of n and is of order $q = \Omega(1/\delta)$. Bertrand's postulate guarantees that for every positive integer m, there exists a prime number between m and $2m$. Therefore,

$$\frac{\sqrt{P/N}}{2^{\delta/8}-1} \leqslant q \leqslant 2\frac{\sqrt{P/N}}{2^{\delta/8}-1} + 2. \tag{18}$$

Let $R = \frac{1}{2}\log_2 \frac{P}{N} - \delta$, and κ be an integer such that[14]

$$\frac{\kappa}{n}\log_2 q = R. \tag{19}$$

We define an ensemble of fine lattices as follows: Choose an $n \times \kappa$ generator matrix $\mathbf{G}_{\text{lin}}$ uniformly over $\mathbb{F}_q^{n\times\kappa}$. This defines a linear code $\mathcal{C}(\mathbf{G}_{\text{lin}}) = \mathbf{G}_{\text{lin}}\mathbb{F}_q^{\kappa}$ where the arithmetics are over $\mathbb{F}_q$. Let $\Lambda' := \frac{1}{q}\Phi(\mathcal{C}(\mathbf{G}_{\text{lin}})) + \mathbb{Z}^n$, where Φ is the natural embedding of $\mathbb{F}_q^n$ into $\mathbb{R}^n$ and the arithmetics are over $\mathbb{R}$. In other words, Φ operates componentwise on vectors, and maps $0, 1, \ldots, q-1 \in \mathbb{F}_q$ to $0, 1, \ldots, q-1 \in \mathbb{R}$. Note that $\mathbb{Z}^n \subset \Lambda' \subset q^{-1}\mathbb{Z}^n$. Our fine lattice is $\Lambda := \mathbf{G}_{\Lambda_0}\Lambda'$. It is easy to verify that Λ_0 is always a sublattice of Λ. In fact, $\Lambda_0 \subset \Lambda \subset q^{-1}\Lambda_0$ forms a chain of nested lattices. Our nested lattice codebook is then $\mathcal{C} := \Lambda \cap \mathcal{V}(\Lambda_0)$.

We will show the following result, which implies Theorem 20.

Theorem 21. *If $P > N$, then*

$$\Pr[\Lambda \cap \mathcal{V}(\Lambda_0) \textit{ is not } (P, N, 2^{\mathcal{O}(\frac{1}{\delta}\log_2^2\frac{1}{\delta})})\textit{-list decodable}] \\ 2^{-\Omega(n)}.$$

Note that the only randomness involved is in the choice of the generator matrix $\mathbf{G}_{\text{lin}}$ that is used to construct the fine lattice Λ.

We now discuss some intermediate lemmas which will be used to prove Theorem 20. The formal proofs will be given in Sec. V-E.

Fix any $\underline{y} \in \mathbb{R}^n$. Fundamental to the proof is counting the number of lattice points within a ball of radius r around $\underline{y}$. We will need bounds on $\left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}(\underline{y}, r)\right|$. We can write it as $\left|\left\{\underline{x} \in \mathbb{Z}^n : \left\|\underline{y} - \frac{1}{q}\mathbf{G}_{\Lambda_0}\underline{x}\right\|_2 \leqslant r\right\}\right|$. A simple argument generalizing [58, Lemma 1] can be used to show that this is upper (resp. lower) bounded by the volume ratio between the ball (whose radius is lengthened (resp. shortened) by the covering radius of $q^{-1}\Lambda_0$) and the fundamental Voronoi region of $q^{-1}\Lambda_0$. This can be formally stated as follows:

Lemma 22 (Generalization of [58, Lemma 1])**.** *Let V_n denote the volume of the unit ball in $\mathbb{R}^n$, and Λ_0 be a full-rank lattice in $\mathbb{R}^n$. Then, for any $r > r_{\text{cov}}(\Lambda_0)/q = r_{\text{cov}}(q^{-1}\Lambda_0)$ and $\underline{y} \in \mathbb{R}^n$, we have*

$$\frac{q^n V_n}{\text{Vol}(\mathcal{V}(\Lambda_0))}\left(r - \frac{r_{\text{cov}}(\Lambda_0)}{q}\right)^n \leqslant \left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}(\underline{y}, r)\right| \\ \leqslant \frac{q^n V_n}{\text{Vol}(\mathcal{V}(\Lambda_0))}\left(r + \frac{r_{\text{cov}}(\Lambda_0)}{q}\right)^n. \tag{20}$$

Observe that there is a bijection between $\mathbb{F}_q^{\kappa}$ and $\Lambda \cap \mathcal{V}(\Lambda_0)$. The encoder maps $m \in \mathbb{F}_q^{\kappa}$ to a nested lattice codeword (with slight abuse of notation[15])

$$\psi(m) := \left[\frac{1}{q}\mathbf{G}_{\Lambda_0}\left([\mathbf{G}_{\text{lin}}m] \bmod (q\mathbb{Z}^n)\right)\right] \bmod \Lambda_0,$$

where all arithmetics are over $\mathbb{R}$.

Lemma 23. *Fix $m \in \mathbb{F}_q^{\kappa}\backslash\{0\}$ and $\underline{y} \in \mathbb{R}^n$. We have*

$$\Pr[\psi(m) \in \mathcal{B}(\underline{y}, r)] \leqslant \left(\frac{r}{\sqrt{nP}}2^{\delta/8}\left(1 + \frac{\sqrt{nP}}{qr}\right)\right)^n. \tag{21}$$

Proof. Since $\mathbf{G}_{\text{lin}} \in \{0, 1, \cdots, q-1\}^{n\times\kappa}$ is a uniformly random matrix, $[\mathbf{G}_{\text{lin}}m] \bmod (q\mathbb{Z}^n)$ is uniformly distributed in $\{0, 1, \cdots, q-1\}^n$. Consequently $\psi(m)$ is uniformly distributed in $q^{-1}\Lambda_0 \cap \mathcal{V}(\Lambda_0)$. Since the codeword $\psi(m)$ is guaranteed to be in $\mathcal{V}(\Lambda_0)$, we have

$$\begin{aligned}\Pr[\psi(m) \in \mathcal{B}(\underline{y}, r)] &\leqslant \Pr\left[\psi(m) \in [\mathcal{B}(\underline{y}, r)] \bmod \Lambda_0\right] \\ &= \frac{1}{q^n}\left|\frac{1}{q}\Lambda_0 \cap [\mathcal{B}(\underline{y}, r)] \bmod \Lambda_0\right| \\ &\leqslant \frac{1}{q^n}\left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}(\underline{y}, r)\right| \\ &\leqslant \frac{V_n}{\text{Vol}(\mathcal{V}(\Lambda_0))}\left(r + \frac{r_{\text{cov}}(\Lambda_0)}{q}\right)^n\end{aligned}$$

using Lemma 22. Simplifying this, we get

$$\begin{aligned}\Pr[\psi(m) \in \mathcal{B}(\underline{y}, r)] &\leqslant \frac{r^n}{(r_{\text{eff}}(\Lambda_0))^n}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)^n \\ &\leqslant \frac{r^n}{(r_{\text{cov}}(\Lambda_0))^n}2^{n\delta/8}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)^n \\ &= \left(\frac{r}{\sqrt{nP}}2^{\delta/8}\left(1 + \frac{\sqrt{nP}}{qr}\right)\right)^n,\end{aligned}$$

where we have used Eqn. (15) in the second step. □

[14] More accurately, κ is the integer closest to $nR/\log_2 q$. But we assume that κ as defined above is an integer so that our proofs are cleaner.

[15] Here, we use the natural embedding of $\mathbb{F}_q$ in $\mathbb{Z}$, Φ, to identify elements in $\mathbf{G}_{\text{lin}}$ with the corresponding values in $\mathbb{Z}$. To be rigorous, we should have written $\Phi(\mathbf{G}_{\text{lin}})$ instead of $\mathbf{G}_{\text{lin}}$.

E. Proof of Theorem 21

If $m_1, \ldots, m_\ell$ are linearly independent vectors in $\mathbb{F}_q^\kappa$ and $\mathbf{G}_{\text{lin}}$ is uniform, then $\psi(m_1), \ldots, \psi(m_\ell)$ are statistically independent and hence,

$$\Pr[\psi(m_1), \ldots, \psi(m_\ell) \in \mathcal{B}(\underline{y}, r)] = \left(\Pr[\psi(m) \in \mathcal{B}(\underline{y}, r)]\right)^\ell. \tag{22}$$

Every set of $L+1$ distinct vectors $m_1, \ldots m_{L+1}$ in $\mathbb{F}_q^\kappa$ contains a subset of $\ell := \log_q(L+1)$ linearly independent vectors.

$$\begin{aligned}
&\Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, r) \text{ for some distinct } m_1, \ldots m_{L+1}] \\
&\leqslant \Pr[\psi(m_1), \ldots, \psi(m_\ell) \in \mathcal{B}(\underline{y}, r) \text{ for some linearly independent } m_1, \ldots m_{L+1}] \\
&\leqslant \binom{2^{nR}}{\ell} \Pr[\psi(m_1), \ldots, \psi(m_\ell) \in \mathcal{B}(\underline{y}, r)] \qquad (23) \\
&\leqslant 2^{nR\ell} \Pr[\psi(m_1), \ldots, \psi(m_\ell) \in \mathcal{B}(\underline{y}, r)],
\end{aligned}$$

where in Eqn. (23), $m_1, \ldots, m_\ell$ is a fixed (but arbitrary) set of linearly independent vectors in $\mathbb{F}_q^\kappa$. Using Eqn. (21) and (22) in the above, we get

$$\begin{aligned}
&\Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, r) \text{ for some distinct } m_1, \ldots m_{L+1}] \\
&\leqslant 2^{nR\ell} \left(\frac{r}{\sqrt{nP}} 2^{\delta/8} \left(1 + \frac{\sqrt{nP}}{qr}\right)\right)^{n\ell},
\end{aligned}$$

and hence,

$$\begin{aligned}
&\frac{1}{n} \log_2 \Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, r) \text{ for some distinct } m_1, \ldots m_{L+1}] \\
&\leqslant \ell \left(R - \log_2 \left(\frac{\sqrt{nP}}{r} \right) + \frac{\delta}{8} + \log_2 \left(1 + \frac{\sqrt{nP}}{qr} \right) \right). \qquad (24)
\end{aligned}$$

This suggests that if R and r are not too large, then for any fixed but arbitrary $\underline{y}$, the probability that there are more than L lattice points within distance r of $\underline{y}$ is small. We want to show that this happens for every $\underline{y} \in \mathbb{R}^n$. First, observe that if $\underline{y} \notin \mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN})$, then all codewords are at least $\sqrt{nN}$-away from $\underline{y}$. Therefore, it is enough to consider only those $\underline{y}$ in $\mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN})$. A second observation is that if (for a positive integer α) $Q(\underline{y})$ denotes the closest point in $\frac{1}{\alpha}\Lambda_0$ to $\underline{y}$, then

$$\begin{aligned}
&\Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, \sqrt{nN}) \text{ for some distinct } m_1, \ldots m_{L+1}] \\
&\leqslant \Pr\left[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}\left(Q(\underline{y}), \sqrt{nN} + \frac{r_{\text{cov}}(\Lambda_0)}{\alpha}\right) \text{ for some distinct } m_1, \ldots m_{L+1}\right].
\end{aligned}$$

The idea here is to quantize the $\underline{y}$'s using $\frac{1}{\alpha}\Lambda_0$ and then use a union bound. We want to make sure that α is sufficiently large, but not too large. Specifically, α is the smallest integer greater than $\sqrt{P/N}/(2^{\delta/8} - 1)$. Therefore, α satisfies

$$1 + \frac{1}{\alpha}\sqrt{\frac{P}{N}} < 2^{\delta/8}, \tag{25}$$

and

$$\alpha < \frac{\sqrt{P/N}}{(2^{\delta/8} - 1)} + 2. \tag{26}$$

Note that $\alpha = \Theta(1/\delta)$.

Letting $r = \sqrt{nN} + \frac{r_{\text{cov}}(\Lambda_0)}{\alpha} = \sqrt{nN} + \frac{\sqrt{nP}}{\alpha}$, we have

$$\begin{aligned}
&\Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, \sqrt{nN}) \text{ for some distinct } m_1, \ldots m_{L+1} \text{ and } \underline{y} \in \mathbb{R}^n] \\
&= \Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, \sqrt{nN}) \text{ for some distinct } m_1, \ldots m_{L+1} \text{ and } \underline{y} \in \mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN})] \\
&\leqslant \Pr\left[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, r) \text{ for some distinct } m_1, \ldots m_{L+1} \text{ and } \underline{y} \in \frac{1}{\alpha}\Lambda_0 \cap (\mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN}))\right].
\end{aligned}$$

From Eqn. (15), (16) and the fact that $P > N$, we have

$$\begin{aligned}
\sqrt{nN} &\leqslant \sqrt{nP} = r_{\text{cov}}(\Lambda_0) \\
&\leqslant 2^{\delta/8} r_{\text{eff}}(\Lambda_0) \leqslant 4 \cdot 2^{\delta/8} r_{\text{pack}}(\Lambda_0) \\
&\leqslant 4 \cdot 2^{0.9/8} r_{\text{pack}}(\Lambda_0) < 4.4 r_{\text{pack}}(\Lambda_0).
\end{aligned}$$

Therefore, $\mathcal{B}(0, \sqrt{nN}) \subset 4.4\mathcal{B}(0, r_{\text{eff}}(\Lambda_0)) \subset 4.4\mathcal{V}(\Lambda_0)$. We can therefore take a union bound over $\frac{1}{\alpha}\Lambda_0 \cap (5.4\mathcal{V}(\Lambda_0))$ which gives us[16]

$$\begin{aligned}
&\Pr[\Lambda \cap \mathcal{V}(\Lambda_0) \text{ is not } (P, N, L)\text{-list decodable}] \\
&= \Pr[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, \sqrt{nN}) \text{ for some distinct } m_1, \ldots m_{L+1} \text{ and } \underline{y} \in \mathbb{R}^n] \\
&\leqslant \Pr\left[\psi(m_1), \ldots, \psi(m_{L+1}) \in \mathcal{B}(\underline{y}, r) \text{ for some distinct } m_1, \ldots m_{L+1} \text{ and } \underline{y} \in \frac{1}{\alpha}\Lambda_0 \cap (5.4\mathcal{V}(\Lambda_0))\right].
\end{aligned}$$

Using Eqn. (24) and applying the union bound over $\underline{y}$'s, we have

$$\begin{aligned}
&\frac{1}{n} \log_2 \Pr[\Lambda \cap \mathcal{V}(\Lambda_0) \text{ is not } (P, N, L)\text{-list decodable}] \\
&\leqslant \frac{1}{n} \log_2 \left| \frac{1}{\alpha}\Lambda_0 \cap (5.4\mathcal{V}(\Lambda_0)) \right| \\
&\quad + \ell \left(R - \log_2 \left(\frac{\sqrt{nP}}{r} \right) + \frac{\delta}{8} + \log_2 \left(1 + \frac{\sqrt{nP}}{qr} \right) \right).
\end{aligned}$$

Since $\left|\frac{1}{\alpha}\Lambda_0 \cap (5.4\mathcal{V}(\Lambda_0))\right| \leqslant (5.4\alpha)^n$ and $r \geqslant \sqrt{nN}$, we get

$$\begin{aligned}
&\frac{1}{n} \log_2 \Pr[\Lambda \cap \mathcal{V}(\Lambda_0) \text{ is not } (P, N, L)\text{-list decodable}] \\
&\leqslant \log_2(5.4\alpha) + \ell \left(R - \log_2 \left(\frac{\sqrt{nP}}{\sqrt{nN} + \sqrt{nP}/\alpha} \right) + \frac{\delta}{8} + \log_2 \left(1 + \frac{\sqrt{P}}{q\sqrt{N}} \right) \right) \\
&\leqslant \log_2(5.4\alpha) + \ell \left(R - \frac{1}{2} \log_2 \left(\frac{P}{N} \right) + \log_2 \left(1 + \frac{\sqrt{P}}{\alpha\sqrt{N}} \right) \right.
\end{aligned}$$

[16] We would like to remark that we have not optimized these constants, and the bounds obtained may be loose. However, this will not qualitatively change the overall result.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

$$+\frac{\delta}{8}+\log_2\left(1+\frac{\sqrt{P}}{q\sqrt{N}}\right)\Bigg)$$
$$\leqslant \log_2(5.4\alpha)+\ell\Bigg(-\delta+\log_2\left(1+\frac{\sqrt{p}}{\alpha\sqrt{n}}\right)$$
$$+\frac{\delta}{8}+\log_2\left(1+\frac{\sqrt{P}}{q\sqrt{N}}\right)\Bigg).$$

Using Eqn. (25) and (17),

$$\begin{aligned}&\frac{1}{n}\log_2\Pr[\Lambda\cap\mathcal{V}(\Lambda_0)\text{ is not }(P,N,L)\text{-list decodable}]\\&\leqslant\log_2(5.4\alpha)+\ell\left(-\delta+\frac{\delta}{8}+\frac{\delta}{8}+\frac{\delta}{8}\right)\\&=\log_2(5.4\alpha)-5\ell\frac{\delta}{8}.\end{aligned}\tag{27}$$

From Eqn. (27), we can say that if $\ell > c_1\log_2(\alpha)/\delta$, the probability that a random lattice code is not list decodable goes to zero exponentially in n. For $0<\delta<0.9$, there exist positive constants c_2, c_3 (that could depend on P, N but not on δ) so that $c_2\delta < 2^{\delta/8}-1 < c_3\delta$ and using Eqn. (26), we can see that $\log_2(\alpha)\leqslant c_4\log_2\frac{1}{\delta}$ for some positive c_4. Likewise, using Eqn. (18), we can show that there exist c_5, c_6 such that $c_5\log_2\frac{1}{\delta}\leqslant\log_2 q\leqslant c_6\log_2\frac{1}{\delta}$ for $\delta\in(0,0.9)$. This implies that we can choose $L=q^\ell-1=2^{\ell\log_2 q}-1$ to be less than $2^{\frac{c}{\delta}\log_2^2\frac{1}{\delta}}$ for a sufficiently large constant c, so that the probability that a random nested lattice code is not (P,N,L)-list decodable goes to zero as $2^{-\Omega(n)}$. This concludes the proof of Theorem 21. □

VI. List Decodability of Infinite Lattices

We now direct our attention to infinite constellations. Recall that an infinite constellation $\mathcal{C}$ is a countably infinite subset of $\mathbb{R}^n$, and every lattice is an infinite constellation.

A. List decoding capacity theorem for infinite constellations

In subsequent sections (see Proposition 28), we show that there exist random ICs with NLD $\frac{1}{2}\log\frac{1}{2\pi eN}-\delta$ which are $\left(N,\mathcal{O}\left(\frac{1}{\delta}\log\frac{1}{\delta}\right)\right)$-list decodable. We also show that there exist Construction-A lattices with NLD $\frac{1}{2}\log\frac{1}{2\pi eN}-\delta$ which are $\left(N,2^{\mathcal{O}\left(\frac{1}{\delta}\log^2\frac{1}{\delta}\right)}\right)$-list decodable.

In Appendix D, we also give a converse argument showing that *no* IC of NLD $\frac{1}{2}\log\frac{1}{2\pi eN}+\delta$ can be $\left(N,2^{\mathcal{O}(\delta n)}\right)$-list decodable. Therefore, combining Proposition 28 and Proposition 44, we get the following list decoding capacity theorem for ICs.

Theorem 24. *For any $N>0$, $C(N)=\frac{1}{2}\log\frac{1}{2\pi eN}$.*

B. List size upper bound for infinite lattices

We claim that list decodability of nested Construction-A lattice codes implies a list decoding result for infinite Construction-A lattices.

Definition 7 ($\bmod\Lambda_0$ list decodability)**.** Let (Λ_0,Λ) be a nested lattice pair with $\Lambda\subset\Lambda_0$. The nested lattice code $\Lambda\cap\mathcal{V}(\Lambda_0)$ is said to be (P,N,L)-*list decodable* $\bmod\Lambda_0$ if $r_{\text{cov}}(\Lambda_0)\leqslant\sqrt{nP}$ and for every $\underline{y}\in\mathbb{R}^n$, $\left|\Lambda\cap\mathcal{V}(\Lambda_0)\cap\left([\mathcal{B}(\underline{y},\sqrt{nN})]\bmod\Lambda_0\right)\right|\leqslant L$.

We observe that if $r_{\text{pack}}(\Lambda_0)>\sqrt{nN}$, the ball $\mathcal{B}(\underline{y},\sqrt{nN})$ centered at any point $\underline{y}\in\mathbb{R}^n$ will have no overlap after the $\bmod\Lambda_0$ operation. That is, the following lemma holds.

Lemma 25. *Let $\Lambda_0\subset\mathbb{R}^n$ be a lattice of packing radius*

$$r_{\text{pack}}(\Lambda_0)>\sqrt{nN}\tag{28}$$

for some constant $N>0$. Then for any $\underline{y}\in\mathbb{R}^n$ and $r\leqslant\sqrt{nN}$, we have $\left|[\mathcal{B}(\underline{y},r)]\bmod\Lambda_0\right|=\left|\mathcal{B}(\underline{y},r)\right|$.

Proof. Fix an arbitrary $\underline{y}\in\mathbb{R}^n$ and $r\leqslant\sqrt{nN}$. To show that the $\bmod\Lambda_0$ operation does not create overlap, it suffices to show that no two distinct points in $\mathcal{B}(\underline{y},r)$ will be mapped to the same point. The proof is by contradiction. Suppose there are two distinct points $\underline{x}_1\neq\underline{x}_2\in\mathcal{B}(\underline{y},r)$ such that $[\underline{x}_1]\bmod\Lambda_0=[\underline{x}_2]\bmod\Lambda_0$. One can decompose $\underline{x}_1$ and $\underline{x}_2$ as $\underline{x}_1=\widetilde{\underline{x}}_1+[\underline{x}_1]\bmod\Lambda_0$ and $\underline{x}_2=\widetilde{\underline{x}}_2+[\underline{x}_2]\bmod\Lambda_0$ for unique $\widetilde{\underline{x}}_1,\widetilde{\underline{x}}_2\in\Lambda_0$, respectively. Therefore, $\|\underline{x}_1-\underline{x}_2\|_2=\|\widetilde{\underline{x}}_1-\widetilde{\underline{x}}_2\|_2$, by $[\underline{x}_1]\bmod\Lambda_0=[\underline{x}_1]\bmod\Lambda_0$. On one hand, since both $\underline{x}_1$ and $\underline{x}_2$ are in $\mathcal{B}(\underline{y},r)$, $\|\underline{x}_1-\underline{x}_2\|_2\leqslant 2r\leqslant 2\sqrt{nN}$. On the other hand, since both $\widetilde{\underline{x}}_1$ and $\widetilde{\underline{x}}_2$ are in Λ_0, $\|\widetilde{\underline{x}}_1-\widetilde{\underline{x}}_2\|_2\geqslant 2r_{\text{pack}}(\Lambda_0)>2\sqrt{nN}$. This leads to a contradiction and the proof is finished. □

Lemma 26. *Let (Λ_0,Λ) be a pair of nested lattices with $\Lambda_0\subset\Lambda$ and $r_{\text{pack}}(\Lambda_0)>\sqrt{nN}$. Suppose that the nested lattice code $\Lambda\cap\mathcal{V}(\Lambda_0)$ is (P,N,L)-list decodable $\bmod\Lambda_0$. Then, the infinite lattice Λ is (N,L)-list decodable.*

Proof. The infinite lattice Λ is (N,L)-list decodable if for every $\underline{y}\in\mathbb{R}^n$, we have $\left|\Lambda\cap\mathcal{B}(\underline{y},\sqrt{nN})\right|\leqslant L$. Due to the periodic structure of Λ, we have the property that $\{\underline{x}+\mathcal{V}(\Lambda):\underline{x}\in\Lambda\}$ forms a partition of $\mathbb{R}^n$. Therefore, it suffices to show that $\left|\Lambda\cap\mathcal{B}(\underline{y},\sqrt{nN})\right|\leqslant L$ for all $\underline{y}\in\mathcal{V}(\Lambda)$ in order to prove (N,L)-list decodability of Λ.

But we already have this from the $\bmod\Lambda_0$ list decodability of $\Lambda\cap\mathcal{V}(\Lambda_0)$. Specifically, since $r_{\text{pack}}(\Lambda_0)>\sqrt{nN}$, by Lemma 25 , $\left|\Lambda\cap\mathcal{V}(\Lambda_0)\cap\left([\mathcal{B}(\underline{y},\sqrt{nN})]\bmod\Lambda_0\right)\right|=\left|\Lambda\cap\mathcal{B}(\underline{y},\sqrt{nN})\right|\leqslant L$. Since $\mathcal{V}(\Lambda)\subset\mathcal{V}(\Lambda_0)$, $\bmod\Lambda_0$ (P,N,L)-list decodability of $\Lambda\cap\mathcal{V}(\Lambda_0)$ implies (N,L)-list decodability of Λ. □

A careful inspection of the proof (in particular, the proof of Lemma 23) shows that Theorem 20 actually holds under the notion of $\bmod\Lambda_0$ list decodability (as per Definition 7). Using Lemma 26 and Theorem 20, we have

Theorem 27. *For any constant $0<\delta<0.9$, let Λ be a random Construction-A lattice drawn from the ensemble of Sec. V-D with Λ_0 having covering radius $2\sqrt{nN}$ and $r_{\text{cov}}(\Lambda_0), r_{\text{pack}}(\Lambda_0), q, \kappa$ satisfying (15), (16), (17), (19), (28), respectively. Then,*

1) *the normalized logarithmic density of Λ is $R(\Lambda)\geqslant\frac{1}{2}\log\frac{1}{2\pi eN}-\delta$;*
2) *there exists a constant $c>0$ independent of n,δ such that*

$$\Pr[\Lambda\text{ is not }(N,2^{c\frac{1}{\delta}\log^2\frac{1}{\delta}})\text{-list decodable}]=o(1).$$

Proof. Fix $P = 4N$. Since Λ_0 is good for covering, we have $r_{\text{eff}}(\Lambda_0) \geqslant r_{\text{cov}}(\Lambda_0)2^{-\delta/8} = 2^{1-\delta/8}\sqrt{nN}$. We know that with high probability $\Lambda \cap \mathcal{V}(\Lambda_0)$ is $(P, N, 2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})})$-list decodable, where the implied constant can only depend on N. From Lemma 26, we know that Λ is also $(N, 2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})})$-list decodable with high probability. To complete the proof, it suffices to compute the NLD of Λ.

The rate of the nested lattice code

$$R(\Lambda \cap \mathcal{V}(\Lambda_0)) = \frac{1}{2}\log\frac{P}{N} - \delta = \frac{1}{2}\log\frac{4N}{N} - \delta = 1 - \delta.$$

The NLD of the infinite lattice can hence be bounded as follows,

$$\begin{aligned}
R(\Lambda) &= \frac{1}{n}\log\frac{2^{nR}}{|\mathcal{V}(\Lambda_0)|} \\
&= \frac{1}{n}\log\frac{2^{n(1-\delta)}}{|\mathcal{B}^n(0, r_{\text{eff}}(\Lambda_0))|} \\
&= \log\frac{2^{1-\delta}}{V_n^{1/n} r_{\text{eff}}(\Lambda_0)} \\
&\asymp \log\frac{2^{1-\delta}}{\sqrt{\pi n}^{-1/n}\sqrt{2\pi e/n}\, r_{\text{eff}}(\Lambda_0)} \\
&\geqslant \log\frac{2^{1-\delta}}{\sqrt{2\pi e/n}\sqrt{4nN}} + o(1) \\
&= \frac{1}{2}\log\frac{1}{2\pi e N} - \delta + o(1).
\end{aligned}$$

This completes the proof. □

C. Remark

We proved the above theorem for the random infinite lattice $\mathbf{G}_{\Lambda_0}\Lambda'(\mathcal{C}_{\text{lin}})$, where $\Lambda'(\mathcal{C}_{\text{lin}}) = \Phi(\mathcal{C}_{\text{lin}}) + q\mathbb{Z}^n$ is the "standard" Construction-A lattice obtained from a random linear code $\mathcal{C}_{\text{lin}}$, and $\mathbf{G}_{\Lambda_0}$ is a generator matrix of the coarse lattice. We could have instead proved a similar list decoding result for $\Lambda'(\mathcal{C}_{\text{lin}})$ by following the same approach as in Sec. V, but instead taking a union bound on $\underline{y}$'s within $[0, q)^n$. Doing so would also give a list size of $2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})}$ for all NLD satisfying $R(\Lambda'(\mathcal{C}_{\text{lin}})) \leqslant \frac{1}{2}\log\frac{1}{2\pi e N} - \delta$.

Similarly, for lattice codes, via essentially the same arguments, it can be shown that nested random Construction-A lattice codes $\frac{1}{\alpha}\Lambda'(\mathcal{C}_{\text{lin}}) \cap \Lambda'(\mathcal{C}_{\text{lin}})$ and random Construction-A lattices with ball shaping $\frac{1}{\alpha}\Lambda'(\mathcal{C}_{\text{lin}}) \cap \mathcal{B}(0, \sqrt{nP})$ for proper scaling $1/\alpha$ so as to achieve rate $\frac{1}{2}\log\frac{P}{N} - \delta$ are also $(P, N, 2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})})$-list decodable whp.

VII. List Decodability of Regular Infinite Constellations

Having established a list decoding result for infinite lattices, we now turn to the problem of determining optimal list sizes for infinite constellations. Do there exist ICs $\mathcal{C}$ for which the list size is at most $\mathcal{O}(\text{poly}(1/\delta))$ for all $R(\mathcal{C}) \leqslant \frac{1}{2}\log\frac{1}{2\pi e N} - \delta$?

To study this, we define an ensemble of periodic infinite constellations. We call this a (Λ_0, q, M) infinite constellation (IC) which is defined as follows. Let Λ_0 be a (full rank) lattice satisfying Eqn. (15) and (16). Let q be a prime. A (Λ_0, q, M) random IC $\mathcal{C}$ is obtained by selecting M points $\mathcal{C}' = \{\underline{\mathbf{x}}_1, \ldots, \underline{\mathbf{x}}_M\}$ independently and uniformly at random from $\mathcal{V}(\Lambda_0) \cap \frac{1}{q}\Lambda_0$ and then tiling. Therefore, $\mathcal{C} = \mathcal{C}' + \Lambda_0$. See Fig. 2 for a pictorial illustration of the construction of such an IC ensemble.

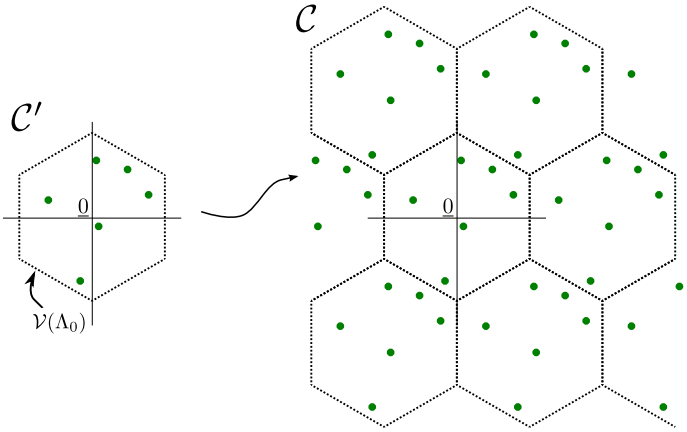


Fig. 2: Illustration of the class of infinite constellations studied in Sec. VII. The finite set $\mathcal{C}'$ is tessellated using Λ_0 to obtain $\mathcal{C} = \mathcal{C}' + \Lambda_0$

The reason why we introduce this new class of ICs is because it is very simple to work with. We can very easily prove several nice properties that would be otherwise complicated for random lattices. We feel that this is a natural counterpart of uniformly random codes over finite fields. Moreover, we can obtain a code for the power-constrained channel by taking the intersection of the IC with $\mathcal{B}(0, \sqrt{nP})$.

Remark 6. In fact, one can define a hierarchy of more and more "uniform" ICs in a similar manner. Under the same construction $\mathcal{C} = \mathcal{C}' + \Lambda_0$ as above, we can choose $\mathcal{C}'$ such that

1) each point in $\mathcal{C}'$ is independent and uniformly distributed in $\mathcal{V}(\Lambda_0)$;
2) each point in $\mathcal{C}'$ is independent and uniformly distributed in $\mathcal{V}(\Lambda_0) \cap \frac{1}{q}\Lambda_0$; (This choice is the same as that in the above paragraph.)
3) $\mathcal{C}'$ is a random subset of $\mathcal{V}(\Lambda_0) \cap \frac{1}{q}\Lambda_0$ that forms a group under addition modulo Λ_0.

The above three constructions are decreasingly "uniform". In particular, the last construction of $\mathcal{C}$ forms a lattice.

We will study the list decodability property of the second construction. The same quantitative results in this section hold for the first construction as well since the latter is more uniform. In Appendix C, we study other goodness properties of the first construction. Properties are easiest to prove under this construction.

Before presenting our formal results, we need one more definition. The effective radius of an infinite constellation is defined as the radius of the n-dimensional ball having volume equal to $1/\Delta(\mathcal{C})$, i.e.,

$$r_{\text{eff}}(\mathcal{C}) := \left(\frac{1}{V_n \Delta(\mathcal{C})}\right)^{1/n}$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

where V_n denotes the volume of a unit n-ball.

A. List size upper bound

We now study list decodability properties of (Λ_0, q, M) random ICs.

An infinite constellation $\mathcal{C}$ is (N, L)-list decodable if for every $\underline{y} \in \mathbb{R}^n$, we have $|\mathcal{C} \cap \mathcal{B}(\underline{y}, \sqrt{nN})| \leqslant L$.

Proposition 28. *Fix a small $\delta > 0$. Let $C := \frac{1}{2}\log\frac{1}{2\pi e N}$. For any $N > 0$, the ensemble of random (Λ_0, q, M) IC $\mathcal{C}$ defined at the beginning of this section with $M = 2^{n(C-\delta)}|\mathcal{V}(\Lambda_0)|$ chosen so as to satisfy $R(\mathcal{C}) = C - \delta$ is $(N, \mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta}))$-list decodable with probability at least $1 - 2^{-\Theta(n)}$.*

Proof. We choose Λ_0 such that $r_{\text{pack}}(\Lambda_0) = 2\sqrt{nN}$. Let q be the smallest integer such that

$$\log\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{q\sqrt{nN}}\right) < 2^{\delta/8}. \tag{29}$$

This implies the following lower and upper bounds on q:

$$\begin{aligned} q &\geqslant \frac{r_{\text{cov}}(\Lambda_0)/\sqrt{nN}}{2^{\delta/8}-1} \geqslant \frac{r_{\text{pack}}(\Lambda_0)/\sqrt{nN}}{2^{\delta/8}-1} \\ &= \frac{2}{2^{\delta/8}-1}; \end{aligned} \tag{30}$$

$$\begin{aligned} q &\leqslant \frac{r_{\text{cov}}(\Lambda_0)/\sqrt{nN}}{2^{\delta/8}-1} + 2 \\ &\leqslant \frac{2^{\delta/8}\cdot r_{\text{eff}}(\Lambda_0)/\sqrt{nN}}{2^{\delta/8}-1} + 2 \\ &\leqslant \frac{2^{\delta/8}\cdot 4\cdot r_{\text{pack}}(\Lambda_0)/\sqrt{nN}}{2^{\delta/8}-1} + 2 \\ &= \frac{2^{3+\delta/8}}{2^{\delta/8}-1} + 2. \end{aligned} \tag{31}$$

Using the elementary inequality

$$\frac{11}{\delta} < \frac{1}{2^{\delta/8}-1} < \frac{12}{\delta} \quad \text{for } 0 < \delta \leqslant 1, \tag{32}$$

we get the following looser bounds:

$$q \geqslant \frac{22}{\delta}, \quad q \leqslant 2^{4+1/8}\cdot\frac{12}{\delta} + \frac{2}{\delta} \leqslant \frac{107}{\delta}. \tag{33}$$

Fix a $\delta > 0$, and choose M such that $R(\mathcal{C}) = C - \delta$. We will show that such random (Λ_0, q, M) ICs are list decodable with constant list sizes. The proof is quite standard so we only give a brief outline. Using q, we define a net for $\mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN})$ as follows: $\mathcal{Y} := \frac{1}{q}\Lambda_0 \cap (\mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN}))$. By the same argument as in the proof of Theorem 21, we have

$$\mathcal{Y} \subset \frac{1}{q}\Lambda_0 \cap (5.4\mathcal{V}(\Lambda_0)),$$

and $|\mathcal{Y}| \leqslant (5.4q)^n$. Let $r := \sqrt{nN} + q^{-1}r_{\text{cov}}(\Lambda_0)$. For a set $\mathcal{A} \subset \mathbb{R}^n$, define the shorthand notation $\mathcal{A}^* := [\mathcal{A}] \bmod \Lambda_0$. For any $\underline{\mathbf{x}} \in \mathcal{C}'$, by Lemma 22,

$$\Pr\left[\underline{\mathbf{x}} \in \mathcal{B}^*(\underline{y}, r)\right] \leqslant \left(\frac{r}{r_{\text{cov}}(\Lambda_0)}2^{\delta/8}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right)^n.$$

For any $\underline{y}$, since each point in $\mathcal{C}'$ is independent, for any $1 \leqslant i_1 < \cdots < i_{L+1} \leqslant M$,

$$\begin{aligned} \Pr\left[\forall j \in [L+1],\ \underline{\mathbf{x}}_{i_j} \in \mathcal{B}^*(\underline{y}, r)\right] &= \prod_{j\in[L]} \Pr\left[\underline{\mathbf{x}}_{i_j} \in \mathcal{B}^*(\underline{y}, r)\right] \\ &\leqslant \left(\frac{r}{r_{\text{cov}}(\Lambda_0)}2^{\delta/8}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right)^{n(L+1)}. \end{aligned}$$

By union bound,

$$\begin{aligned} &\Pr\left[\exists L+1 \text{ codewords in } \mathcal{B}^*(\underline{y}, r)\right] \\ &\leqslant \binom{M}{L+1}\left(\frac{r}{r_{\text{cov}}(\Lambda_0)}2^{\delta/8}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right)^{n(L+1)}. \end{aligned}$$

Finally,

$$\begin{aligned} &\Pr\left[\text{The IC is not } (N, L) \text{ list decodable}\right] \\ &\leqslant \Pr[\exists \underline{y} \in \mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN}),\ \exists L+1 \text{ distinct codewords in } \mathcal{B}(\underline{y}, \sqrt{nN})] \\ &\leqslant \Pr[\exists \underline{y} \in \mathcal{Y},\ \exists L+1 \text{ distinct codewords in } \mathcal{B}(\underline{y}, \sqrt{nN} + r_{\text{cov}}(q^{-1}\Lambda_0))] \\ &\leqslant \sum_{\underline{y}\in\mathcal{Y}} \Pr\left[\exists L+1 \text{ codewords in } \mathcal{B}^*(\underline{y}, \sqrt{nN} + q^{-1}r_{\text{cov}}(\Lambda_0))\right] \\ &\leqslant |\mathcal{Y}|\binom{M}{L+1}\left(\frac{r}{r_{\text{cov}}(\Lambda_0)}2^{\delta/8}\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right)^{n(L+1)}. \end{aligned}$$

We then work with the exponent.

$$\begin{aligned} &\frac{1}{n}\log\Pr\left[\text{The IC is not } (N, L) \text{ list decodable}\right] \\ &\leqslant \log(5.4q) + (L+1)\left[R + \frac{1}{n}\log\left(V_n r_{\text{eff}}(\Lambda_0)^n\right)\right. \\ &\qquad \left. + \log\frac{r}{r_{\text{cov}}(\Lambda_0)} + \frac{\delta}{8} + \log\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right]. \end{aligned}$$

We note that

$$\begin{aligned} &\frac{1}{n}\log\left(V_n r_{\text{eff}}(\Lambda_0)^n\right) \\ &\leqslant \frac{1}{n}\log\left(V_n r_{\text{cov}}(\Lambda_0)^n\right) \\ &= \frac{1}{n}\log V_n + \log r_{\text{cov}}(\Lambda_0) \\ &\overset{n\to\infty}{\asymp} \frac{1}{n}\log\left[\frac{1}{\sqrt{\pi n}}\left(\frac{2\pi e}{n}\right)^{n/2}\right] + \log r_{\text{cov}}(\Lambda_0) \\ &\overset{n\to\infty}{\asymp} \frac{1}{2}\log\frac{2\pi e}{n} + \log r_{\text{cov}}(\Lambda_0) \\ &= \frac{1}{2}\log\left(\frac{2\pi e}{n}r_{\text{cov}}(\Lambda_0)^2\right). \end{aligned}$$

Therefore

$$\begin{aligned} &\frac{1}{n}\log\Pr\left[\text{The IC is not } (N, L) \text{ list decodable}\right] \\ &\leqslant \log(5.4q) + (L+1)\left[R + \frac{1}{2}\log\left(\frac{2\pi e}{n}r_{\text{cov}}(\Lambda_0)^2\right)\right. \\ &\qquad \left. + \log\frac{r}{r_{\text{cov}}(\Lambda_0)} + \frac{\delta}{8} + \log\left(1 + \frac{r_{\text{cov}}(\Lambda_0)}{qr}\right)\right] + o(1) \end{aligned}$$

$$= \log(5.4q) + (L+1)\left[R + \frac{1}{2}\log\left(\frac{2\pi e r^2}{n}\right) + \frac{\delta}{8} + \log\left(1 + \frac{r_{\mathrm{cov}}(\Lambda_0)}{qr}\right)\right] + o(1).$$

By the choice of r, we know

$$\frac{r^2}{n} = \frac{1}{n}\left(\sqrt{nN} + \frac{r_{\mathrm{cov}}(\Lambda_0)}{q}\right)^2 = N\left(1 + \frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)^2.$$

The exponent can be simplified to

$$\begin{aligned}
&\log(5.4q) + (L+1)\left[R + \frac{1}{2}\log(2\pi e N) + \log\left(1 + \frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right) + \frac{\delta}{8} + \log\left(1 + \frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right] \\
&\leqslant \log(5.4q) + (L+1)\left(-\delta + \frac{\delta}{8} + \frac{\delta}{8} + \frac{\delta}{8}\right) \\
&= \log(5.4q) - \frac{5}{8}\delta(L+1).
\end{aligned}$$

Observe that the above exponent is negative if $L > \frac{c_1}{\delta}\log q$ for some constant $c_1 > 0$. Using the condition on q (Eqn. (33)) and following similar calculations that appear at the end of the proof of Theorem 21, we conclude that there exists a constant $c > 0$ such that the probability that a random IC is not list decodable is exponentially small in n if $L > c\frac{1}{\delta}\log\frac{1}{\delta}$. This completes the proof. □

B. List size lower bound

Lemma 29. *Let $\mathcal{C}$ be an (α, M) random IC chosen so as to satisfy $\alpha = 4\sqrt{nN}$ and $R(\mathcal{C}) = C - \delta$ where $C = \frac{1}{2}\log\frac{1}{2\pi e N}$. Then,*

$$\Pr\left[\mathcal{C} \text{ is } \left(N, \mathcal{O}\left(\frac{1}{\delta}\right)\right)\text{-list decodable}\right] = o(1).$$

Proof. The proof is almost identical to that of Proposition 18. We only highlight the main differences here.

Let q be the smallest integer satisfying Eqn. (29) (recall that this implies Eqn. (30) and (31), or more loosely, Eqn. (33)). Define $\mathcal{C} := \mathcal{C}' + \frac{1}{q}\Lambda_0$, where Λ_0 is simultaneously good for covering and packing, i.e., it satisfies both Eqn. (15) and (16); $\mathcal{C}' \subseteq \frac{1}{q}\Lambda_0 \cap \mathcal{V}(\Lambda_0)$ is a set of M uniformly random and independent points $\underline{\mathbf{x}}_1, \cdots, \underline{\mathbf{x}}_M$ in $\frac{1}{q}\Lambda_0 \cap \mathcal{V}(\Lambda_0)$. Scale Λ_0 such that $r_{\mathrm{pack}}(\Lambda_0) = 2\sqrt{nN}$. Let $M = 2^{n(C-\delta)}|\mathcal{V}(\Lambda_0)|$ for some $0 < \delta < 0.9$.

Let $\mathcal{M} := [M]$ and $\mathcal{Y} := \frac{1}{q}\Lambda_0 \cap (\mathcal{V}(\Lambda_0) + \mathcal{B}(0, \sqrt{nN}))$. Define the random variable

$$W := \sum_{\underline{y}\in\mathcal{Y}} \sum_{\{m_1,\cdots,m_L\}\in\binom{\mathcal{M}}{L}} \mathbb{1}\left\{\underline{\mathbf{x}}_{m_1}, \cdots, \underline{\mathbf{x}}_{m_L} \in \mathcal{B}^*(\underline{y}, \sqrt{nN})\right\}, \tag{34}$$

where we use the following notation $\mathcal{A}^* := [\mathcal{A}] \bmod \Lambda_0$ for any $\mathcal{A} \subset \mathbb{R}^n$.

We will upper bound

$$\begin{aligned}
&\Pr\left[\mathcal{C} \text{ is } (N, L)\text{-list decodable}\right] \\
&= \Pr\left[\bigcap_{\underline{y}\in\mathbb{R}^n}\left\{\left|\mathcal{C} \cap \mathcal{B}\left(\underline{y}, \sqrt{nN}\right)\right| \leqslant L\right\}\right] \\
&\leqslant \Pr\left[\bigcap_{\underline{y}\in\mathcal{Y}}\left\{\left|\mathcal{C}' \cap \mathcal{B}^*\left(\underline{y}, \sqrt{nN}\right)\right| \leqslant L\right\}\right], \\
&\leqslant \mathrm{Var}\left[W\right]/\mathbb{E}\left[W\right]^2.
\end{aligned}$$

1) Lower bounding $\mathbb{E}[W]$: It turns out that the expectation

$$\mathbb{E}\left[W\right] = \sum_{\underline{y}\in\mathcal{Y}} \sum_{\mathcal{L}\in\binom{\mathcal{M}}{L}} \Pr\left[\underline{\mathbf{x}}_{\mathcal{L}} \subset \mathcal{B}^*\left(\underline{y}, \sqrt{nN}\right)\right], \tag{35}$$

where $\underline{\mathbf{x}}_{\mathcal{L}} := \{\underline{\mathbf{x}}_m : m \in \mathcal{L}\}$, can be computed precisely.

The probability in the summand of the right-hand side (RHS) of Eqn. (35) is

$$\mu^L := \left(\frac{\left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}^*(\underline{y}, \sqrt{nN})\right|}{\left|\frac{1}{q}\Lambda_0 \cap \mathcal{V}(\Lambda_0)\right|}\right)^L.$$

Overall the expectation in Eqn. (35) equals

$$\mathbb{E}\left[W\right] = |\mathcal{Y}|\binom{M}{L}\mu^L.$$

2) Upper bounding $\mathrm{Var}[W]$: As in the proof of Proposition 18, $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{E}_3$ are similarly defined and their probabilities are similarly bounded.

$$\Pr\left[\mathcal{E}_1\right] = \left(\frac{\left|\mathcal{B}\left(\sqrt{nN}\right) \cap \frac{1}{q}\Lambda_0\right|}{|\mathcal{Y}|}\right)^2 =: \eta^2,$$

$$\Pr\left[\mathcal{E}_2 \cap \mathcal{E}_3 | \mathcal{E}_1\right] \leqslant \mu^{L-1}\mu^{L-\ell} = \mu^{2L-\ell-1}.$$

Overall we have

$$\mathrm{Var}\left[W\right] \leqslant |\mathcal{Y}|^2 \sum_{\ell=1}^{L} M^{2L-\ell}\eta^2\mu^{2L-\ell-1} \leqslant |\mathcal{Y}|^2 L M^L \eta^2 \mu^{L-1}.$$

3) Wrapping things up: The probability that a random infinite constellation is list decodable is at most

$$\frac{\mathrm{Var}\left[W\right]}{\mathbb{E}\left[W\right]^2} \leqslant \frac{|\mathcal{Y}|^2 L M^L \eta^2 \mu^{L-1}}{|\mathcal{Y}|^2 (M/L)^{2L}\mu^{2L}} = L^{2L+1} M^{-L}\eta^2\mu^{-L-1}.$$

We shall upper bound η and lower bound μ.

For η, we have

$$\eta = \frac{\left|\mathcal{B}(\sqrt{nN}) \cap \frac{1}{q}\Lambda_0\right|}{|\mathcal{Y}|} \leqslant \frac{\left|\mathcal{B}(\sqrt{nN}) \cap \frac{1}{q}\Lambda_0\right|}{q^n} \leqslant \left[\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)} 2^{\delta/8}\left(1 + \frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]^n,$$

where the first inequality is because

$$\mathcal{Y} \supset \frac{1}{q}\Lambda_0 \cap \mathcal{V}(\Lambda_0) \implies |\mathcal{Y}| \geqslant q^n.$$

For μ, we have

$$\mu = \frac{\left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}^*(\underline{y}, \sqrt{nN})\right|}{\left|\frac{1}{q}\Lambda_0 \cap \mathcal{V}(\Lambda_0)\right|} = \frac{\left|\frac{1}{q}\Lambda_0 \cap \mathcal{B}(\underline{y}, \sqrt{nN})\right|}{q^n}$$

$$\geqslant \left[\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)} 2^{\delta/8}\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]^n.$$

Therefore,

$$\begin{aligned}
&\frac{1}{n}\log\frac{\mathrm{Var}\left[W\right]}{\left(\mathbb{E}\left[W\right]\right)^2}\\
&\leqslant -L\left[R+\frac{1}{2}\log\left(\frac{2\pi e}{n}r_{\mathrm{cov}}(\Lambda_0)^2\right)\right]\\
&\quad+2\left[\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}+\frac{\delta}{8}+\log\left(1+\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&\quad-(L+1)\left[\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}+\frac{\delta}{8}+\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&=-L\left[R+\frac{1}{2}\log\left(\frac{2\pi e}{n}r_{\mathrm{cov}}(\Lambda_0)^2\right)+\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}\right.\\
&\qquad\left.+\frac{\delta}{8}+\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&\quad+2\left[\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}+\frac{\delta}{8}+\log\left(1+\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&\quad-\left[\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}+\frac{\delta}{8}+\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&=-L\left[R+\frac{1}{2}\log(2\pi eN)+\frac{\delta}{8}+\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\right]\\
&\quad+\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}+\frac{\delta}{8}+2\log\left(1+\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right)\\
&\quad-\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)}{q\sqrt{nN}}\right).
\end{aligned}$$

Recall that q satisfies Eqn. (29) which implies

$$\begin{aligned}
&\log\left(1-\frac{r_{\mathrm{cov}}(\Lambda_0)/\sqrt{nN}}{q}\right)\\
&\quad>\log\left[1-\frac{r_{\mathrm{cov}}(\Lambda_0)/\sqrt{nN}}{(r_{\mathrm{cov}}(\Lambda_0)/\sqrt{nN})/(2^{\delta/8}-1)}\right]\\
&\quad=\log(2-2^{\delta/8})\geqslant-\frac{\delta}{7},
\end{aligned}$$

where the last inequality is true for any $\delta\in(0,1)$. Using the bounds on q, we get

$$\begin{aligned}
\frac{1}{n}\log\frac{\mathrm{Var}\left[W\right]}{\left(\mathbb{E}\left[W\right]\right)^2}&\leqslant -L\left(-\delta+\frac{\delta}{8}-\frac{\delta}{7}\right)+\log\frac{\sqrt{nN}}{r_{\mathrm{cov}}(\Lambda_0)}\\
&\quad+\frac{\delta}{8}+\frac{\delta}{4}+\frac{\delta}{7}.
\end{aligned}$$

Recall the relation $r_{\mathrm{cov}}(\Lambda_0)\geqslant r_{\mathrm{pack}}(\Lambda_0)=2\sqrt{nN}$. Then

$$\frac{1}{n}\log\frac{\mathrm{Var}\left[W\right]}{\left(\mathbb{E}\left[W\right]\right)^2}\leqslant\frac{57}{56}\delta L-1+\frac{29}{56}\delta.$$

This exponent is negative if $L<\frac{1-\frac{29}{56}\delta}{\frac{57}{56}\delta}$, or more loosely, $L<\frac{9}{19\delta}$. □

C. Other goodness properties

The random ICs defined in this section have other interesting geometric properties which are much harder to prove for lattices [11], for instance, packing goodness, AWGN goodness and covering goodness. See Appendix C for statements and proofs.

VIII. Haar Measure on $\mathscr{L}_n$

Let us first ask ourselves: how do we define a *random* lattice? To sample a random lattice from a certain ensemble, we need to define a distribution on the set of all lattices. As we know, a lattice is specified by its generator matrix and thus it suffices to define a distribution over matrices.[17] There are several ensembles of matrices that are extensively studied in the literature of random matrix theory. Such ensembles, including the Gaussian ensemble, the Bernoulli ensemble, etc., [59] are mostly defined by sampling entries iid from simple distributions. However, we believe that such ensembles will *not* give rise to interesting lattices, in the sense that the resulting lattices are not likely to have nontrivial packing and covering efficiencies simultaneously.

We give a heuristic argument to justify the above statement. Suppose that we sample an n by n random matrix $\mathbf{G}$ over $\mathbb{R}$ by sampling each entry iid according to $\mathcal{N}(0,\sigma^2)$ for some fixed constant deviation $\sigma>0$. By the high-dimensional geometry of Gaussian random vectors, each column of $\mathbf{G}$ has L^2-norm highly concentrated around $\sqrt{n\sigma^2}$ and is approximately orthogonal to other columns. That is to say, the columns $\{\underline{\mathbf{g}}_1,\cdots,\underline{\mathbf{g}}_n\}$ of $\mathbf{G}$ are basically a mildly perturbed version of the standard orthonormal basis $\{\underline{e}_1,\cdots,\underline{e}_n\}$ scaled by $\sqrt{n\sigma^2}$ (and potentially also rotated, which does not affect most goodness properties of the resulting lattices we are interested in). The lattice $\mathbf{G}\mathbb{Z}^n$ cannot be good for packing whp since $\mathbb{Z}^n$ (and also its scaling and rotation) has vanishing packing efficiency $\asymp\sqrt{\frac{\pi e}{2n}}$ as the dimension n tends to infinity. Indeed, there is not much study on lattices resulting from those canonical matrix ensembles. One can find related results along this direction in [60]. Not surprisingly, it boils down to understanding the singular value spectrum of $\mathbf{G}$.

In the case of finite fields, it is known that a *uniformly random* linear code has good list decoding properties. It would therefore be a natural choice to study a random lattice drawn uniformly over the set of all lattices of a fixed determinant. Unfortunately, the space of such lattices is unbounded[18] and

[17] Strictly speaking, each lattice is *non-uniquely* identified with a generator matrix. Those matrices giving rise to the same lattice should be quotiented out when one wants to define a distribution on lattices by defining it on matrices. See Sec. IX for more formalisms.

[18] The unboundedness (wrt L^2-distance) of the set of determinant-1 lattices can be seen from the following example. The matrix $\begin{bmatrix} K^{-\frac{1}{n-1}} & & \\ & \ddots & \\ & & K^{-\frac{1}{n-1}} \end{bmatrix}_{n\times n}$ generates a determinant-1 lattice for every value of $K\geqslant 0$. However, the L^2-norm of (the vectorization of) the matrix is $\sqrt{\left(K^{-\frac{1}{n-1}}\right)^2(n-1)+K^2}\to\infty$ as K approaches infinity.

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

hence it does not make sense to talk about uniform distribution on it.

In the following section, we introduce the Haar distribution over lattices, and survey some of the important results pertaining to our discussions. For small enough subsets $\mathcal{B}$ of $\mathbb{R}^n$, we conjecture that the distribution of the number of lattice points (which is a random variable if the lattice is drawn according to the Haar distribution) in $\mathcal{B}$ looks like a Poisson distribution. Expressions for the first $o(n)$ moments of the number of lattice points have been derived in the literature. Encouraged by these results, we make a conjecture about the first $\mathcal{O}(n)$ moments. We then show that if this conjecture is true, then Haar lattices achieve $\mathrm{poly}(1/\delta)$ list sizes.

IX. Prior Work on Haar Lattices

Let us first introduce the Haar distribution on the set of all lattices. A more detailed exposition can be found in the thesis of Kim [61].

For the convenience of illustration, let us collect all rank-n lattices $\Lambda \subset \mathbb{R}^n$ with covolume one[19] into a set $\mathscr{L}_n$:

$$\mathscr{L}_n := \{\Lambda \leqslant \mathbb{R}^n \text{ lattice}\colon \det(\Lambda) = 1\}.$$

A lattice in $\mathscr{L}_n$ is specified by its generator matrix $\mathbf{G} \in \mathrm{SL}(n,\mathbb{R})$. However, one lattice Λ can have multiple different generator matrices. Indeed, two matrices $\mathbf{G}$ and $\widetilde{\mathbf{G}}$ give rise to the same lattice (i.e., $\mathbf{G}\mathbb{Z}^n = \widetilde{\mathbf{G}}\mathbb{Z}^n$) iff they differ by an $\mathrm{SL}(n,\mathbb{Z})$ matrix, i.e., $\mathbf{G}\mathbf{G}' = \widetilde{\mathbf{G}}$ where $\mathbf{G}' \in \mathrm{SL}(n,\mathbb{Z})$. Hence $\mathscr{L}_n$ can be identified with the quotient space

$$\mathscr{L}_n = \mathrm{SL}(n,R)/\mathrm{SL}(n,\mathbb{Z}).$$

Crucial to us is Haar's seminal result on the existence of *Haar measure* on any locally compact topological group. Specialized to our setting, it was shown by Siegel [62] the existence and finiteness of a certain nicely-behaved distribution on $\mathscr{L}_n$.

Theorem 30 ([62])**.** *There is a unique (up to a multiplicative constant factor) measure μ (called the Haar measure or the Haar–Siegel measure) on $\mathrm{SL}(n,\mathbb{R})$ which satisfies the following properties:*

1) *μ is left-$\mathrm{SL}(n,\mathbb{R})$-invariant, i.e., for any Borel subset $\mathcal{K} \subseteq \mathrm{SL}(n,\mathbb{R})$ and any $\mathbf{G} \in \mathrm{SL}(n,\mathbb{R})$, $\mu(\mathcal{K}) = \mu(\mathbf{G}\mathcal{K})$;*
2) *μ is finite, i.e., for any compact subset $\mathcal{K} \subseteq \mathrm{SL}(n,\mathbb{R})$, $\mu(\mathcal{K}) < \infty$.*

Note that we can normalize the Haar measure μ to make it a probability distribution, i.e., $\mu(\mathrm{SL}(n,\mathbb{R})) = 1$. In this paper we always refer to the normalized version when talking about μ or Haar measure. The Haar distribution on $\mathscr{L}_n$ naturally inherits that on $\mathrm{SL}(n,\mathbb{R})$. We do not specify measure-theoretic details which can be found in, e.g., [63]. With abuse of notation, we use the same notation μ for the Haar measure on $\mathrm{SL}(n,\mathbb{R})$ and the induced Haar measure on $\mathscr{L}_n$. Most of the time we refer to the former one which will be clear from the context.

The above result only provides the existence and properties of the Haar measure but does not provide an explicit form of this measure. What does the Haar measure μ on $\mathrm{SL}(n,\mathbb{R})$ look like? It can be checked that the Lebesgue measure on $\mathbb{R}^{n^2}$ satisfies the properties 1 and 2 required in Theorem 30, and hence is the Haar measure. Given any measure, besides that we can use it to measure a compact subset of the space, we can also integrate functions on the same space against this measure. Since Haar measure is unique, we know that for $\mathbf{G} \in \mathrm{SL}(n,\mathbb{R})$,

$$\mathrm{d}\mu(\mathbf{G}) = \mathrm{dvec}\,(\mathbf{G}),$$

where $\mathrm{vec}\,(\mathbf{G}) \in \mathbb{R}^{n^2}$ denotes the vectorization of $\mathbf{G}$. Namely, the Haar measure of a (measurable) set of matrices in $\mathrm{SL}(n,\mathbb{R})$ is equal to the Lebesgue measure of it when viewed as a set of vectors in $\mathbb{R}^{n^2}$.

As a byproduct of the above reasoning, we also know that the Haar measure on $\mathrm{GL}(n,\mathbb{R})$ is just the normalized Lebesgue measure on $\mathbb{R}^{n^2}$. For $\mathbf{G} \in \mathrm{GL}(n,\mathbb{R})$,

$$\mathrm{d}\mu(\mathbf{G}) = \frac{\mathrm{dvec}\,(\mathbf{G})}{\det(\mathbf{G})^{1/n}}.$$

It is a valid definition since

$$\det\left(\frac{\mathbf{G}}{\det(\mathbf{G})^{1/n}}\right) = \left(\frac{1}{\det(\mathbf{G})^{1/n}}\right)^n \det(\mathbf{G}) = 1,$$

and the definition is reduced to the one on $\mathrm{SL}(n,\mathbb{R})$. Here again with abuse of notation, we use the same notation for Haar measure on $\mathrm{SL}(n,\mathbb{R})$ and $\mathrm{GL}(n,\mathbb{R})$.

One may resort to Iwasawa (KAN) decomposition [63] for a more explicit characterization of the Haar measure.

A. Siegel's and Rogers's averaging formulas

Let us first recall two fundamental averaging formulas which are heavily used in the literature for understanding the distribution of short vectors of a random lattice drawn from the Haar distribution.

In the same seminal paper [62] in which Siegel showed the existence and uniqueness of Haar distribution on the space of unit-covolume lattices, he also proved the following averaging formula.

Theorem 31 ([62])**.** *Let $\rho : \mathbb{R}^n \to \mathbb{R}$ be a bounded, measurable, compactly supported function. Then*

$$\mathop{\mathbb{E}}_{\Lambda\sim\mu}\left[\sum_{\underline{x}\in\Lambda\setminus\{0\}}\rho(\underline{x})\right] = \int_{\mathscr{L}_n}\sum_{\underline{x}\in\Lambda\setminus\{0\}}\rho(\underline{x})\mathrm{d}\mu(\Lambda) = \int_{\mathbb{R}^n}\rho(\underline{x})\mathrm{d}\underline{x}. \tag{36}$$

Remark 7. The requirement that we are allowed to evaluate the function ρ only at nonzero lattice points could be potentially inconvenient in applications. One can drop this condition by paying an extra term, i.e., the value of ρ at the origin, on the RHS of Eqn. (36) and the formula becomes

$$\mathop{\mathbb{E}}_{\Lambda\sim\mu}\left[\sum_{\underline{x}\in\Lambda}\rho(\underline{x})\right] = \int_{\mathscr{L}_n}\sum_{\underline{x}\in\Lambda}\rho(\underline{x})\mathrm{d}\mu(\Lambda) = \rho(0) + \int_{\mathbb{R}^n}\rho(\underline{x})\mathrm{d}\underline{x}. \tag{37}$$

These two forms are completely equivalent and we will state only one of them but potentially use any of them without further explanation depending on whichever is convenient.

[19] This is without loss of generality since normalization does not affect goodness properties.

The identity holds in large generality for any reasonably nice function ρ. Perhaps the most important consequence of this formula is that it gives a way to estimate the number of lattice points in a measurable set, which is in turn an ubiquitous primitive in applications. Specifically, for our list decoding purposes, essentially the only thing we need to control is the number of lattice points in a ball. If we take

$$\rho(\underline{x}) \coloneqq \mathbb{1}\{\underline{x} \in \mathcal{B}(\underline{y}, r)\}$$

to be the indicator function of an Euclidean ball centered at $\underline{y}$ of radius r (which obviously satisfies the conditions required by Theorem 31), then the left-hand side (LHS) of (36) is nothing but the expected number of nonzero Haar lattice points in the ball. Siegel's formula tells us that this is equal to the RHS of (36) which is actually the volume of the ball. This matches our intuition that the number of lattice points in any measurable set $\mathcal{B}$ should be roughly the ratio between the volume of $\mathcal{B}$ and the volume of a Voronoi cell of the lattice, i.e., $|\Lambda \cap \mathcal{B}| \approx \mathrm{Vol}(\mathcal{B})/\det(\Lambda) = \mathrm{Vol}(\mathcal{B})$ since we consider normalized lattices. Siegel's formula indicates that the Haar distribution on $\mathscr{L}_n$ behaves typically in a sense that such intuition is indeed true in expectation.

One simple application of Theorem 31 is that it allows us to control the rate of a Haar lattice code. For a lattice $\Lambda \sim \mathscr{L}_n$, if we define the lattice code $\mathcal{C}$ to be $\Lambda \cap \mathcal{B}(0, \sqrt{nP})$, then Theorem 31 lets us conclude that

$$\frac{1}{n} \log \mathbb{E}_{\Lambda} \left[|\mathcal{C}|\right] = \frac{1}{2} \log P + o(1).$$

It turns out there is a higher-order generalization of Siegel's formula due to Rogers [64] which we introduce below.

Theorem 32 ([64], Theorem 4)**.** *Let $k < n$ be a positive integer. Let*

$$\rho\colon \quad (\mathbb{R}^n)^k \quad \to \quad \mathbb{R}$$

be a bounded Borel measurable function with compact support. Then

$$\begin{aligned} &\mathop{\mathbb{E}}_{\Lambda\sim\mu} \left[\sum_{\underline{x}_1,\cdots,\underline{x}_k \in \Lambda} \rho(\underline{x}_1, \cdots, \underline{x}_k) \right] \\ &= \int_{\mathscr{L}_n} \sum_{\underline{x}_1,\cdots,\underline{x}_k \in \Lambda} \rho\left(\underline{x}_1, \cdots, \underline{x}_k\right) \mathrm{d}\mu\left(\Lambda\right) \qquad (38) \\ &= \rho\left(0, \cdots, 0\right) + \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} \rho\left(\underline{x}_1, \cdots, \underline{x}_k\right) \mathrm{d}\underline{x}_1 \cdots \mathrm{d}\underline{x}_k + \mathcal{E}, \end{aligned}$$

where $\mathcal{E}$ is an error term defined as follows:

$$\begin{aligned} \mathcal{E} \coloneqq \sum_{(\vec{\alpha},\vec{\beta})} \sum_{\ell=1}^{\infty} \sum_{\mathbf{D}} \Bigg[&\left(\frac{e_1}{\ell} \cdots \frac{e_m}{\ell}\right)^n \times \\ &\int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} \rho\left(\sum_{i=1}^{m} \frac{\mathbf{D}_{i1}}{\ell} \underline{x}_i, \cdots, \sum_{i=1}^{m} \frac{\mathbf{D}_{ik}}{\ell} \underline{x}_i\right) \mathrm{d}\underline{x}_1 \cdots \mathrm{d}\underline{x}_m \Bigg]. \end{aligned}$$

Here the first sum is over all divisions $(\vec{\alpha}, \vec{\beta}) = (\alpha_1, \cdots, \alpha_m; \beta_1, \cdots, \beta_{k-m})$ of the numbers $1, \cdots, k$ into two sequences $1 \leqslant \alpha_1 < \cdots < \alpha_m \leqslant k$ and $1 \leqslant \beta_1 < \cdots < \beta_{k-m} \leqslant k$ with $1 \leqslant m \leqslant k-1$ and $\alpha_i \neq \beta_j$ for any i, j. The third sum is taken over all integral $m \times k$ matrices $\mathbf{D} \in \mathbb{Z}^{m\times k}$ such that

1) *no column of $\mathbf{D}$ vanishes;*
2) *the greatest common divisor of all entries is 1;*
3) *for all $i \in [m], s \in [m], t \in [k-m]$, $\mathbf{D}_{i\alpha_s} = \ell \mathbb{1}\{i = s\}$ and $\mathbf{D}_{i\beta_t} = 0$ if $\beta_t < \alpha_i$.*

Finally, $e_i = (\gamma_i, \ell)$, where $\gamma_1, \cdots, \gamma_m$ are the elementary divisors (cf. [65]) of $\mathbf{D}$.

If we take

$$\rho\left(\underline{x}_1, \cdots, \underline{x}_k\right) \coloneqq \mathbb{1}\{\underline{x}_1 \in \mathcal{B}\} \cdots \mathbb{1}\{\underline{x}_k \in \mathcal{B}\},$$

where $\mathcal{B} \coloneqq \mathcal{B}(\underline{y}, r)$ is a ball, then Rogers's formula is precisely computing

$$\mathop{\mathbb{E}}_{\Lambda\sim\mu} \left[|\Lambda \cap \mathcal{B}|^k\right]$$

for $1 \leqslant k \leqslant n-1$.

The proof of Rogers's averaging formula is highly nontrivial and can be divided into three steps. Since the proof contains several ingenious ideas and can be instructive for other purposes, we sketch it below.

Step I. Consider any real-valued bounded Borel measurable function of bounded support on unit-covolume lattices,

$$f\colon \quad \mathscr{L}_n \quad \to \quad \mathbb{R}.$$

We will interchangeably think of f as a function on $\mathrm{SL}\left(n, \mathbb{R}\right)$,

$$f\colon \quad \mathrm{SL}\left(n, \mathbb{R}\right) \quad \to \quad \mathbb{R}$$

by interchangeably thinking of Λ as a lattice or its generator matrix. The function f can be naturally extended from $\mathrm{SL}\left(n, \mathbb{R}\right)$ to $\mathrm{GL}\left(n, \mathbb{R}\right)$ by defining, for $\Lambda \in \mathrm{GL}\left(n, \mathbb{R}\right)$,

$$f\left(\Lambda\right) \coloneqq f|_{\mathrm{SL}(n,\mathbb{R})} \left(\det\left(\Lambda\right)^{-1/n} \Lambda\right). \qquad (39)$$

Note that $\det\left(\Lambda\right)^{-1/n} \Lambda$ always has determinant one.

Fix $\omega \in \mathbb{R}_{>0}$. Let $\mathbf{\Theta} = \mathbf{\Theta}\left(\theta_1, \cdots, \theta_{n-1}, \omega\right) \in \mathbb{R}^{n\times n}$ be drawn from the following ensemble

$$\begin{bmatrix} \omega & & & & \\ & \omega & & & \\ & & \ddots & & \\ & & & \omega & \\ \omega^{-(n-1)}\theta_1 & \omega^{-(n-1)}\theta_2 & \cdots & \omega^{-(n-1)}\theta_{n-1} & \omega^{-(n-1)} \end{bmatrix}, \qquad (40)$$

where each $\theta_i \sim \mathcal{U}\left([0,1]\right)$. Note that any matrix of the above form has determinant one.

Remark 8. The reason behind the choice of this ensemble has connections to number theory. This is well beyond the scope of this paper and we refer interested readers to [66], [67] for relevant background.

Let $\vec{\theta} \coloneqq (\theta_1, \cdots, \theta_{n-1})$. The average of f wrt such an ensemble can be written as

$$\begin{aligned} &\mathop{\mathbb{E}}_{\theta\sim\mathcal{U}([0,1])} \left[f\left(\mathbf{\Theta}(\vec{\theta}, \omega)\mathbb{Z}^n\right)\right] \\ &\qquad = \int_0^1 \cdots \int_0^1 f\left(\mathbf{\Theta}(\vec{\theta}, \omega)\mathbb{Z}^n\right) \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{n-1}. \end{aligned}$$

Let

$$M(f) := \lim_{\omega\to 0+} \mathop{\mathbb{E}}_{\theta\sim\mathcal{U}([0,1])} \left[f\left(\mathbf{\Theta}(\vec{\theta},\omega)\mathbb{Z}^n\right)\right].$$

Rogers [64] showed the following (perhaps surprising) identity.

Theorem 33 ([64], Theorem 1)**.** *Let $\rho : \mathscr{L}_n \to \mathbb{R}$ be a bounded, measurable, compactly supported function. Suppose that the limit $M(f)$ exists. Then*

$$\mathop{\mathbb{E}}_{\Lambda\sim\mu}[f(\Lambda)] = \int_{\mathscr{L}_n} f(\Lambda)\,\mathrm{d}\mu(\Lambda) = M(f).$$

A similar averaging result holds for Construction-A lattices. See [68].

Step II. Equipped with the powerful Theorem 33, computation regarding expectations wrt Haar distribution can be turned into computation wrt the concrete ensemble defined in Eqn. (40). Rogers then gave a formula for the expectation of functions of a particular form by computing $M(\cdot)$. It can be shown that Eqn. (38) holds exactly true without the error term if we only sum over *linearly independent/full-rank* k-tuples.

Theorem 34 ([64], Theorem 2, Lemma 1 and Theorem 3)**.** *Let k and ρ be as in the setting of Theorem 32. Let*

$$f'(\Lambda) := \sum_{\substack{\underline{x}_1,\cdots,\underline{x}_k\in\Lambda\\ \mathrm{rk}\{\underline{x}_1,\cdots,\underline{x}_k\}=k}} \rho(\underline{x}_1,\cdots,\underline{x}_k).$$

Then

$$M(f') = \rho(0,\cdots,0) + \int_{\mathbb{R}^n}\cdots\int_{\mathbb{R}^n} \rho(\underline{x}_1,\cdots,\underline{x}_k)\,\mathrm{d}\underline{x}_1\cdots\mathrm{d}\underline{x}_k. \tag{41}$$

Step III. Rogers finally completed the proof of Theorem 32 by dropping the linear independence condition and lifting Theorem 34 from f' to

$$f(\Lambda) := \sum_{\underline{x}_1,\cdots,\underline{x}_k\in\Lambda} \rho(\underline{x}_1,\cdots,\underline{x}_k)$$

as promised in Theorem 32 at the cost of an extremely complicated error term $\mathcal{E}$.

B. Improvement on Rogers's formula

Although we have Rogers's higher-order averaging formula, it turns out that the error term $\mathcal{E}$ is very tricky to control even if we just plug in simple product functions. In the original paper by Rogers [69], [70], he was only able to show convergence of the first *few* moments of number of random lattice points in a *symmetric* set of *fixed* volume. Nevertheless, an intriguing Poisson behaviour was discovered and has been pushed to a greater generality in recent years.[20] We state below, as far as we know, the strongest results along this direction.

[20]Actually, Rogers showed that, asymptotically in the number of dimensions n, the first $\mathcal{O}(\sqrt{n})$ moments of the number of random lattices points in a set $\mathcal{S}$ which is centrally symmetric wrt the origin exhibit the same behaviour as a Poisson moment of the same degree with mean $V/2$, where $V := \mathrm{Vol}(\mathcal{S})$ is a constant independent of n. As we will see later, this is too weak for our purpose of list decoding. However, it is the earliest result which kicks off a fantastic adventure towards understanding the statistics of random lattices.

Let $Y \sim \mathrm{Pois}(V/2)$ be a Poisson random variable of mean $V/2$ for some V to be specified later.

Kim showed the following improvement upon Rogers results.

Theorem 35 (Proposition 3.3 of [71])**.** *Let $\mathcal{B}$ be a centrally symmetric set in $\mathbb{R}^n$ of volume V. There exists constants $C, c > 0$ such that, if n is sufficiently large and $V, k \leqslant Cn$, then*

$$\begin{aligned}\Pr[Y\geqslant k] - e^{-cn} &\leqslant \Pr\left[\frac{1}{2}\left|(\Lambda\backslash\{0\})\cap\mathcal{B}\right| \geqslant k\right]\\ &\leqslant \Pr[Y\geqslant k] + e^{-cn}.\end{aligned}$$

Note that the number of *pairs* of lattice points is considered since if $\underline{x}\in\Lambda$ then so is $-\underline{x}$. That is why there is a normalization factor $1/2$ in front of the number of nonzero lattice points in $\mathcal{B}$.

Strömbergsson and Södergren provided another improvement on the distribution of short vectors in a random lattice.

Theorem 36 (Theorem 1.2 of [65])**.** *Let $\mathcal{B}$ be an n-dimensional Euclidean ball centered at the origin of volume V. For any $\varepsilon > 0$,*

$$\Pr\left[\frac{1}{2}\left|(\Lambda\backslash\{0\})\cap\mathcal{B}\right| \leqslant k\right] - \Pr[Y\leqslant k] \overset{n\to\infty}{\to} 0,$$

uniformly wrt all $k, V \geqslant 0$ satisfying $\min\{k, V\} \leqslant \mathcal{O}_\varepsilon(e^{\varepsilon n})$.

We remark that though both results by Kim and Strömbergsson–Södergren are great extensions of Rogers's results to higher-order averaging formulas, they are not directly comparable. In Kim's Theorem 35, the set $\mathcal{B}$ can be any symmetric body, not necessarily convex. This is a good news since in list decoding we care about the number of lattice points in $\mathcal{B}(\underline{y}, r)$ for any possible received vector $\underline{y}\in\mathcal{B}(0,\sqrt{nP}+\sqrt{nN})\backslash\mathcal{B}(0,\sqrt{nP}-\sqrt{nN})$. Kim's result allows us to control that by taking $\mathcal{B} = \mathcal{B}(\underline{y}, r)\sqcup\mathcal{B}(-\underline{y}, r)$ (assuming $\mathcal{B}(\underline{y}, r)\cap\mathcal{B}(-\underline{y}, r) = \varnothing$). Obviously the configuration of lattice points are symmetric in $\mathcal{B}(\underline{y}, r)$ and $\mathcal{B}(-\underline{y}, r)$. Hence $|\Lambda\cap\mathcal{B}| = 2|\Lambda\cap\mathcal{B}(\underline{y}, r)|$. Also, Kim's result holds for $k = \mathcal{O}(n)$ which is also sufficient in our application, as we will see. Kim also quantified an exponential convergence rate. Unfortunately, his result requires V to be $\mathcal{O}(n)$, which is not enough for us. On the other hand, Strömbergsson–Södergren's result pushed the volume V to exponentially large in n but insists on $\mathcal{B}$ being a ball centered at the origin.

It should be intuitively clear that the Poissonianity behaviour of the moments will not hold for arbitrarily large degrees and for sets of arbitrarily large volume. The dimension that the lattice is living in is only n. If we compute the moments of very high degrees, we should expect to encounter some nontrivial correlation which makes the moments tricky to understand. Moreover, if we compute the moments of the number of lattice points in a very large set, it should not be surprising that at some point linearity of the lattices will kick in and dominate the behaviour of the moments.

X. List decodability of Haar lattices

Given the state of the art of bounds on moments of the number of Haar lattice points, we pose the following conjec-

ture and use it to show conditional results on list decodability of Haar lattices in the next section. The known properties of the Haar distribution that we have outlined previously should hopefully provide reasonable justification for why we believe that our conjectures are true.

Conjecture 37 (Poisson moment assumption)**.** *Let $\mathcal{B}$ be any symmetric set in $\mathbb{R}^n$. Then there exist constants $0 < c < 1$ and $C > 0$ large, such that if n is sufficiently large, $|\mathcal{B}| = V \leqslant 2^{Cn}$ and $0 \leqslant k \leqslant cn$, the following holds*

$$\begin{aligned}\mathop{\mathbb{E}}_{Y\sim\mathrm{Pois}(V/2)}\left[Y^k\right] - e_n &\leqslant \mathop{\mathbb{E}}_{\Lambda\sim\mu}\left[\left(\frac{|\Lambda\cap\mathcal{B}|}{2}\right)^k\right]\\ &\leqslant \mathop{\mathbb{E}}_{Y\sim\mathrm{Pois}(V/2)}\left[Y^k\right] + E_n,\end{aligned}$$

for some $e_n, E_n > 0$ such that $e_n, E_n \xrightarrow{n\to\infty} 0$. Recall that the k-th moment of a Poisson random variable (Fact 9) is

$$\mathop{\mathbb{E}}_{Y\sim\mathrm{Pois}(V/2)}\left[Y^k\right] = e^{-V/2}\sum_{i=0}^{\infty}\frac{i^k}{i!}(V/2)^i.$$

Note that results/conjectures phrased using tail bounds or moment bounds are essentially equivalent since one can be converted to another using the well-known relation between tails and moments. For any (continuous) random variable X with known tails, we can estimate its moment via

$$\mathbb{E}\left[|X|^k\right] = \int_0^\infty kt^{k-1}\Pr\left[|X| > t\right]\mathrm{d}t.$$

For any (continuous) random variable X with known moments, we can bound its tail via the Chernoff-type inequality,

$$\Pr\left[|X| > t\right] \leqslant \frac{\mathbb{E}\left[|X|^k\right]}{t^k}.$$

Previously, we showed that lattices and nested lattice codes can achieve $2^{\mathcal{O}(\frac{1}{\delta}\log^2\frac{1}{\delta})}$ list sizes whereas random spherical codes and periodic ICs achieve list sizes that grow as $\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})$. This leads to the natural question: Do there exist lattices/nested lattice codes that achieve $\mathcal{O}(\mathrm{poly}(1/\delta))$ list sizes? Are the exponential growth of the list sizes a consequence of structural regularity (i.e., linearity of the lattices) or is it an artifact of our proof? We conjecture that lattices can indeed achieve $\mathcal{O}(\mathrm{poly}(1/\delta))$ although we are unable to supply a complete proof at present. However, based on some heuristic assumptions, we can "prove" that a different ensemble of lattice codes (based on Haar lattices) achieve $\mathcal{O}(\mathrm{poly}(1/\delta))$ list sizes.

A. Conditional list decodability of Haar lattices

1) Codebook construction: Let $R = \frac{1}{2}\log\frac{P}{N} - \delta$ for some small constant $\delta > 0$. Sample a lattice Λ from the Haar distribution on $\mathscr{L}_n$. The lattice codebook is nothing but $\mathcal{C} := \alpha\Lambda\cap\mathcal{B}(0,\sqrt{nP})$ where α is defined below. Note that

$$|\mathcal{C}| = \left|\alpha\Lambda\cap\mathcal{B}(0,\sqrt{nP})\right| = \left|\Lambda\cap\alpha^{-1}\mathcal{B}(0,\sqrt{nP})\right|.$$

By Siegel's formula (Theorem 31), the expected number of codewords in the codebook is

$$\mathbb{E}\left[|\mathcal{C}|\right] = \frac{\mathrm{Vol}\left(\mathcal{B}^n\left(0,\sqrt{nP}\right)\right)}{\alpha^n} = \frac{\sqrt{nP}^n V_n}{\alpha^n}.$$

Setting this equal to 2^{nR}, we have

$$\alpha = \frac{\sqrt{nP}V_n^{1/n}}{2^R} \asymp \frac{\sqrt{2\pi eP}}{2^R}.$$

This coupled with the proceeding computation will provide the (conditional) *existence* of a $(P, N, \mathrm{poly}(1/\delta))$-list decodable lattice code.

2) Under distribution assumption: Heuristically and unrealistically, we first assume that the number of lattice points follows exactly a Poisson distribution, i.e., *every* moment of it is Poissonian.

Heuristic 38 (Poisson distribution assumption)**.** *Let Λ be a random lattice drawn from the Haar distribution on $\mathscr{L}_n$. If $\mathcal{B}$ is any centrally symmetric measurable set with nonempty interior, then $|\Lambda\cap\mathcal{B}|/2 \sim \mathrm{Pois}(\mathrm{Vol}(\mathcal{B})/2)$.*

This assumption is *not* believed to be true. As we mentioned before, at some point the linearity of the lattice should kick in and the moments are expected to diverge from Poissons as the order of the moments grows. Nevertheless, in this section we still conduct computation under this assumption that seems too good to be true. The result sets the bar for the "best" list decoding performance one can hope for, though it may never be reached in reality.

Another motivation for doing these calculations is that the same quantitative results under the the distributional assumption can be viewed as *rigorous* results for another code ensemble, that is, a Poisson point process (PPP) restricted to a ball. A homogeneous PPP has the property that the number of points in any compact set $\mathcal{B}$ with nonempty interior is distributed according to $\mathrm{Pois}(\mathrm{Vol}(\mathcal{B}))$. One subtle difference between this and the distribution assumption is that for lattices we need to normalize the number of lattice points in $\mathcal{B}$ by $1/2$. This is due to the linear structure of Λ – if $\underline{x}\in\Lambda$, then $-\underline{x}\in\Lambda$ with probability 1. Therefore, for any conjecture of this kind to make sense, the normalization factor $1/2$ is necessary.

Under the construction in Sec. X-A1, invoking Heuristic 38, we can get a high-probability guarantee on the size of the codebook. First note that

$$\mathrm{Vol}\left(\alpha^{-1}\mathcal{B}\left(0,\sqrt{nP}\right)\right) = \mathbb{E}\left[|\mathcal{C}|\right] = 2^{nR}.$$

By the Poisson tail bound (Lemma 12),

$$\begin{aligned}\Pr\left[\left|\frac{|\mathcal{C}|}{2} - 2^{nR}\right| \geqslant \frac{1}{2}2^{nR}\right] &\leqslant 2\exp\left(-\frac{(0.5\cdot 2^{nR})^2}{2\cdot(2^{nR}+0.5\cdot 2^{nR})}\right)\\ &= 2\exp\left(-\frac{2^{nR}}{24}\right).\end{aligned}$$

That is to say, with probability at least $1-e^{-\Omega(2^{nR})}$, $0.5\cdot 2^{nR} < |\mathcal{C}|/2 < 1.5\cdot 2^{nR}$, i.e., $2^{nR} < |\mathcal{C}| < 3\cdot 2^{nR}$. Therefore, the rate $R(\mathcal{C})$ of the code is $\frac{1}{2}\log\frac{P}{N} - \delta + o(1)$.

We then upper bound the following probability of failure of list decoding:

$$\Pr\left[\exists\underline{y}\in\mathcal{B}^n\left(0,\sqrt{nP}+\sqrt{nN}\right),\ \left|\alpha\Lambda\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right| > L\right]. \tag{42}$$

Take an optimal $\sqrt{n\varepsilon}$-covering $\mathcal{Y}$ of $\mathcal{B}^n\left(0,\sqrt{nP}+\sqrt{nN}\right)$. It can be achieved that

$$\begin{aligned}|\mathcal{Y}| &= \left(\frac{\mathrm{Vol}\left(\mathcal{B}^n\left(0,\sqrt{nP}+\sqrt{nN}+\sqrt{n\varepsilon}\right)\right)}{\mathrm{Vol}\left(\mathcal{B}^n\left(0,\sqrt{n\varepsilon}\right)\right)}\right)^{1+o(1)}\\ &= \left(\frac{\sqrt{P}+\sqrt{N}+\sqrt{\varepsilon}}{\sqrt{\varepsilon}}\right)^{(1+o(1))n}\\ &\leqslant \left(\frac{c_2}{\delta}\right)^n,\end{aligned}$$

where in the last step we set $\varepsilon := c_1\delta^2$. Then the probability (42) is upper bounded by

$$\begin{aligned}&\Pr\left[\exists \underline{y}\in\mathcal{Y},\ \left|\alpha\Lambda\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right)\right|>L\right]\\ &\quad\leqslant \sum_{\underline{y}\in\mathcal{Y}}\Pr\left[\left|\alpha\Lambda\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right)\right|>L\right]. \quad (43)\end{aligned}$$

Let

$$\begin{aligned}\mathcal{B}_1 &:= \frac{1}{\alpha}\mathcal{B}^n\left(\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right)\cup\frac{1}{\alpha}\mathcal{B}^n\left(-\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right),\\ \mathcal{B}_2 &:= \frac{1}{\alpha}\mathcal{B}^n\left(\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right)\cap\frac{1}{\alpha}\mathcal{B}^n\left(-\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right).\end{aligned}$$

Note that

$$\mathrm{Vol}(\mathcal{B}_1)+\mathrm{Vol}(\mathcal{B}_2)=2\,\mathrm{Vol}\left(\frac{1}{\alpha}\mathcal{B}^n\left(\sqrt{nN}+\sqrt{n\varepsilon}\right)\right). \quad (44)$$

By our assumption (Heuristic 38) in this section,

$$\begin{aligned}&\frac{1}{2}\left|\Lambda\cap\alpha^{-1}\mathcal{B}^n\left(\underline{y},\sqrt{nN}+\sqrt{n\varepsilon}\right)\right|\\ &\leqslant \frac{1}{2}\left|\Lambda\cap\mathcal{B}_1\right|\\ &\sim \mathrm{Pois}(\mathrm{Vol}(\mathcal{B}_1)/2)\\ &\leqslant \mathrm{Pois}(\mathrm{Vol}(\mathcal{B}_1)/2)+\mathrm{Pois}(\mathrm{Vol}(\mathcal{B}_2)/2)\\ &= \mathrm{Pois}\left(\frac{\mathrm{Vol}(\mathcal{B}_1)+\mathrm{Vol}(\mathcal{B}_2)}{2}\right) \quad (45)\\ &= \mathrm{Pois}\left(\mathrm{Vol}\left(\alpha^{-1}\mathcal{B}^n\left(\sqrt{nN}+\sqrt{n\varepsilon}\right)\right)\right). \quad (46)\end{aligned}$$

Eqn. (45) and (46) follow from Fact 10 and Eqn. (44), respectively. Plugging the parameters into the bound in Lemma 11, we can upper bound the probability in Eqn. (43) by

$$\begin{aligned}&\Pr\left[\left|\Lambda\cap\mathcal{B}^n\left(\alpha^{-1}\underline{y},\frac{\sqrt{nN}+\sqrt{n\varepsilon}}{\alpha}\right)\right|>L\right]\\ &\qquad<\frac{e^{-V}\left(eV\right)^{L/2}}{(L/2)^{L/2}}, \quad (47)\end{aligned}$$

where

$$\begin{aligned}V &:= \mathrm{Vol}\left(\mathcal{B}^n\left(\frac{\sqrt{nN}+\sqrt{n\varepsilon}}{\alpha}\right)\right)=\left(\frac{\sqrt{nN}+\sqrt{n\varepsilon}}{\alpha}\right)^n V_n\\ &\asymp 2^{n\left(R-\frac{1}{2}\log\frac{P}{N+2\sqrt{N\varepsilon}+\varepsilon}\right)}\approx 2^{-c_3 n\sqrt{\varepsilon}}<L. \quad (48)\end{aligned}$$

In the last step of the above chain of equalities, we set $c_3\approx 1/\sqrt{c_1}-(\log e)/\sqrt{N}$ and use that $R=\frac{1}{2}\log\frac{P}{N}-\delta$, $\varepsilon=c_1\delta^2$ and $\log(1+x)\approx(\log e)x$. Hence the RHS of the tail (47) is

$$\begin{aligned}&\exp\left(-2^{-c_3 n\sqrt{\varepsilon}}\right)\left(e2^{-c_3 n\sqrt{\varepsilon}}\right)^{L/2}/(L/2)^{L/2}\\ &\qquad\asymp\left(\frac{e}{L/2}\right)^{L/2}2^{-c_3 n\sqrt{\varepsilon}L/2}. \quad (49)\end{aligned}$$

Taking a union bound over $\mathcal{Y}$, the overall probability of failure of list decoding (Eqn. (42)) is at most

$$\begin{aligned}&\left(\frac{e}{L/2}\right)^{L/2}2^{-c_3 n\sqrt{\varepsilon}L/2}\left(\frac{c_2}{\delta}\right)^n\\ &\qquad=\left(\frac{e}{L/2}\right)^{L/2}2^{-n\left(c_3\sqrt{c_1}\delta L/2-\log\frac{c_2}{\delta}\right)}.\end{aligned}$$

The multiplicative factor $\left(\frac{e}{L/2}\right)^{L/2}$ is going to be negligible once n is sent to infinity. The exponent is negative if we set L to be $c'\frac{1}{\delta}\log\frac{1}{\delta}$ for some appropriate constant c'.

The above calculations indicate that, under the Poisson distributional assumption of the number of lattice points in a set, a random lattice (appropriately scaled) drawn from the Haar measure performs as well as uniformly random spherical codes. We therefore have the following result:

Lemma 39. *If Heuristic 38 is true, then there exists a lattice Λ such that $\Lambda\cap\mathcal{B}(0,\sqrt{nP})$ has rate $C(P,N)-\delta$ and is $\left(P,N,\mathcal{O}(\frac{1}{\delta}\log\frac{1}{\delta})\right)$-list decodable.*

3) Under moment assumption: Now instead of assuming that the number of lattice points in any symmetric body has Poisson distribution, we only assume that its first k moments match Poisson moments.

First note that the rate of the code is still well concentrated:

$$\begin{aligned}\Pr\left[\left|\frac{|\mathcal{C}|}{2}-2^{nR}\right|\geqslant 2^{n(R+\delta/2)}\right] &\leqslant \frac{\mathrm{Var}\left[|\mathcal{C}|/2\right]}{\left(2^{n(R+\delta/2)}\right)^2}\\ &\leqslant \frac{2^{nR}}{2^{n(2R+\delta)}}\\ &= 2^{-n(R+\delta)}.\end{aligned}$$

The second inequality follows since the first and second moments of $|\mathcal{C}|/2$ are the same as those of $\mathrm{Pois}(2^{nR})$. Therefore $R(\mathcal{C})=R+\delta/2+o(1)=\frac{1}{2}\log\frac{P}{N}-\delta/2+o(1)$.

Let

$$X=\frac{1}{2}\left|\Lambda\cap\mathcal{B}^n\left(\frac{\sqrt{nN}+\sqrt{n\varepsilon}}{\alpha}\right)\right|. \quad (50)$$

By Conjecture 37 and Fact 9, for any $0\leqslant m\leqslant k$,

$$\mathbb{E}\left[X^m\right]\asymp\mathbb{E}\left[Y^m\right]=e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j j^m}{j!},$$

where $Y\sim\mathrm{Pois}(\lambda)$, $\lambda=V/2$ and V is given by formula (48). Indeed,

$$\lambda\approx\frac{1}{2}\cdot 2^{-c_3 n\sqrt{\varepsilon}}\overset{n\to\infty}{\to}0. \quad (51)$$

Then the probability in Eqn. (43) can be upper bounded by

$$\Pr\left[X^k>(L/2)^k\right]<\mathbb{E}\left[X^k\right]/(L/2)^k=\mathbb{E}\left[X^k\right]e^{-k\ln(L/2)}.$$

Let $k=:cn$ where $0<c<1$ is a constant. If $\mathbb{E}\left[X^k\right]\leqslant e^{-nD}$ for some $D>0$, then after taking a union bound over $\underline{y}\in\mathcal{Y}$, we are in good shape if

$$e^{-n\left(D+c\ln\frac{L}{2}-\ln\frac{c_2}{\delta}\right)}\overset{n\to\infty}{\to}0. \quad (52)$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542

Now let us compute the k-th moment.

$$\begin{aligned}\mathbb{E}\left[X^k\right] =& e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j j^k}{j!}\\ \asymp& \sum_{j\geqslant 0}\frac{\lambda^j j^k}{\sqrt{2\pi j}\,(j/e)^j} \qquad (53)\\ =& \sum_{j\geqslant 0}\exp\left(j\ln\lambda + k\ln j - j\ln j + j - \frac{1}{2}\ln(2\pi j)\right),\end{aligned}$$

where in Eqn. (53) we use Stirling's approximation (Lemma 13). As we know, a sum of exponentials is dominated by the largest term. Let us compute the largest one. Define function

$$f(j) \coloneqq -j\ln j + (\ln\lambda + 1)\, j + (k-1/2)\ln j - \frac{1}{2}\ln(2\pi).$$

Its first derivative is given by

$$\frac{\mathrm{d}f}{\mathrm{d}j} = \ln\lambda + \frac{k-1/2}{j} - \ln j.$$

Setting it equal to zero and solving the equation, we get the critical point

$$j^* \coloneqq \frac{k-1/2}{W\left(\frac{k-1/2}{\lambda}\right)},$$

where $W(\cdot)$ is the *Lambert W function* which is the inverse of $g(x) = xe^x$. The function $W(\cdot)$ satisfies the following estimate for sufficiently large x,

$$W(x) = \ln x - \ln\ln x + o(1).$$

Note that, by Eqn. (51),

$$\frac{k-1/2}{\lambda} = (cn-1/2)\cdot 2\cdot 2^{c_3 n\sqrt{\varepsilon}} \overset{n\to\infty}{\to} \infty.$$

Hence

$$\begin{aligned}W\left(\frac{k-1/2}{\lambda}\right) \asymp& \ln\left((cn-1/2)\cdot 2\right) + c_3 n\sqrt{\varepsilon}\ln 2\\ &+ \ln\left(\ln\left((cn-1/2)\cdot 2\right) + c_3 n\sqrt{\varepsilon}\ln 2\right)\\ =& \ln 2\cdot c_3\sqrt{\varepsilon}\cdot n(1+o(1)).\end{aligned}$$

We thus have

$$j^* \asymp \frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}. \qquad (54)$$

Furthermore,

$$\frac{\mathrm{d}^2 f}{\mathrm{d}j^2} = -\frac{1}{j} - \frac{k-1/2}{j^2} < 0$$

since $k = cn > 1/2$ (where $c > 0$ is a constant) for sufficiently large n. Therefore, f is concave and attains its maximum at j^*. Plug j^* (Eqn. (54)) into f,

$$\begin{aligned}f(j^*) =& -j^*\ln j^* + (\ln\lambda+1)\, j^* + (cn-1/2)\ln j^*\\ &- \frac{1}{2}\ln(2\pi)\\ =& -\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}\ln\left(\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}\right)\\ &+ \left(-\ln 2\cdot c_3\sqrt{\varepsilon}\cdot n + 1\right)\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}\\ &+ (cn-1/2)\ln\left(\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}\right) - \frac{1}{2}\ln(2\pi)\\ =& -n(1+o(1))\left(c - c\ln\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}}\right).\end{aligned}$$

Finally, the exponent of expression (52) is

$$\begin{aligned}&D + c\ln\frac{L}{2} - \ln\frac{c_2}{\delta}\\ &\approx c - c\ln\frac{c}{\ln 2\cdot c_3\sqrt{\varepsilon}} + c\ln\frac{L}{2} - \ln\frac{c_2}{\delta}\\ &= c\ln L - (c+1)\ln\frac{1}{\delta} + c - c\ln\frac{2c}{\ln 2\cdot c_3\sqrt{c_1}} - \ln c_2.\end{aligned}$$

In order for it to be positive as $\delta\to 0$, we had better set $L = (1/\delta)^a$, where $ac > c+1$, i.e., $a > 1 + 1/c$. If we only assume the first $k = cn < n$ moments are Poissonian for some $c = \frac{1}{1+\gamma} < 1$ where $\gamma > 0$ is a some small positive constant, then we need to take $a > 2+\gamma$.

Therefore, we have proved the following lemma.

Lemma 40. *If Conjecture 37 is true, then there exists a lattice Λ such that $\Lambda\cap\mathcal{B}(0,\sqrt{nP})$ has rate $C(P,N)-\delta$ and is $\left(P,N,\mathcal{O}(1/\delta^{1+1/c})\right)$-list decodable where $0<c<1$ is given in Conjecture 37.*

B. Remark

Careful readers might have observed that in order to prove Lemma 40, we do not really need the first cn moments to be Poisson. It suffices to show that the first cn moments are bounded from above by a quantity that is subexponential in n. However, we are optimistic that a result similar to Conjecture 37 can indeed be proved for the Haar distribution on $\mathscr{L}_n$.

XI. Concluding Remarks and Open Problems

In this paper we initiate a systematic study of the list size problem for codes over $\mathbb{R}$. In particular, upper bounds on list sizes of nested Construction-A lattice codes and infinite Construction-A lattices are exhibited. Similar upper bounds are also obtained for an ensemble of regular infinite constellations. Matching lower bounds for such an ensemble are provided. Other coding-theoretic properties are studied by the way. Our lower bound for random spherical codes also matches the upper bound in previous work. A caveat is that all of our bounds are concerned with *typical* scaling of the list sizes of *random* codes sampled from the ensembles of interest. The extremal list sizes *may* be smaller than our lower bounds. We conclude the paper with several open questions.

1) Careful readers might have already noted that a missing piece in this work is a list size lower bound for random Construction-A lattice codes. We had trouble replicating the arguments in [19]. We leave it as an open question to get a $\mathrm{poly}(1/\delta)$ list size lower bound.
2) Can one sample efficiently from the Haar distribution on the spaces of our interest? In particular, can one sample efficiently a generator matrix $\mathbf{G}$ from the Haar distribution μ on $\mathrm{SL}(n,\mathbb{R})$? Can one sample eficiently a lattice Λ from the Haar distribution μ on $\mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$? To

this end, we can think of $\mathrm{SL}(n,\mathbb{R})$ as a codimensional-one hypersurface in $\mathbb{R}^{n^2}$ cut off by the equation $\det(\mathbf{G}) = 1$. Readers from the Monte Carlo Markov Chain (MCMC) community may be interested in such problems.

3) A very intriguing question which we are unable to resolve in this work is to bring down the exponential list size of random Construction-A lattice codes. We do not believe that our upper bound is tight. A starting step towards this goal is probably to obtain an averaging formula custom tailored for Construction-A lattices. Indeed, Loeliger [68] has proved a first-order averaging formula for (appropriately scaled) Construction-A lattices as an analog of Siegel's formula for Haar lattices. Specifically, consider an ensemble of Construction-A-type lattices $\Lambda := \frac{1}{\alpha}(\mathcal{C} + q\mathbb{Z}^n)$ where $\mathcal{C} \sim \mathbf{Gr}(\kappa, \mathbb{F}_q^n)$ is a uniformly random κ-dimensional subspace of $\mathbb{F}_q^n$. Then for any bounded measurable compactly supported function $\rho\colon \mathbb{R}^n \to \mathbb{R}$, it holds that

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathcal{C}\sim\mathscr{C}_{n,\kappa}}\left[\sum_{\underline{x}\in\Lambda\setminus\{0\}}\rho(\underline{x})\right] &= \frac{1}{|\mathscr{C}_{n,\kappa}|}\sum_{\mathcal{C}\in\mathscr{C}_{n,\kappa}}\sum_{\underline{x}\in\Lambda\setminus\{0\}}\rho(\underline{x}) \\ &\xrightarrow{\alpha\to\infty, q/\alpha\to\infty} \det(\Lambda)^{-1}\int_{\mathbb{R}^n}\rho(\underline{x})\mathrm{d}\underline{x},\end{aligned}$$

where $\mathscr{C}_{n,\kappa} := \mathbf{Gr}(\kappa, \mathbb{F}_q^n)$ and the covolume

$$\begin{aligned}\det(\Lambda) &= \left(\frac{1}{\alpha}\right)^n [\mathbb{Z}^n\colon (\mathcal{C} + q\mathbb{Z}^n)] \\ &= |\mathbb{Z}^n/(\mathcal{C} + q\mathbb{Z}^n)|\,/\alpha^n = q^{n-\kappa}/\alpha^n,\end{aligned}$$

is kept fixed. Can one lift Loeliger's formula to k-variate functions $\rho\colon (\mathbb{R}^n)^k \to \mathbb{R}$ and get a higher-order averaging formula for Construction-A lattices as an analog of Rogers's formula for Haar lattices?

4) Can one compute similar moments for random Construction-A lattices? Given a random Construction-A lattice $\Lambda = q^{-1}\mathcal{C} + \mathbb{Z}^n$ where $\mathcal{C}$ is a κ-dimensional random linear code in $\mathbb{F}_q^n$, compute the k-th moment $\mathbb{E}\left[\left|\Lambda \cap \mathcal{B}_2^n(\underline{y}, \sqrt{nN})\right|^k\right]$ for any $\underline{y} \in \mathbb{R}^n$ and for k as large as possible. For *random binary linear code* over $\mathbb{F}_2^n$ of rate $1 - H(p) + \delta$,[21] Linial and Mosheiff [72] recently managed to *characterize* the first $\mathcal{O}(n/\log n)$ moments of the number of codewords in a Hamming *sphere* of radius np. It turns out that the normalized centered moment

$$\frac{\mathbb{E}\left[\left(|\mathcal{C}\cap\mathcal{S}_{\mathrm{H}}(0,np)| - \mathbb{E}\left[|\mathcal{C}\cap\mathcal{S}_{\mathrm{H}}(0,np)|\right]\right)^k\right]}{\mathrm{Var}\left[|\mathcal{C}\cap\mathcal{S}_{\mathrm{H}}(0,np)|\right]^{k/2}}$$

behaves like the moment of a Gaussian (recall Fact 8) up to some threshold $k < k_0$, where k_0 is 3 or 4 for δ not too small. From k_0 on, linearity quickly kicks in and dominates the behaviour of the moments.

5) We showed that Haar lattices of rate $\frac{1}{2}\log\frac{P}{N} - \delta$ are $(P, N, \mathrm{poly}(1/\delta))$-list decodable whp conditioned on Conjecture 37. Can one show other coding-theoretic goodness properties under the conjecture? It is known that Haar lattices are good for packing [47]. Are they also good for covering, AWGN, quantization, etc.?
6) In this paper, the list decodability of two ensembles (Construction-A and Haar) of lattices are considered. The ultimate goal is to find an *explicit* $(P, N, \mathrm{poly}(1/\delta))$-list decodable lattice code of rate $\frac{1}{2}\log\frac{P}{N} - \delta$. Recently Kaufman and Mass [73] constructed explicit lattices of good *distance* from high dimensional expanders. However, there is no explicit bound on the covolume of the lattice. Therefore, it is unclear whether their construction is good for packing. Moreover, their results are conditioned on the conjecture that the cohomology group of Ramanujan complexes with *integer* coefficients is large.
7) Our lower bounds on list sizes only indicate typical behaviours of ensembles of random lattices. This does not exclude the existence of codes with smaller list sizes. Can one prove a lower bound on list sizes of *general* codes over reals? Namely, for any 2^{nR} points on $\mathcal{S}^{n-1}\left(0, \sqrt{nP}\right)$, how large can L be such that one can always find a position $\underline{y}$ to which there are at least L points that are $\sqrt{nN}$-close?

[21] Note that such a code operates at a rate *above* capacity and the corresponding moments they are interested in are exponentially large. Indeed, they instead consider *centered* moments $\mathbb{E}\left[(X - \mathbb{E}[X])^k\right]$.

XII. Acknowledgement

YZ thanks Noah Stephens-Davidowitz for sharing his expertise on lattices, in particular, introducing him the Poisson heuristics as a prediction of the behaviors of random lattices when he was visiting MIT and for exchanging multiple informative emails afterwards. YZ thanks Mary Wootters for clarifying the state of the art of list decodability of random linear codes. YZ also wants to thank Boris Bukh, Chris Cox, Sidharth Jaggi, Nicolas Resch and Tomasz Tkocz for several inspiring discussions, respectively. Part of this work was done when YZ was visiting CMU under the mentorship of Venkatesan Guruswami who listened to the progress and provided generous encouragement at the early stage. Part of this work was done when SV was a postdoc at CUHK.

APPENDIX A
TABLE OF NOTATION

Symbol	Section	Description	Definition/Value/Range
A_{n-1}	Throughout the paper	Area of an $(n-1)$-dimensioinal unit sphere	$A_{n-1} := \mathrm{Area}(\mathcal{S}_2^{n-1})$
$\mathcal{A}$	Sec. VI, VII	Cube of side length α	$\mathcal{A} := [-\alpha/2, \alpha/2)^n$
C	Sec. II	List decoding capacity	$C := 1 - H_q(p) \in [0,1]$
		List decoding capacity	$C := 1 - p \in [0,1]$
	Throughout the paper	List decoding capacity	$C := \frac{1}{2}\log\frac{P}{N} \in \mathbb{R}_{\geqslant 0}$
$\mathcal{C}$	Sec. II	Code	$\mathcal{C} \in \binom{\mathbb{F}_q^n}{q^{nR}}$
	Throughout the paper	Code	$\mathcal{C} \subset \mathbb{R}^n$ of size 2^{nR}
	Sec. I-A, VI, VII	IC	$\mathcal{C} \subset \mathbb{R}^n$
$C(L)$	Sec. II	List-L capacity	See Eqn. (1)
$C_{\mathrm{rand}}(\mathcal{W})$	Sec. II	Random code capacity	$C_{\mathrm{rand}}(\mathcal{W}) := \max_P \min_{P_{\mathbf{xsy}}=P_{\mathbf{x}}P_{\mathbf{s}}W_{\mathbf{y}\|\mathbf{xs}}:\ P_{\mathbf{x}}=P} I(\mathbf{x};\mathbf{y})$
$\mathbf{G}$	Throughout the paper	Generator matrix of a linear code	$\mathbf{G} \in \mathbb{F}_q^{n\times\kappa}$
		Generator matrix of a lattice	$\mathbf{G} \in \mathbb{R}^{n\times\kappa}$
k	Sec. IX, X	Degree of moments	$k = cn$
	Sec. XI	Degree of moments	$k = \mathcal{O}(n/\log n)$
ℓ	Sec. V	Log of list size	$\ell := \log_q(L+1)$
L	Throughout the paper	List size	$L \in [q^{nR}]$
$\mathcal{L}$	Throughout the paper	List	$\mathcal{L} \in \binom{\mathcal{C}}{\leqslant L}$
$\mathscr{L}_n$	Sec. IX, X	Space of determinant-1 lattices	$\mathscr{L}_n := \{\Lambda \leqslant \mathbb{R}^n \text{ lattice}:\ \det(\Lambda) = 1\}$
m	Sec. IV, V	Message	$m \in [q^{nR}]$
M	Sec. IV, V	Number of messages/size of codebook	$M := \|\mathcal{M}\| = q^{nR}$
	Sec. II	Symmetrizability	See Eqn. (2)
$\mathcal{M}$	Sec. IV, V	Set of messages	$\mathcal{M} := \{0, 1, \cdots, 2^{nR}-1\}$
n	Throughout the paper	Blocklength	$n \in \mathbb{Z}_{>0}$
N	Throughout the paper	Adversary's power constraint	$N \in \mathbb{R}_{>0}$
p	Sec. II	Adversary's power constraint	$p \in [0,1]$
P	Throughout the paper	Transmit power constraint	$P \in \mathbb{R}_{>0}$
$\mathcal{P}(\Lambda)$	Sec. V	Fundamental parallelepiped	$\mathcal{P}(\Lambda) := \{\mathbf{G}\underline{x} : \underline{x} \in [0,1)^n\}$
q	Throughout the paper	Characteristic of finite field	Prime number
$Q_\Lambda(\cdot)$	Sec. V	Lattice quantizer	See Eqn. (14)
r_{cov}	Sec. V	Covering radius of a lattice	See Appendix V-B
r_{eff}	Sec. V	Effective radius of a lattice	See Appendix V-B
	Sec. I-A, VI, VII	Effective radius of an infinite constellation	See Def. 3
r_{pack}	Sec. V	Packing radius of a lattice	See Appendix V-B
R	Throughout the paper	Rate of a code	$\frac{\log\|\mathcal{C}\|}{n} \in \mathbb{R}_{>0}$
$\underline{s}$	Throughout the paper	Jamming vector	$\underline{s} \in \mathcal{B}(0, \sqrt{nN})$
$V(\mathcal{C})$	Sec. I-A, VI, VII	Effective volume of an IC	$V(\mathcal{C}) = 1/\Delta(\mathcal{C})$
V_n	Throughout the paper	Volume of an n-dimensioinal unit ball	$V_n := \mathrm{Vol}(\mathcal{B}_2^n)$
$\mathcal{V}(\Lambda)$	Sec. V, VI	Fundamental Voronoi region	$\mathcal{V}(\Lambda) := \{\underline{x} \in \mathbb{R}^n : Q_\Lambda(\underline{x}) = 0\}$
W	Sec. IV, VII	Witness of list decodability	See Eqn. (8), (34)
$W(\cdot\|\cdot,\cdot)$	Sec. II	Transition probability of an AVC	$W\colon \mathcal{Y}\times\mathcal{X}\times\mathcal{S} \to [0,1]$
$\mathcal{W}$	Sec. II	AVC	$\mathcal{W} := \{W(\cdot\|\cdot, s), \underline{s} \in \mathcal{S}\}$
$\underline{x}$	Throughout the paper	Transmitted codeword	$\underline{\mathbf{x}} \in \mathcal{C}$
X	Sec. X	Number of lattice points in a ball	See Eqn. (50)
$\underline{y}$	Throughout the paper	Received word	$\underline{y} = \underline{x} + \underline{s} \in \mathcal{B}(0, \sqrt{nP}+\sqrt{nN}) \backslash \mathcal{B}(0, \sqrt{nP}-\sqrt{nN})$

Y	Sec. IX, X	Poisson random variable	$Y \sim \mathrm{Pois}(V/2)$
$\mathcal{Y}$	Throughout the paper	Net for $\underline{y}$'s	See specific definitions
$\underline{\mathbf{z}}$	Sec. VII	AWGN	$\mathbb{R}^n \ni \underline{\mathbf{z}} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$
α	Throughout the paper	Side length of $\mathcal{A}$	$\alpha \in \mathbb{R}_{>0}$
δ	Throughout the paper	Gap to capacity	$\delta := C - R \in \mathbb{R}_{>0}$
	Sec. I-A, VI, VII	Gap between $r_{\mathrm{eff}}(\mathcal{C})$ and $\sqrt{nN}$	$\delta := \log \frac{r_{\mathrm{eff}}(\mathcal{C})}{\sqrt{nN}}$
$\Delta(\mathcal{C})$	Sec. I-A, VI, VII	Density of an IC	$\Delta(\mathcal{C}) := \limsup_{a\to\infty} \frac{\lvert\mathcal{C}\cap[0,a]^n\rvert}{a^n}$
ε	Throughout the paper	Parameter of a net	See specific definitions
$\mathbf{\Theta}$	Sec. IX	Rogers's ensemble	$\mathbf{\Theta} = \mathbf{\Theta}(\theta_1, \cdots, \theta_{n-1}, \omega)$ (Eqn. (40))
κ	Throughout the paper	Dimension of a linear code or a lattice	$\kappa \in \{0, 1, \cdots, n\}$
Λ	Throughout the paper	Lattice	$\Lambda \leqslant \mathbb{R}^n$
μ	Sec. IX, X	Haar measure on $\mathrm{SL}(n,\mathbb{R})$, $\mathscr{L}_n$ or $\mathrm{GL}(n,\mathbb{R})$	See Theorem 30
τ	Sec. II	Gap to list decoding radius	$\tau := 1 - 1/q - p \in \mathbb{R}_{>0}$
Φ	Sec. V	Natural embedding	$\Phi\colon \mathbb{F}_q \to \mathbb{Z}$
ψ	Sec. IV, V, VII	Encoding function	$\psi\colon \mathcal{M} \to \mathcal{C}$
$[\cdot] \bmod \mathcal{A}$	Sec. VI, VII	Quantization error wrt $\alpha\mathbb{Z}^n$	$[\cdot] \bmod \mathcal{A} := \cdot \bmod \alpha\mathbb{Z}^n$
$[\cdot] \bmod \Lambda$	Sec. V	Lattice quantization error	$[\cdot] \bmod \Lambda := \cdot - Q_\Lambda(\cdot)$
$(\cdot)^*$	Sec. VI, VII	Set modulo $\alpha\mathbb{Z}^n$	$(\cdot)^* := \cdot \bmod \mathcal{A}$

APPENDIX B
PROOF OF EQN. (10) AND (11)

A. A covering lemma

Before proving Eqn. (10) and (11), we need the following lemma. It guarantees the existence of a covering of a sphere which is sufficiently spread out in the sense that the fraction of points in any spherical cap does not deviate much from the corresponding volume ratio.

Lemma 41. *Let $r > 0$ and $\varepsilon > 0$ sufficiently small. There exists a subset $\mathcal{Y}$ of the sphere $\mathcal{S}^{n-1}(0, \sqrt{nr})$ such that*

1) *for every $\underline{y} \in \mathcal{S}^{n-1}(0, \sqrt{nr})$, there exists $\underline{y}' \in \mathcal{Y}$ with $\|\underline{y} - \underline{y}'\| \leqslant \sqrt{n\varepsilon}$;*
2) *$|\mathcal{Y}| = (c/\sqrt{\varepsilon})^{n(1+o(1))}$ for some $c > 0$ that is independent of n and ε but depends on r;*
3) *for every $\underline{y} \in \mathcal{S}^{n-1}(0, \sqrt{nr})$ and every $0 < \rho < r$,*

$$\frac{\left|\mathcal{Y} \cap \mathcal{C}ap^{n-1}(\underline{y}, \sqrt{n\rho})\right|}{|\mathcal{Y}|} \geqslant \frac{1}{2} \frac{\text{Area}\left(\mathcal{C}ap^{n-1}(\underline{z}, \sqrt{n}(\sqrt{\rho} - \sqrt{\varepsilon_\ell}))\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)},$$
$$\frac{\left|\mathcal{Y} \cap \mathcal{C}ap^{n-1}(\underline{y}, \sqrt{n\rho})\right|}{|\mathcal{Y}|} \leqslant \frac{3}{2} \frac{\text{Area}\left(\mathcal{C}ap^{n-1}(\underline{z}, \sqrt{n}(\sqrt{\rho} + \sqrt{\varepsilon_u}))\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)},$$

where

$$\varepsilon_\ell := \left(\frac{1}{4\sqrt{r}} + \frac{3}{2}\right)^2 \varepsilon, \quad \varepsilon_u := \frac{9}{4}\varepsilon.$$

Proof. Let $\mathcal{Y}$ be a set of $M = (1 + o(1))\sqrt{2\pi n}\left(4\sqrt{\frac{r}{\varepsilon}}\right)^{n-1}$ points $\underline{\mathbf{y}}_1, \cdots, \underline{\mathbf{y}}_M$ each independent and uniformly distributed on $\mathcal{S}^{n-1}(0, \sqrt{nr})$. Note that $M = (c/\sqrt{\varepsilon})^{n+o(n)}$ for some c independent of n and ε (but dependent on r), which satisfies property 2. We will show that such a $\mathcal{Y}$ satisfies all properties in Lemma 41 with high probability.

By a standard volume argument, there exists a $\sqrt{n\varepsilon_1}$-net $\mathcal{Z}$ of $\mathcal{S}^{n-1}(0, \sqrt{n\varepsilon_1})$ satisfying properties 1 and 2 with ε replaced with $\varepsilon_1 = \varepsilon/4$ (and the constant c needs to be adjusted accordingly).

$$\begin{aligned}
&\Pr\left[\exists \underline{y} \in \mathcal{S}^{n-1}(0, \sqrt{nr}),\ \forall i \in [M],\ \|\underline{y} - \underline{\mathbf{y}}_i\| > \sqrt{n\varepsilon}\right] \\
&\leqslant \Pr\left[\exists \underline{z} \in \mathcal{Z},\ \forall i \in [M],\ \|\underline{z} - \underline{\mathbf{y}}_i\| > \sqrt{n\varepsilon} - \sqrt{n\varepsilon_1}\right] \\
&\leqslant \sum_{\underline{z} \in \mathcal{Z}} \prod_{i \in [M]} \Pr\left[\|\underline{z} - \underline{\mathbf{y}}_i\| > \sqrt{n\varepsilon/4}\right] \\
&\leqslant \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} \left(1 - \frac{\text{Area}\left(\mathcal{C}ap^{n-1}(\sqrt{n\varepsilon/8})\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)}\right)^M \qquad (55) \\
&\leqslant \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} \left(1 - \frac{\text{Vol}\left(\mathcal{B}^{n-1}(\sqrt{n\varepsilon/8})\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)}\right)^M \qquad (56) \\
&= \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} \left(1 - \frac{V_{n-1}}{A_{n-1}}\sqrt{\frac{\varepsilon}{8r}}^{n-1}\right)^M \\
&= \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} \left(1 - \frac{1 + o(1)}{\sqrt{2\pi n}}\sqrt{\frac{\varepsilon}{8r}}^{n-1}\right)^M \qquad (57) \\
&= \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} \times \\
&\quad \left(1 - \frac{1 + o(1)}{\sqrt{2\pi n}}\sqrt{\frac{\varepsilon}{8r}}^{n-1}\right)^{(1+o(1))\sqrt{2\pi n}\sqrt{\frac{8r}{\varepsilon}}^{n-1}\sqrt{2}^{n-1}} \qquad (58) \\
&\leqslant \left(\frac{c}{\sqrt{\varepsilon_1}}\right)^{n+o(n)} e^{-\sqrt{2}^{n-1}}. \qquad (59)
\end{aligned}$$

Eqn. (55) follows since the set $\left\{\underline{y} \in \mathcal{S}^{n-1}(\sqrt{nr}) : \|\underline{z} - \underline{y}\| \leqslant \sqrt{n\varepsilon/4}\right\}$ forms a cap of radius $\sqrt{n\varepsilon'}$ where ε' can be determined by inspecting the geometry. Specifically, $\sqrt{\varepsilon'} = \sqrt{r}\sin\theta$, where θ satisfies

$$\cos\theta = \frac{r + r - \varepsilon/4}{2r} = 1 - \frac{\varepsilon}{8r}.$$

Therefore,

$$\varepsilon' = \sqrt{r}\sqrt{1 - \left(1 - \frac{\varepsilon}{8r}\right)^2} \geqslant \sqrt{r}\sqrt{\frac{\varepsilon}{8r}} = \sqrt{\varepsilon/8},$$

where the inequality follows since $1 - (1 - x)^2 \geqslant x$ for any $0 \leqslant x \leqslant 1$. Eqn. (56) follows since $\text{Area}\left(\mathcal{C}ap^{n-1}(\gamma)\right) \geqslant \text{Vol}\left(\mathcal{B}^{n-1}(\gamma)\right)$ for any $\gamma > 0$. In Eqn. (57), the ratio V_{n-1}/A_{n-1} is given by

$$\begin{aligned}
\frac{V_{n-1}}{A_{n-1}} &= \frac{\frac{1}{\sqrt{\pi(n-1)}}\left(\frac{2\pi e}{n-1}\right)^{\frac{n-1}{2}}}{\sqrt{\frac{n}{\pi}}\left(\frac{2\pi e}{n}\right)^{\frac{n}{2}}}(1 + o(1)) \\
&= \frac{1}{\sqrt{2\pi e n}}\left(\frac{n}{n-1}\right)^{n/2}(1 + o(1)) \\
&\to \frac{1}{\sqrt{2\pi n}}(1 + o(1)).
\end{aligned}$$

Eqn. (58) is by the choice of M. Eqn. (59) follows from the inequality $(1 - 1/x)^x \leqslant 1/e$ for $x \geqslant 1$. Therefore, property 1 holds with probability at least $1 - e^{-e^{\Omega(n)}}$.

To show property 3, we quantize the interval $[0, r]$ using an ε_2-net $\{0, \varepsilon_2, 2\varepsilon_2, \cdots, \lfloor r/\varepsilon_2 \rfloor \varepsilon_2\}$. Such a net satisfies that for any $\rho \in [0, r]$, there exists $\tau \in \mathcal{I}$ with $|\tau - \rho| \leqslant \varepsilon_2$. Let

$$\gamma_\ell := \frac{\text{Area}\left(\mathcal{C}ap^{n-1}(\sqrt{n}(\sqrt{\rho} - \sqrt{\varepsilon_\ell}))\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)},$$
$$\gamma_u := \frac{\text{Area}\left(\mathcal{C}ap^{n-1}(\sqrt{n}(\sqrt{\rho} + \sqrt{\varepsilon_u}))\right)}{\text{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)}.$$

We then bound the probability that property 3 is violated.

$$\begin{aligned}
&\Pr\Bigg[\exists \underline{y} \in \mathcal{S}^{n-1}(0, \sqrt{nr}),\ \exists \rho \in [0, r], \\
&\qquad \frac{\left|\mathcal{Y} \cap \mathcal{C}ap^{n-1}(\underline{y}, \sqrt{n\rho})\right|}{M} \notin \left[\frac{1}{2}\gamma_\ell, \frac{3}{2}\gamma_u\right]\Bigg] \qquad (60) \\
&\leqslant \Pr\Bigg[\exists \underline{z} \in \mathcal{Z},\ \exists \tau \in \mathcal{I},\ \|\underline{z} - \underline{y}\| \leqslant \sqrt{n\varepsilon_1},\ |\tau - \rho| \leqslant \varepsilon_2,
\end{aligned}$$

$$\left.\left.\frac{\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau})\right|}{M}\notin\left[\frac{1}{2}\gamma_\ell,\frac{3}{2}\gamma_u\right]\right]$$
$$\leqslant \sum_{\underline{z}\in\mathcal{Z}}\sum_{\tau\in\mathcal{I}}\left(\Pr\left[\frac{\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau_\ell})\right|}{M}<\frac{1}{2}\gamma_\ell\right]\right.$$
$$\left.+\Pr\left[\frac{\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau_u})\right|}{M}>\frac{3}{2}\gamma_u\right]\right). \quad (61)$$

We choose radii τ_ℓ and τ_u such that

$$\mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau_\ell})\subset \mathcal{C}ap^{n-1}(\underline{y},\sqrt{n\rho})\subset \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau_u}).$$

In what follows, we derive lower and upper bounds on τ_ℓ and τ_u, respectively. The geometry is depicted in Fig. 3.

For notational convenience, let $\rho_\ell := \rho-\varepsilon_2$ and $\rho_u := \rho+\varepsilon_2$. Let $\alpha_\ell,\alpha_u,\beta$ be defined as:

$$\sin\alpha_\ell=\sqrt{\frac{\rho_\ell}{r}},$$
$$\sin\alpha_u=\sqrt{\frac{\rho_u}{r}},$$
$$\cos\beta=\frac{r+r-\varepsilon_1}{2r}=1-\frac{\varepsilon_1}{2r}.$$

For τ_ℓ, we have

$$\begin{aligned}
\sqrt{\tau_\ell}&=\sqrt{r}\sin(\alpha_\ell-\beta)\\
&=\sqrt{r}(\sin\alpha_\ell\cos\beta-\cos\alpha_\ell\sin\beta)\\
&=\sqrt{r}\left(\sqrt{\frac{\rho_\ell}{r}}\left(1-\frac{\varepsilon_1}{2r}\right)-\sqrt{1-\frac{\rho_\ell}{r}}\sqrt{1-\left(1-\frac{\varepsilon_1}{2r}\right)^2}\right)\\
&\geqslant\sqrt{\rho_\ell}\left(1-\frac{\varepsilon_1}{2r}\right)-\sqrt{r-\rho_\ell}\sqrt{\frac{\varepsilon_1}{r}} &(62)\\
&\geqslant\sqrt{\rho-\varepsilon_2}\left(1-\frac{\varepsilon_1}{2r}\right)-\sqrt{\varepsilon_1} &(63)\\
&\geqslant(\sqrt{\rho}-\sqrt{\varepsilon_2})\left(1-\frac{\varepsilon_1}{2r}\right)-\sqrt{\varepsilon_1} &(64)\\
&\geqslant\sqrt{\rho}-\frac{\sqrt{\rho}\varepsilon_1}{2r}-\sqrt{\varepsilon_2}-\sqrt{\varepsilon_1} &(65)\\
&\geqslant\sqrt{\rho}-\left(\frac{1}{2\sqrt{r}}+1\right)\sqrt{\varepsilon_1}-\sqrt{\varepsilon_2}. &(66)
\end{aligned}$$

Eqn. (62) follows from the elementary inequality: $1-(1-x)^2\leqslant 2x$ for any $x\geqslant 0$. Eqn. (63) is by the assumption $\rho<r$. Eqn. (64) follows from the fact that $\sqrt{x-y}\geqslant\sqrt{x}-\sqrt{y}$ for any $x\geqslant y\geqslant 0$. In Eqn. (65), we drop the term $\frac{\sqrt{\varepsilon_2}\varepsilon_1}{2r}$. Eqn. (66) follows since $\rho<r$ and $\varepsilon_1\leqslant\sqrt{\varepsilon_1}$ for $0\leqslant\varepsilon_1\leqslant 1$.

For τ_u, we have

$$\begin{aligned}
\sqrt{\tau_u}&=\sqrt{r}\sin(\alpha_u+\beta)\\
&=\sqrt{r}(\sin\alpha_u\cos\beta+\cos\alpha_u\sin\beta)\\
&=\sqrt{r}\left(\sqrt{\frac{\rho_u}{r}}\left(1-\frac{\varepsilon_1}{2r}\right)+\sqrt{1-\frac{\rho_u}{r}}\sqrt{1-\left(1-\frac{\varepsilon_1}{2r}\right)^2}\right)\\
&\leqslant\sqrt{\rho_u}+\sqrt{r}\sqrt{\frac{\varepsilon_1}{r}}\\
&=\sqrt{\rho+\varepsilon_2}+\sqrt{\varepsilon_1}\\
&\leqslant\sqrt{\rho}+\sqrt{\varepsilon_2}+\sqrt{\varepsilon_1}. &(67)
\end{aligned}$$

Eqn. (67) follows from the elementary inequality: $\sqrt{x+y}\leqslant\sqrt{x}+\sqrt{y}$ for any $x,y\geqslant 0$.

Set $\varepsilon_2=\varepsilon$ and

$$\begin{aligned}
\sqrt{\varepsilon_\ell}&:=\left(\frac{1}{2\sqrt{r}}+1\right)\sqrt{\varepsilon_1}+\sqrt{\varepsilon_2}\\
&=\left(\frac{1}{2\sqrt{r}}+1\right)\sqrt{\varepsilon/2}+\sqrt{\varepsilon}\\
&=\left(\frac{1}{4\sqrt{r}}+\frac{3}{2}\right)\sqrt{\varepsilon},\\
\sqrt{\varepsilon_u}&:=\sqrt{\varepsilon_1}+\sqrt{\varepsilon_2}=\sqrt{\varepsilon/4}+\sqrt{\varepsilon}=\frac{3}{2}\sqrt{\varepsilon}.
\end{aligned}$$

Then the first term of the summand in Eqn. (61) is at most

$$\Pr\left[\frac{\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n\tau_\ell})\right|}{M}<\frac{1}{2}\gamma_\ell\right]$$
$$\leqslant\Pr\left[\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n}(\sqrt{\rho}-\sqrt{\varepsilon_\ell}))\right|<\frac{1}{2}\gamma_\ell M\right].$$

Note that

$$\begin{aligned}
&\mathbb{E}\left[\left|\mathcal{Y}\cap \mathcal{C}ap^{n-1}(\underline{z},\sqrt{n}(\sqrt{\rho}-\sqrt{\varepsilon_\ell}))\right|\right]\\
&=\frac{\mathrm{Area}\left(\mathcal{C}ap^{n-1}(\sqrt{n}(\sqrt{\rho}-\sqrt{\varepsilon_\ell}))\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}(\sqrt{nr})\right)}M\\
&=\gamma_\ell M.
\end{aligned}$$

Hence by the Chernoff bound, the above probability is at most

$$\exp\left(-\frac{(1/2)^2}{3}\gamma_\ell M\right)=e^{-e^{\Omega(n)}}.$$

By similar reasoning, the second term of the summand in Eqn. (61) is at most $\exp\left(-\frac{(1/2)^2}{3}\gamma_u M\right)=e^{-e^{\Omega(n)}}$. Since the concentration bounds are doubly exponentially small, the probability in Eqn. (60) is still $e^{-e^{\Omega(n)}}$ once we take union bounds over $\underline{z}\in\mathcal{Z}$ and $\tau\in\mathcal{I}$ which are (singly) exponential in total.

Finally, a union bound shows that with probability doubly exponentially close to 1, $\mathcal{Y}$ of size $M=(c/\sqrt{\varepsilon})^{n-o(n)}$ satisfies properties 1 and 3 simultaneously. This completes the proof. □

B. *Proof of Eqn.* (10)

$$\mathbb{E}[W]=\sum_{\mathcal{L}\in\binom{\mathcal{M}}{L}}\sum_{\underline{y}\in\mathcal{Y}}\Pr\left[\psi(\mathcal{L})\subset\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right] \quad (68)$$
$$=\binom{M}{L}|\mathcal{Y}|\,\mu^L \quad (69)$$
$$\geqslant(M/L)^L|\mathcal{Y}|\,\mu^L,$$

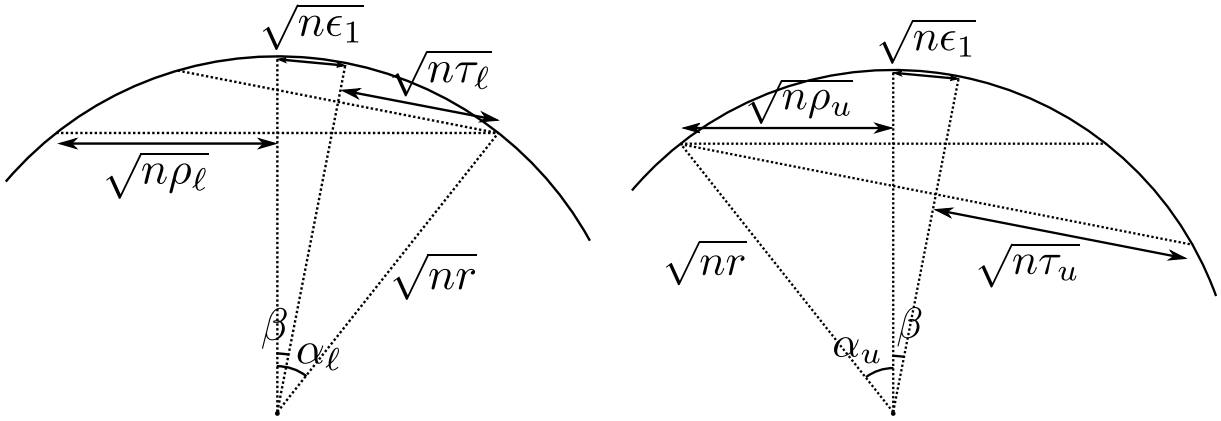
where in Eqn. (68) we use the shorthand notation

$$\psi(\mathcal{L}):=\{\psi(m)\colon m\in\mathcal{L}\},$$

and in Eqn. (69), μ is defined as follows,

$$\mu:=\frac{\mathrm{Area}\left(\mathcal{C}ap^{n-1}\left(\sqrt{nN}\right)\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}\left(\sqrt{nP}\right)\right)}.$$

This article has been accepted for publication in IEEE Transactions on Information Theory. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIT.2022.3189542



(a) (b)

Fig. 3: We consider a cap of radius $\sqrt{n\rho}$ centered at $\underline{y}$ on the sphere $\mathcal{S}^{n-1}(\sqrt{nr})$. Due to quantization error, the center of the cap is distorted to $\underline{z}$ which is at most $\sqrt{n\varepsilon_1}$ from $\underline{y}$ and the radius of the cap is distorted by ε_2. Let $\rho_\ell := \rho - \varepsilon_1$ and $\rho_u := \rho + \varepsilon_2$. The radii τ_ℓ and τ_u are such that the original cap is sandwiched between two tilted caps: $\mathcal{C}ap^{n-1}(\underline{z}, \sqrt{n\tau_\ell}) \subset \mathcal{C}ap^{n-1}(\underline{y}, \sqrt{n\rho}) \subset \mathcal{C}ap^{n-1}(\underline{z}, \sqrt{n\tau_u})$.

C. *Proof of Eqn.* (11)

For $\mathcal{L} = \{m_1, \cdots, m_L\}$ and $\underline{y} \in \mathcal{Y}$, define

$$\mathbb{I}\left(\underline{y}, \mathcal{L}\right) := \mathbb{1}\Big\{\psi\left(\mathcal{L}\right) \subset \mathcal{B}^n\left(\underline{y}, \sqrt{nN}\right)\Big\}.$$

Now the variance of W can be bounded from above as follows,

$$\begin{aligned}
\mathrm{Var}\left[W\right] =& \mathbb{E}\left[W^2\right] - \mathbb{E}\left[W\right]^2 \\
=& \sum_{\underline{y}_1, \underline{y}_2 \in \mathcal{Y}} \sum_{\mathcal{L}_1, \mathcal{L}_2 \in \binom{\mathcal{M}}{L}} \Bigg(\mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right) \mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] \\
& \qquad - \mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right)\right] \mathbb{E}\left[\mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] \Bigg) \\
\leqslant & \sum_{\substack{\mathcal{L}_1, \mathcal{L}_2 \in \binom{\mathcal{M}}{L} \\ \mathcal{L}_1 \cap \mathcal{L}_2 \neq \varnothing}} \sum_{\underline{y}_1, \underline{y}_2 \in \mathcal{Y}} \mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right) \mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] \qquad (70) \\
=& \left|\mathcal{Y}\right|^2 \sum_{\ell=1}^{L} \sum_{\substack{\mathcal{L}_1, \mathcal{L}_2 \in \binom{\mathcal{M}}{L} \\ |\mathcal{L}_1 \cap \mathcal{L}_2| = \ell}} \Pr_{\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2, \mathcal{C}}\left[\mathcal{E}(\mathcal{L}_1, \mathcal{L}_2, \underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2)\right], \qquad (71)
\end{aligned}$$

where

1) Eqn. (70) follows since for disjoint $\mathcal{L}_1$ and $\mathcal{L}_2$, $\mathbb{I}\left(y_1, \mathcal{L}_1\right)$ and $\mathbb{I}\left(y_2, \mathbb{L}_2\right)$ are independent and hence the corresponding summand vanishes; we upper bound the variance by dropping the negative term;
2) in Eqn. (71) the probability is taken over the code construction and the pair $\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2$ sampled independently and uniformly from $\mathcal{Y}$; the event $\mathcal{E}(\mathcal{L}_1, \mathcal{L}_2)$ is defined as

$$\mathcal{E}(\mathcal{L}_1, \mathcal{L}_2) := \Bigg\{ \psi\left(\mathcal{L}_1\right) \subset \mathcal{B}^n\left(\underline{\mathbf{y}}_1, \sqrt{nN}\right), \\ \psi\left(\mathcal{L}_2\right) \subset \mathcal{B}^n\left(\underline{\mathbf{y}}_2, \sqrt{nN}\right) \Bigg\}.$$

It is easy to verify that for any $m \in \mathcal{L}_1 \cap \mathcal{L}_2$,

$$\mathcal{E}(\mathcal{L}_1, \mathcal{L}_2) \subset \mathcal{E}_1(m, \mathcal{L}_1, \mathcal{L}_2) \cap \mathcal{E}_2(m, \mathcal{L}_1, \mathcal{L}_2) \cap \mathcal{E}_3(\mathcal{L}_1, \mathcal{L}_2),$$

where

$$\begin{aligned}
\mathcal{E}_1(m, \mathcal{L}_1, \mathcal{L}_2) :=& \Big\{ \underline{\mathbf{y}}_1 \in \mathcal{B}^n\left(\psi\left(m\right), \sqrt{nN}\right), \\
& \qquad \underline{\mathbf{y}}_2 \in \mathcal{B}^n\left(\psi\left(m\right), \sqrt{nN}\right) \Big\}, \\
\mathcal{E}_2(m, \mathcal{L}_1, \mathcal{L}_2) :=& \Big\{ \forall m_1 \in \mathcal{L}_1 \backslash \{m\},\ \psi\left(m_1\right) \in \mathcal{B}^n\left(\underline{\mathbf{y}}_1, \sqrt{nN}\right) \Big\}, \\
\mathcal{E}_3(\mathcal{L}_1, \mathcal{L}_2) :=& \Big\{ \forall m_2 \in \mathcal{L}_2 \backslash \mathcal{L}_1,\ \psi\left(m_2\right) \in \mathcal{B}^n\left(\underline{\mathbf{y}}_2, \sqrt{nN}\right) \Big\}.
\end{aligned}$$

Note that conditioned on $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_3$ are independent, and

$$\begin{aligned}
\Pr\left[\mathcal{E}_1\right] &= \left(\frac{\left|\mathcal{Y} \cap \mathcal{C}ap^{n-1}\left(\sqrt{n\rho}\right)\right|}{\left|\mathcal{Y}\right|} \right)^2 =: \nu^2, \\
\Pr\left[\mathcal{E}_2 \cap \mathcal{E}_3 | \mathcal{E}_1\right] &= \Pr\left[\mathcal{E}_2 | \mathcal{E}_1\right] \Pr\left[\mathcal{E}_3 | \mathcal{E}_1\right] \\
&= \mu^{L-1} \mu^{L-\ell} = \mu^{2L-\ell-1},
\end{aligned}$$

where $\rho := N\left(P - N\right)/P$ as shown in Fig. 4.

Let us upper bound ν and μ. For μ, we have

$$\mu \leqslant \frac{\mathrm{Area}\left(\mathcal{S}^{n-1}\left(\sqrt{nN}\right)\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}\left(\sqrt{nP}\right)\right)} = c_1 2^{-n\frac{1}{2}\log\frac{P}{N}}, \qquad (72)$$

where $c_1 := \sqrt{P/N}$.

For ν, we choose $\mathcal{Y}$ to be a $\sqrt{n\varepsilon}$-covering of $\mathcal{S}^{n-1}(\sqrt{nP})$ for some $\varepsilon > 0$ to be determined as given by Lemma 41 (with the choice $r = P - N$). Then

$$\begin{aligned}
\nu &\leqslant \frac{3}{2} \frac{\mathrm{Area}\left(\mathcal{C}ap^{n-1}(\sqrt{n}(\sqrt{\rho} + 3\sqrt{\varepsilon}/2))\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}(\sqrt{n(P-N)})\right)} \\
&\leqslant \frac{3}{2} \frac{\mathrm{Area}\left(\mathcal{S}^{n-1}(\sqrt{n}(\sqrt{\rho} + 3\sqrt{\varepsilon}/2))\right)}{\mathrm{Area}\left(\mathcal{S}^{n-1}(\sqrt{n(P-N)})\right)} \\
&= c_2 \left(\frac{\sqrt{\rho} + 3\sqrt{\varepsilon}/2}{\sqrt{P-N}} \right)^n, \qquad (73)
\end{aligned}$$

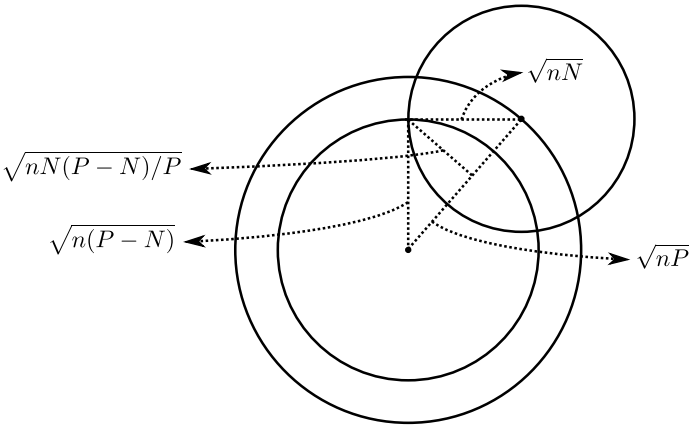


Fig. 4: If $\psi(m)$ is a codeword on $\mathcal{S}^{n-1}(\sqrt{nP})$, then a ball of radius $\sqrt{nN}$ around $\psi(m)$ induces a cap of radius $\sqrt{nN(P-N)/P}$ on the sphere $\mathcal{S}^{n-1}(\sqrt{n(P-N)})$ on which $\mathcal{Y}$ lives.

where $c_2 := \frac{3\sqrt{P}}{2(\sqrt{\rho}+3\sqrt{\varepsilon}/2)}$.

Note that the number of pairs $(\mathcal{L}_1, \mathcal{L}_2)$ with intersection size ℓ is

$$K_\ell := \binom{M}{\ell}\binom{M-\ell}{L-\ell}\binom{M-L}{L-\ell} \leqslant M^{2L-\ell}.$$

Hence overall we have

$$\begin{aligned}\mathrm{Var}\left[W\right] &\leqslant |\mathcal{Y}|^2 \sum_{\ell=1}^{L} K_\ell \nu^2 \mu^{2L-\ell-1} \\ &\leqslant |\mathcal{Y}|^2 \nu^2 \mu^{-1} \sum_{\ell=1}^{L} (M\mu)^{2L-\ell} \\ &\leqslant |\mathcal{Y}|^2 \nu^2 \mu^{-1} L\,(M\mu)^{L} \qquad (74) \\ &= |\mathcal{Y}|^2 L \nu^2 M^L \mu^{L-1},\end{aligned}$$

where Eqn. (74) is obtained by noting that $M\mu \leqslant 2^{nR} c_1 2^{-n\frac{1}{2}\log\frac{P}{N}} = c_1 2^{-\delta n}$ and taking the dominating term corresponding to $\ell = L$.

Appendix C
Other "Goodness" Properties of Regular Infinite Constellations

In Sec. VII, for technical reasons[22], we require the lattice Λ_0 to be simultaneously good for packing and covering. The existence of such lattices was established in [11]. We now give simpler proofs of the existence of (nonlattice) infinite constellations that satisfy these properties.

Let $\alpha > 0$. We allow α to be a function of n. Define $\mathcal{A} := [-\alpha/2, \alpha/2)^n$. We will study infinite constellations of the form $\mathcal{C} = \mathcal{C}' + \alpha\mathbb{Z}^n$ for finite sets $\mathcal{C}' \subset \mathcal{A}$. We assume that each point in $\mathcal{C}'$ is independent and uniformly distributed in $\mathcal{A}$. In other words, $\mathcal{C}$ is obtained by tiling a finite subset of random points from within a cube. See Fig. 5 for a pictorial illustration of the construction of such an IC ensemble. Since the IC is a tiling, it suffices to study finite sets of points in the space $\mathbb{R}^n \bmod \mathcal{A}$. [23]

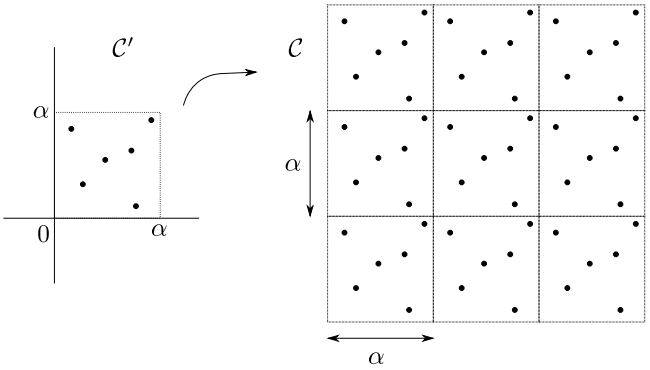


Fig. 5: Illustration of the class of infinite constellations studied in Appendix C.

Note that if $\mathcal{C}'$ forms a group with respect to addition modulo $\mathcal{A}$, then the resulting IC is a lattice. Construction-A lattices are essentially obtained by taking $\mathcal{C}'$ as an embedding of a linear code over a finite field into $\mathcal{A}$.

The density, NLD, effective volume and effective radius of an (α, M) IC are given by

$$\begin{aligned}\Delta(\mathcal{C}) &= M/\alpha^n, \\ R(\mathcal{C}) &= \frac{1}{n}\log\frac{M}{\alpha^n}, \\ V(\mathcal{C}) &= \frac{n}{\log M/\alpha^n}, \\ r_{\text{eff}}(\mathcal{C}) &= \left(\frac{\alpha^n}{V_n M}\right)^{1/n}. \qquad (75)\end{aligned}$$

respectively.

For any set $\mathcal{D} \subset \mathbb{R}^n$, define $\mathcal{D}^* := \mathcal{D} \bmod \mathcal{A}$.

A. Packing goodness

The packing radius of an IC $r_{\text{pack}}(\mathcal{C})$ is defined to be half the minimum distance between two points. We say that an infinite constellation is good for packing if $r_{\text{pack}}(\mathcal{C})/r_{\text{eff}}(\mathcal{C}) \geqslant 0.5 - o(1)$.

We give a greedy construction of ICs which is good for packing.

Choose α to be some constant larger than 4. We will construct an infinite constellation with packing radius at least 1.

The IC is constructed iteratively as follows: Start with an arbitrary point $\underline{x}_1$. At the i-th step, choose $\underline{x}_i$ to be an arbitrary point from $\mathcal{A}\backslash\cup_{j=1}^{i-1}\mathcal{B}^*(\underline{x}_j, 2)$. We repeat this till the $\mathcal{B}^*(\underline{x}_j, 2)$'s cover $\mathcal{A}$. Suppose that the algorithm terminates at the M-th step.

The construction ensures that the packing radius is at least 1. Moreover,

$$M \geqslant \frac{\alpha^n}{\mathrm{Vol}(\mathcal{B}(0,2))}.$$

[22] Specifically, in the proof, we need to take a union bound over centers in a Voronoi region.

[23] Since $\alpha\mathbb{Z}^n$ is a lattice, we slightly abuse notation and define $[\cdot] \bmod \mathcal{A} := [\cdot] \bmod \alpha\mathbb{Z}^n$.

However, $\alpha^n/M = V_n r_{\text{eff}}^n$. Using this in the above gives $r_{\text{pack}}/r_{\text{eff}} \geqslant 0.5$.

B. AWGN goodness

We say that an (α, M) infinite constellation $\mathcal{C}$ is good for AWGN (Additive White Gaussian Noise) [11] if for $\underline{\mathbf{z}} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ and $\underline{\mathbf{x}} \sim \mathcal{U}(\mathcal{C} \cap \mathcal{A})$, we have

$$\Pr[\|\underline{\mathbf{z}}\| > \|\underline{\mathbf{x}} + \underline{\mathbf{z}} - \underline{x}_j\| \text{ for some } \underline{x}_j \in \mathcal{C}] = 2^{-\Theta(n)}$$

where the probability is over the random choice of the codeword $\underline{\mathbf{x}}$ and the noise $\underline{\mathbf{z}}$. This is equal to the probability that a codeword different from the transmitted one is closer to the received vector when a random codeword is transmitted through an AWGN channel.

The following proposition recovers the achievability part of Poltyrev's [10] result:

Proposition 42. *Fix $\delta > 0$ and $N > 0$. A random $(4\sqrt{n\sigma^2}, M)$ constellation with M chosen so as to satisfy $r_{\text{eff}}/\sqrt{n\sigma^2} > 2^{\delta}$ is good for AWGN with probability $1 - 2^{-\Theta(n)}$.*

Proof. Since codewords are chosen uniformly, it suffices to assume that the first codeword is transmitted.

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}} \Pr[\|\underline{\mathbf{z}}\| > \|\underline{\mathbf{x}} + \underline{\mathbf{z}} - \underline{x}_j\| \text{ for some } \underline{x}_j \in \mathcal{C}] \\
&\leqslant \Pr[\|\underline{\mathbf{z}}\|^2 > 2^{\delta} n\sigma^2] \\
&\qquad + \Pr[\underline{x}_j \in \mathcal{B}^*(\underline{x}_1 + \underline{\mathbf{z}}, \sqrt{n\sigma^2}2^{\delta/2}) \text{ for some } j \neq 1] \\
&\leqslant \Pr[\|\underline{\mathbf{z}}\|^2 > 2^{\delta} n\sigma^2] + \sum_{j=2}^{M} \Pr[\underline{x}_j \in \mathcal{B}^*(\underline{x}_1 + \underline{\mathbf{z}}, \sqrt{n\sigma^2}2^{\delta/2})] \\
&\leqslant 2^{-\Theta(n)} + M \frac{\text{Vol}(\mathcal{B}(0, \sqrt{n\sigma^2}2^{n\delta/2}))}{\alpha^n} \\
&= 2^{-\Theta(n)} + \frac{\text{Vol}(\mathcal{B}(0, \sqrt{n\sigma^2}2^{n\delta/2}))}{\text{Vol}(\mathcal{B}(0, r_{\text{eff}}))} \\
&= 2^{-\Theta(n)}.
\end{aligned}$$

□

C. Covering goodness

We say that an infinite constellation $\mathcal{C}$ is a β-covering if for every $\underline{y} \in \mathbb{R}^n$,

$$\min_{\underline{x} \in \mathcal{C}} \|\underline{x} - \underline{y}\| \leqslant \beta.$$

The *covering radius* of an infinite constellation $\mathcal{C}$, denoted $r_{\text{cov}}(\mathcal{C})$, is the smallest $\beta > 0$ such that $\mathcal{C}$ is a β covering.

We say that a sequence of (α, M) ICs $\mathcal{C}$ is $(1+\delta)$-good for covering if

$$\limsup_{n \to \infty} \frac{r_{\text{cov}}(\mathcal{C})}{r_{\text{eff}}(\mathcal{C})} \leqslant 1 + \delta.$$

Proposition 43. *Fix any $N > 0$, and define $C \coloneqq \frac{1}{2}\log_2 \frac{1}{2\pi e N}$. Also choose $\alpha = 2\sqrt{nN}$, $\epsilon_n = \frac{1}{\log n}$, and $M = 2^{nC(1+\epsilon_n)}\alpha^n$.*

Then, a random (α, M) IC is $(1+\epsilon_n)2^{\epsilon_n C}$-good for covering[24] with probability $1 - 2^{-2^{\Theta(n/\log n)}}$.

[24]Essentially, $(1 + o(1))$-good for covering.

Proof. Define $\mathcal{Q} = \frac{\epsilon_n\sqrt{N}}{4}\mathbb{Z}^n$. We can conclude that $\mathcal{C}$ is a $\sqrt{nN}(1+\epsilon_n/4)$ covering if for all $\underline{y} \in \mathcal{Q}$, we have

$$\min_{\underline{x} \in \mathcal{C}} \|\underline{x} - \underline{y}\| \leqslant \sqrt{nN}.$$

Since

$$\begin{aligned}
r_{\text{eff}}(\mathcal{C}) &= \left(\frac{\Gamma(n/2+1)}{\pi^{n/2}} \times \frac{\alpha^n}{M}\right)^{1/n} \\
&= \left(\frac{\sqrt{n}2^{-R}}{\sqrt{2\pi e}}\right)(1 + o(1)) \\
&= \sqrt{nN} \times 2^{-\epsilon_n C},
\end{aligned}$$

this ensures that

$$\frac{r_{\text{cov}}(\mathcal{C})}{r_{\text{eff}}(\mathcal{C})} \leqslant (1+\epsilon_n)2^{\epsilon_n C}.$$

It is therefore sufficient to show that

$$\Pr\left[\max_{\underline{y} \in \mathcal{Q}} \min_{\underline{x} \in \mathcal{C}} \|\underline{x} - \underline{y}\| > \sqrt{nN}\right] = 2^{-2^{\Theta(n)}}.$$

To prove the rest, we simply find the probability that there is no point that is $\sqrt{nN}$-close to $\underline{y} \in \mathcal{Q}$, and then take a union bound over $\underline{y}$. To compute the aforementioned probability, we use a Chernoff bound.

Fix any $\underline{y} \in \mathcal{Q}$. Suppose that $\mathcal{C} = \{\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_M\}$, where each $\underline{\mathbf{x}}_i$ is is independent of the others and uniformly distributed over $[0, \alpha]^n$. Define $\xi_i(\underline{y})$ to be the indicator random variable which is 1 if $\|\underline{\mathbf{x}}_i - \underline{y}\| \leqslant \sqrt{nN}$, and zero otherwise.

We then have

$$\begin{aligned}
\mu &\coloneqq \mathbb{E}\left[\sum_{i=1}^{M} \xi_i(\underline{y})\right] \\
&= M \times \frac{\text{Vol}(\mathcal{B}(\underline{y}, \sqrt{nN}))}{\alpha^n} \\
&= 2^{nR}\,\text{Vol}(\mathcal{B}(\underline{y}, \sqrt{nN})) \\
&= 2^{nC\epsilon_n(1+o(1))}.
\end{aligned}$$

This implies, by the Chernoff bound, that

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{M} \xi_i(\underline{y}) = 0\right] &\leqslant \Pr\left[\sum_{i=1}^{M} \xi_i(\underline{y}) \leqslant \mu/2\right] \\
&\leqslant e^{-\mu/12} = 2^{-2^{\Theta(n/\log n)}}.
\end{aligned}$$

Therefore, the probability that random IC is not a $\sqrt{nN}(1+\epsilon/4)$-covering is upper bounded by

$$\begin{aligned}
\Pr\left[\max_{\underline{y} \in \mathcal{Q}} \min_{\underline{x} \in \mathcal{C}} \|\underline{x} - \underline{y}\| > \sqrt{nN}\right] &\leqslant \left(\frac{4\alpha}{\epsilon_n}\right)^n 2^{-2^{\Theta(n/\log n)}} \\
&= 2^{O(n\log n)} 2^{-2^{\Theta(n/\log n)}} \\
&= 2^{-2^{\Theta(n/\log n)}}.
\end{aligned}$$

This completes the proof. □

APPENDIX D
CONVERSE OF LIST DECODING CAPACITY THEOREM FOR INFINITE CONSTELLATIONS

Proposition 44. *For any $N > 0$ and $\delta > 0$, let $\mathcal{C}$ be an arbitrary (N, L)-list decodable IC of NLD $\frac{1}{2}\log\frac{1}{2\pi eN} + \delta$. Then $L \geqslant 2^{\Omega(\delta n)}$.*

Proof. Let $\mathcal{C} \subset \mathbb{R}^n$ be an arbitrary infinite constellation of NLD $\frac{1}{2}\log\frac{1}{2\pi eN} + \delta$ for some constant $\delta > 0$. Then there must exist a sufficiently large P such that

$$\frac{1}{n}\log\frac{\left|\mathcal{C}\cap\mathcal{B}^n\left(0,\sqrt{nP}\right)\right|}{\mathrm{Vol}\left(\mathcal{B}^n\left(0,\sqrt{nP}\right)\right)} \geqslant \frac{1}{2}\log\frac{1}{2\pi eN} + \frac{\delta}{2}.$$

Therefore,

$$\begin{aligned}
&\frac{1}{n}\log\left|\mathcal{C}\cap\mathcal{B}^n\left(0,\sqrt{nP}\right)\right| \\
&\geqslant \frac{1}{2}\log\frac{1}{2\pi eN} + \frac{1}{n}\log\mathrm{Vol}\left(\mathcal{B}^n\left(0,\sqrt{nP}\right)\right) + \frac{\delta}{2} \\
&= \frac{1}{2}\log\frac{1}{2\pi eN} + \frac{1}{n}\log\left(V_n\sqrt{nP}^n\right) + \frac{\delta}{2} \\
&\asymp \frac{1}{2}\log\frac{1}{2\pi eN} + \frac{1}{2}\log(2\pi eP) + \frac{\delta}{2} \\
&= \frac{1}{2}\log\frac{P}{N} + \frac{\delta}{2}.
\end{aligned}$$

By the list decoding converse for codes with power constraints [8, Lemma 33], since the code $\mathcal{C}\cap\mathcal{B}^n\left(0,\sqrt{nP}\right)$ has rate larger than the list decoding capacity $\frac{1}{2}\log\frac{P}{N}$, it must have exponential list sizes. That is, there must exist $\underline{y}\in\mathbb{R}^n$ such that $\left|\mathcal{C}\cap\mathcal{B}^n\left(0,\sqrt{nN}\right)\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right| \geqslant 2^{\Omega(\delta n)}$. This implies the existence of $\underline{y}\in\mathbb{R}^n$ such that $\left|\mathcal{C}\cap\mathcal{B}^n\left(\underline{y},\sqrt{nN}\right)\right| \geqslant 2^{\Omega(\delta n)}$, which completes the proof. □

REFERENCES

[1] Y. Zhang and S. Vatedka, "List decoding random euclidean codes and infinite constellations," in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1627–1631, IEEE, 2019. 1, I

[2] G. A. Kabatiansky and V. I. Levenshtein, "On bounds for packings on a sphere and in space," *Problemy Peredachi Informatsii*, vol. 14, no. 1, pp. 3–25, 1978. I

[3] N. Blachman, "On the capacity of bandlimited channel perturbed by statistically dependent interference," *IRE Transactions on Information Theory*, vol. 8, pp. 48–55, 1962. I

[4] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011. I

[5] V. Guruswami, *List Decoding of Error Correcting Codes (Lecture Notes in Computer Science),*. Springer-Verlag, NY, 2004. I, 5, II-B, II-B, III

[6] M. Langberg, "Private codes or succinct random codes that are (almost) perfect," in *Proc. IEEE Symp. Found. Comp. Sci.*, (Rome, Italy), 2004. I

[7] B. K. Dey, S. Jaggi, and M. Langberg, "Sufficiently myopic adversaries are blind," *IEEE Transactions on Information Theory*, vol. 65, no. 9, pp. 5718–5736, 2019. I

[8] Y. Zhang, S. Vatedka, S. Jaggi, and A. Sarwate, "Quadratically Constrained Myopic Adversarial Channels," *IEEE Transactions on Information Theory, accepted*, 2021. I, I-A, 1, I-A, 2, 4, I, D

[9] Y. Zhang, S. Vatedka, and S. Jaggi, "Quadratically constrained two-way adversarial channels," *arXiv preprint arXiv:2001.02575*, 2020. I

[10] G. Poltyrev, "On coding without restrictions for the awgn channel," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 409–417, 1994. I, C-B

[11] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3401–3416, 2005. I-A, V-A, 19, V-D, VII-C, C, C-B

[12] H. Groemer, "Existenzsätze für lagerungen im euklidishen raum," *Mathematische Zeitschrift*, vol. 81, no. 3, pp. 260–278, 1963. 2

[13] H. Cohn and N. Elkies, "New upper bounds on sphere packings i," *Annals of Mathematics*, pp. 689–714, 2003. 2

[14] F. Hosseinigoki and O. Kosut, "List-decoding capacity of the gaussian arbitrarily-varying channel," *Entropy*, vol. 21, no. 6, p. 575, 2019. 4

[15] F. Hosseinigoki, *Fundamental Limits of Gaussian Communication Networks in the Presence of Intelligent Jammers*. PhD thesis, Arizona State University, 2019. 4

[16] A. S. Mansour, H. Boche, and R. F. Schaefer, "Secrecy capacity under list decoding for a channel with a passive eavesdropper and an active jammer," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1972–1976, 2018. 4

[17] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert communication over adversarially jammed channels," *IEEE Transactions on Information Theory*, vol. 67, no. 9, pp. 6096–6121, 2021. 4

[18] V. V. Zyablov and M. S. Pinsker, "List concatenated decoding," *Problemy Peredachi Informatsii*, vol. 17, no. 4, pp. 29–33, 1981. II-A, 3, 1, II

[19] V. Guruswami and S. Narayanan, "Combinatorial limitations of average-radius list decoding," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pp. 591–606, Springer, 2013. II-A, II-A, II-A, II-B, II-B, III, IV-B, 1

[20] R. Li and M. Wootters, "Improved list-decodability of random linear binary codes," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, 2018. II-A, II-A, 7, 9, 7, II, III

[21] V. Guruswami, J. Hastad, and S. Kopparty, "On the List-Decodability of Random Linear Codes," in *Proc. ACM Symp. on Theory of Comp.*, 2010. 1, 3, 5, II, III

[22] V. Guruswami, J. Hastad, M. Sudan, and D. Zuckerman, "Combinatorial bounds for list decoding," *IEEE Transactions on Information Theory*, vol. 45, pp. 1021–1034, 2002. 2, 7, II

[23] M. Cheraghchi, V. Guruswami, and A. Velingker, "Restricted isometry of Fourier matrices and list decodability of random linear codes," *SIAM Journal on Computing*, vol. 42, pp. 1888–1914, 2013. 4, 5, II

[24] M. Wootters, "On the list decodability of random linear codes with large error rates," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 2013. 5, II

[25] A. Rudra and M. Wootters, "Every list-decodable code for high noise has abundant near-optimal rate puncturings," in *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 2014. 6

[26] A. Rudra and M. Wootters, "It'll probably work out: improved list-decoding through random operations," in *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, 2015. 6

[27] A. Rudra and M. Wootters, "Average-radius list-recoverability of random linear codes," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2018. 6

[28] J. Mosheiff, N. Resch, N. Ron-Zewi, S. Silas, and M. Wootters, "Ldpc codes achieve list decoding capacity," *arXiv preprint arXiv:1909.06430*, 2019. 8

[29] V. Guruswami, R. Li, J. Mosheiff, N. Resch, S. Silas, and M. Wootters, "Bounds for list-decoding and list-recovery of random linear codes," *arXiv preprint arXiv:2004.13247*, 2020. 8, II, III

[30] V. Guruswami, J. Mosheiff, N. Resch, S. Silas, and M. Wootters, "Sharp threshold rates for random codes," *arXiv preprint arXiv:2009.04553*, 2020. 8

[31] V. M. Blinovsky, "Bounds for codes in the case of list decoding of finite volume," *Problems of Information Transmission*, vol. 22, pp. 7–19, 1986. II-A, III

[32] V. M. Blinovsky, "Code bounds for multiple packings over a nonbinary finite alphabet," *Problems of Information Transmission*, vol. 41, pp. 23–32, 2005. II-A, III

[33] V. M. Blinovsky, "On the convexity of one coding-theory function," *Problems of Information Transmission*, vol. 44, pp. 34–39, 2008. II-A, III

[34] V. Guruswami and S. Vadhan, "A lower bound on list size for list decoding," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques.*, (Berlin, Heidelberg), pp. 318–329, 2005. II-A, III

[35] A. Ashikhmin, A. Barg, and S. Litsyn, "A new upper bound on codes decodable into size-2 lists," in *Numbers, Information and Complexity*, pp. 239–244, Springer, 2000. II-A

[36] Y. Polyanskiy, "Upper bound on list-decoding radius of binary codes," *IEEE Transactions on Information Theory*, vol. 62, no. 3, pp. 1119–1128, 2016. II-A

[37] Y. Zhang, A. J. Budkuley, and S. Jaggi, "Generalized List Decoding," in *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)* (T. Vidick, ed.), vol. 151 of *Leibniz International Proceedings*

in Informatics (LIPIcs), (Dagstuhl, Germany), pp. 51:1–51:83, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. II-A

[38] G. D. Cohen, S. N. Litsyn, and G. Zemor, "Upper bounds on generalized distances," vol. 40, pp. 2090–2092, November 1994. II-B, III

[39] A. Ta-Shma, "Explicit, almost optimal, epsilon-balanced codes," in *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2017. II-B

[40] A. Ben-Aroya, D. Doron, and A. Ta-Shma, "Near-optimal erasure list-decodable codes," in *35th Computational Complexity Conference (CCC 2020)*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. II-B

[41] I. Dinur, P. Harsha, T. Kaufman, I. L. Navon, and A. T. Shma, "List decoding with double samplers," in *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 2134–2153, SIAM, 2019. II-B

[42] V. L. Alev, F. G. Jeronimo, D. Quintana, S. Srivastava, and M. Tulsiani, "List decoding of direct sum codes," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1412–1425, SIAM, 2020. II-B

[43] F. G. Jeronimo, D. Quintana, S. Srivastava, and M. Tulsiani, "Unique decoding of explicit ε-balanced codes near the Gilbert-Varshamov bound," in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 434–445, 2020. II-B

[44] F. G. Jeronimo, S. Srivastava, and M. Tulsiani, "Near-linear time decoding of ta-shma's codes via splittable regularity," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1527–1536, 2021. II-B

[45] B. L. Hughes, "The smallest list for the arbitrarily varying channel," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 803–815, 1997. II-B

[46] Y. Zhang, S. Jaggi, and A. J. Budkuley, "Tight List-Sizes for Oblivious AVCs under Constraints," *arXiv preprint arXiv:2009.03788*, 2020. II-B

[47] S. Shlosman and M. A. Tsfasman, "Random lattices and random sphere packings: typical properties," *arXiv preprint math-ph/0011040*, 2000. 1, 5

[48] V. M. Blinovsky, "Random sphere packing," *Problems of Information Transmission*, vol. 41, no. 4, pp. 319–330, 2005. 2

[49] Y. Zhang and S. Vatedka, "Bounds for multiple packing and list-decoding error exponents," *arXiv preprint arXiv:2107.05161*, 2021. 2, 3, 5

[50] V. Blinovsky, "Multiple packing of euclidean sphere," in *Proceedings of IEEE International Symposium on Information Theory*, p. 18, IEEE, 1997. 3, IV-B

[51] V. Blinovsky and S. Litsyn, "New asymptotic bounds on the size of list codes on euclidean sphere," in *2009 IEEE International Symposium on Information Theory*, pp. 1244–1247, IEEE, 2009. 3, IV-B

[52] E. Grigorescu and C. Peikert, "List decoding barnes-wall lattices," in *2012 IEEE 27th Conference on Computational Complexity*, pp. 316–325, IEEE, 2012. 4

[53] E. Mook and C. Peikert, "Lattice (list) decoding near minkowski's inequality," 2020. 4

[54] C. Canonne, "A short note on Poisson tail bounds." https://github.com/ccanonne/probabilitydistributiontoolbox/blob/master/poissonconcentration.pdf, 2019. 12

[55] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge University Press, 2014. V-A, V-B

[56] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, vol. 290. Springer Science & Business Media, 2013. V-B

[57] A. Barvinok, "Math 669: Combinatorics, Geometry and Complexity of Integer Points," 2013. V-B, V-B

[58] O. Ordentlich and U. Erez, "A simple proof for the existence of "good" pairs of nested lattices," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4439–4453, 2016. V-D, 22

[59] T. Tao, "Topics in random matrix theory," *American Mathematical Soc.*, vol. 132, 2012. VIII

[60] R. Neelamani, S. Dash, and R. G. Baraniuk, "On nearly orthogonal lattice bases and random lattices," *siam journal on discrete mathematics*, vol. 21, pp. 199–219, 2007. VIII

[61] S. Kim, *On the shape of a high-dimensional random lattice*. PhD thesis, Stanford University, 2008. IX

[62] C. L. Siegel, "A mean value theorem in geometry of numbers," *Annals of Mathematics*, pp. 340–347, 1945. IX, 30, IX-A, 31

[63] J.-F. Quint, "An introduction to the study of dynamical systems on homogeneous spaces." https://www.math.u-bordeaux.fr/~jquint/publications/RiggaCourse.pdf, 2013. IX

[64] C. A. Rogers, "Mean values over the space of lattices," *Acta mathematica*, pp. 249–287, 1955. IX-A, 32, IX-A, 33, 34

[65] A. Strömbergsson and A. Södergren, "On the generalized circle problem for a random lattice in large dimension," *Advances in Mathematics*, vol. 345, pp. 1042–1074, 2019. 32, 36

[66] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*. volume 259 of Graduate Texts in Mathematics. 8

[67] D. Goldstein and A. Mayer, "On the equidistribution of Hecke points," in *Forum Mathematicum*, vol. 15, (Berlin), pp. 165–190, New York: De Gruyter, January 2003. 8

[68] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, 1997. IX-A, 3

[69] C. A. Rogers, "The moments of the number of points of a lattice in a bounded set," *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pp. 225–251, 1955. IX-B

[70] C. A. Rogers, "The number of lattice points in a set," *Proceedings of the London Mathematical Society*, pp. 305–320, 1956. IX-B

[71] S. Kim, "Random lattice vectors in a set of size $O(n)$," *International Mathematics Research Notices*, 2016. 35

[72] N. Linial and J. Mosheiff, "On the weight distribution of random binary linear codes," *Random Structures & Algorithms*, vol. 56, no. 1, pp. 5–36, 2020. 4

[73] T. Kaufman and D. Mass, "Good Distance Lattices from High Dimensional Expanders." arXiv preprint arXiv:1803.02849 (2018). 6