Multiple Packing: Lower Bounds via Error Exponents

Yihan Zhang and Shashank Vatedka, Senior Member, IEEE

Abstract—We derive lower bounds on the maximal rates for multiple packings in high-dimensional Euclidean spaces. For any \( N > 0 \) and \( L \in \mathbb{Z}_{\geq 2} \), a multiple packing is a set \( C \) of points in \( \mathbb{R}^n \) such that any point in \( \mathbb{R}^n \) lies in the intersection of at most \( L - 1 \) balls of radius \( \sqrt{nN} \) around points in \( C \). This is a natural generalization of the sphere packing problem. We study the multiple packing problem for both bounded point sets whose points have norm at most \( \sqrt{nN} \) for some constant \( P > 0 \), and unbounded point sets whose points are allowed to be anywhere in \( \mathbb{R}^n \). Given a well-known connection with coding theory, multiple packings can be viewed as the Euclidean analog of list-decodable codes, which are well-studied over finite fields. We derive the best known lower bounds on the optimal multiple packing density. This is accomplished by establishing an inequality which relates the list-decoding error exponent for additive white Gaussian noise channels, a quantity of average-case nature, to the list-decoding radius, a quantity of worst-case nature. We also derive novel bounds on the list-decoding error exponent for infinite constellations and closed-form expressions for the list-decoding error exponents for the power-constrained AWGN channel, which may be of independent interest beyond multiple packing.

Index Terms—Channel coding, communication channels, channel capacity, combinatorial mathematics, error correction codes.

I. INTRODUCTION

We study the problem of multiple packing in Euclidean space, a natural generalization of the sphere packing problem [1]. Let \( P > 0 \), \( N > 0 \) and \( L \in \mathbb{Z}_{\geq 2} \). We say that a point set \( C \) in \( B^n(\sqrt{nN}) \) forms a \((P, N, L - 1)\)-multiple packing if every point in \( \mathbb{R}^n \) lies in the intersection of at most \( L - 1 \) balls of radius \( \sqrt{nN} \) around points in \( C \). Equivalently, the radius of the smallest ball containing any size-L subset of \( C \) is larger than \( \sqrt{nN} \). This radius is known as the Chebyshev radius of the L-sized subset.

If \( L = 2 \), then \( C \) forms a sphere packing, i.e., a point set such that balls of radius \( \sqrt{nN} \) around points in \( C \) are disjoint,

\[
R(C) := \frac{1}{n} \ln |C|.
\]

Denote by \( C_{L-1}(P, N) \) the largest rate of a \((P, N, L - 1)\)-multiple packing as \( n \to \infty \). We will also refer to this as the adversarial list-decoding capacity, or simply the list-decoding capacity. Note that \( C_{L-1}(P, N) \) depends on \( P \) and \( N \) only through their ratio \( N/P \), which we call the noise-to-signal ratio. The goal of this paper is to derive lower bounds on \( C_{L-1}(P, N) \).

The problem of multiple packing is closely related to the list-decoding problem [2], [3] in coding theory. Indeed, a multiple packing can be viewed as the Euclidean analog of a list-decodable code. We will interchangeably use the terms “packing” and “code” to refer to the point set of interest. To see the connection, note that if any point/codeword in a multiple packing is transmitted through an adversarial omniscient jamming channel that can inflict an arbitrary additive noise of Euclidean norm at most \( \sqrt{nN} \), then given the distorted transmission, one can decode to a list of the nearest \( L - 1 \) codewords which is guaranteed to contain the transmitted codeword. The quantity \( C_{L-1}(P, N) \) can therefore be interpreted as the capacity of this channel.

List-decodable codes are also useful in other problems involving communication in the presence of adversaries. It is known that with a small amount of shared secret key between the transmitter and receiver, list-decodable codes can be turned into unique-decodable codes for the adversarial channel mentioned previously, and the receiver can uniquely decode to the correct codeword with a vanishingly small probability of error [4], [5], [6]. List-decoding also serves as a proof technique for deriving bounds on the (unique-decoding) capacity for various adversarial jamming channels; see, e.g., [7] and [8].

1Here we use \( B^n(r) \) to denote an \( n \)-dimensional Euclidean ball of radius \( r \) centered at the origin.

2We choose to stick with \( L - 1 \) rather than \( L \) for notational convenience. This is because in the proof, we need to examine the violation of \((L - 1)\)-packing, i.e., the existence of an \( L \)-sized subset that lies in a ball of radius \( \sqrt{nN} \).

3An omniscient adversary is one who can choose the jamming/additive noise vector that must satisfy a power constraint but otherwise be any function of the codebook and the transmitted codeword (available noncausally to the jammer). This is more powerful than an oblivious jammer, who can transmit a jamming vector that can only depend on the codebook but not the transmitted codeword.
A. Bounded Packings

We now review the results known in the literature.

Let us start with the $L = 2$ case, which corresponds to the sphere packing problem. The best known lower bound is due to Blachman in 1962 [9] using a simple volume packing argument. The best known upper bound is due to Kabatiansky and Levenshtein in 1978 [10] using the seminal Delsarte’s linearm programing framework [11] from coding theory. These bounds meet nowhere except at two points: $N/P = 0$ (where $C_{L-1}(P, N) = \infty$), and $N/P = 1/2$ (where $C_{L-1}(P, N) = 0$).

For $L > 2$, Blinovsky [12] derived a lower bound (Equation (3)) on $C_{L-1}(P, N)$, and in fact our results are closely related to this work. In the present paper, we use a different approach to obtain the same bound, and also extend the results to unbounded packings. Please see Section VIII-F for an in-depth discussion of the connection to [12]. To the best of our knowledge, this is the best known lower bound on $C_{L-1}(P, N)$.

Blinovsky [12] also derived an upper bound using the ideas of the Plotkin bound [13] and the Elias–Bassalygo bound [14] in coding theory. The same upper bound was originally shown by Blachman and Few [15] using a more involved approach. Blinovsky and Litsyn [16] later improved this result in the low-rate regime by a recursive application of a bound on the distance distribution by Ben-Haim and Litsyn [17]. The latter in turn relies on the Kabatiansky–Levenshtein linear programming bound [10]. Blinovsky and Litsyn [16] numerically verified that their bounds improve previous ones when the rate is sufficiently low, but no explicit expression was provided. More recently, Zhang and Vatedka [18] derived various upper and lower bounds on the list-decoding capacity and a related notion known as the average-radius list-decoding capacity.

More recently, Alon et al. [19] derived bounds on the list decoding radius for specified list sizes in the zero-rate regime for spherical codes as well as Hamming space.

B. Unbounded Packings

The above notion of $(P, N, L - 1)$-multiple packing is well defined even if we remove the restriction that all points must lie in $\mathbb{E}^n(\sqrt{n}P)$, and instead allow the packing to contain points anywhere in $\mathbb{R}^n$. The codebook can now be countably infinite, and this leads to the notion of $(N, L - 1)$-multiple packing. The density of such an unbounded packing is measured by the (normalized) number of points per volume

$$R(\mathcal{C}) := \liminf_{K \to \infty} \frac{1}{n} \ln \frac{\mathcal{C} \cap [-K, K]^n}{||[-K, K]^n||}. \quad (2)$$

With slight abuse of terminology, we call $R(\mathcal{C})$ the rate of the unbounded packing $\mathcal{C}$. This is also referred to as the normalized logarithmic density (NLD) in the literature. The largest density of unbounded multiple packings as $n \to \infty$ is denoted by $C_{L-1}(N)$.

For $L = 2$, the unbounded sphere packing problem has a long history since at least the Kepler conjecture [20] in 1611. The best known lower bound is given by a straightforward volume packing argument [21]. The best known upper bound is obtained by reducing it to the bounded case for which we have the Kabatiansky–Levenshtein linear programming-type bound [10]. For $L > 2$, Blinovsky [22] described a lower bound by analyzing an (expurgated) Poisson Point Process (PPP). Further results along similar lines can be found in Zhang and Vatedka [23].

For $L \to \infty$, Zhang and Vatedka [24] determined the limiting value of $C_{L-1}(N)$. The limit of $C_{L-1}(P, N)$ as $L \to \infty$ is a folklore in the literature and a proof can be found in [7].

Very little is known about structured packings. Grigorescu and Peikert [25] initiated the study of list-decodability of lattices. Some recent work can be found in Mook and Peikert [26], and Zhang and Vatedka [24] on list-decodability of random lattices and infinite constellations.

C. Error Exponents

Our lower bounds on $C_{L-1}(P, N)$ and $C_{L-1}(N)$ are derived by making an interesting connection between list-decodable codes for adversarial (omniscient jamming) channels and list-decodable codes for the additive white Gaussian noise (AWGN) channel. As discussed in Section VIII-F, a similar connection was made in [12] and our main contribution is a new proof of this result which is then extended to infinite constellations.

Loosely speaking, we show that any code that is $(L - 1)$-list-decodable over the AWGN $\mathcal{N}(0, \sigma^2)$ channel with exponentially decaying probability of error $e^{-nE + o(n)}$ for some $E > 0$ can be expurgated without loss of rate to give a code with Chebyshev radius $\sqrt{2n\sigma^2E + o(n)}$. Deriving tight upper and lower bounds on the list-decoding error exponent for AWGN channels is a classical problem studied in the literature [27]. For the bounded case, we use the best-known lower bounds for the expurgated exponent [28] to obtain lower bounds on the (adversarial) list-decoding capacity. A similar approach was used to derive lower bounds on the zero-rate threshold of binary channels under (adversarial) list-decoding in [29]. However, no lower bounds on the list-decoding capacity were derived below the zero-rate threshold.

List-decoding error exponents for general discrete memoryless channels (DMCs) were originally studied by Gallager [27] and Viterbi and Omura [30]. A more systematic study of list-decoding error exponents for DMCs was made by Merhav [28]. Merhav [28] gave bounds on the list-decoding random coding and expurgated error exponents for both constant and exponential (in $n$) list sizes. For the input-constrained AWGN channel, we also derive a closed-form expression for the expurgated exponent that numerically evaluates to the one derived by Merhav [28], which could be of independent interest.

For the input-unconstrained case, there is not much work on the list-decoding error exponents for the AWGN channel to the best of our knowledge. We derive closed-form expressions...
for the random coding and expurgated list-decoding error exponents using codebooks designed using Poisson point processes, improving upon the best known results in the literature. We then use these results to construct codes for the adversarial list-decoding problem, and derive lower bounds on the list decoding capacity.

D. List-Decoding

For $L = 2$, the problem of (unbounded) sphere packing has a long history and has been extensively studied, especially for small dimensions. The largest packing density is open for almost every dimension, except for $n = 1$ (trivial), 2 ([31], [32]), 3 (the Kepler conjecture, [33], [34]), 8 ([35]) and 24 ([36]). For $n \to \infty$, the best lower and upper bounds remain the trivial sphere packing bound and Kabatiansky–Levenshtein’s linear programming bound [10]. This paper is only concerned with (multiple) packings in high dimensions and we measure the density in the normalized way as mentioned in Section I.

There is a parallel line of research in combinatorial coding theory. Specifically, a uniquely-decodable code (resp. list-decodable code) is nothing but a sphere packing (resp. multiple packing) which has been extensively studied for $\mathbb{F}_q^n$ equipped with the Hamming metric.

We first list the best known results for sphere packing (i.e., $L = 2$) in Hamming spaces. For $q = 2$, the best lower and upper bounds are the Gilbert–Varshamov bound [37], [38] proved using a trivial volume packing argument and the second MRRW bound [39] proved using the seminal Delarte’s linear programming framework [11], respectively. Surprisingly, the Gilbert–Varshamov bound can be improved using algebraic geometry codes [40], [41] for $q \geq 49$. Note that such a phenomenon is absent in $\mathbb{F}_q^n$; as far as we know, no algebraic constructions of Euclidean sphere packings are known to beat the greedy/random constructions. For $q \geq n$, the largest packing density is known to exactly equal the Singleton bound [42], [43], [44] which is met by, for instance, the Reed–Solomon code [45].

Less is known for multiple packing in Hamming spaces. We first discuss the binary case (i.e., $q = 2$). For every $L \in \mathbb{Z}_{\geq 2}$, the best lower bound appears to be Blinovsky’s bound [46, Theorem 2, Chapter 2] proved under the stronger notion of average-radius list-decoding. The best upper bound for $L = 3$ is due to Ashikhmin, Barg and Litsyn [47] who combined the MRRW bound [39] and Litsyn’s bound [48] on distance distribution. For any $L \geq 4$, the best upper bound is essentially due to Blinovsky again [49], [46, Theorem 3, Chapter 2], though there are some partial improvements. In particular, the idea in [47] was recently generalized to larger $L$ by Polanskyi [50] who improved Blinovsky’s upper bound for even $L$ (i.e., odd $L - 1$) and sufficiently large $R$. Similar to [47], the proof also makes use of a bound on distance distribution due to Kalai and Linial [51] which in turn relies on Delarte’s linear programming bound. Very recently, [52] derived bounds on the size of a binary list decodable code for Hamming radius equal to 1. For larger $q$, Blinovsky’s lower and upper bounds [53], [54], [55, Chapter III, Lecture 9, §1 and 2] remain the best known.

As $L \to \infty$, the limiting value of the largest multiple packing density is a folklore in the literature known as the “list-decoding capacity” theorem. Moreover, the limiting value remains the same under a more general notion of average-radius list-decoding.

The problem of list-decoding was also studied for settings beyond the Hamming errors, e.g., list-decoding against erasures [56], [57], insertions/deletions [58], asymmetric errors [59], etc. Zhang et al. considered list-decoding over general adversarial channels [60]. List-decoding against other types of adversaries with limited knowledge such as oblivious or myopic adversaries were also considered in the literatures [7], [61], [62], [63], and [64].

1) Relation to Conference Version: This work was presented in part at the 2022 IEEE International Symposium on Information Theory [65]. All proofs were omitted in the published 6-page conference paper. The current article contains complete proofs of all results, and also includes several novel results on error exponents and list-decoding for Euclidean codes without power constraints.

II. Our Results

In this paper, we derive lower bounds on the largest multiple packing density for the bounded and the unbounded case. Let $C_{L - 1}(P, N)$ and $C_{L - 1}(N)$ denote the largest possible density of bounded and unbounded multiple packings, respectively.

A. Bounded Packings

In Theorem 3, we derive the following lower bound on the $(P, N, L - 1)$-list-decoding capacity:

$$C_{L - 1}(P, N) \geq \frac{1}{2} \left[ \frac{\ln (L - 1)P}{LN} + \frac{1}{L - 1} \ln \frac{P}{L(P - N)} \right].$$

(3)

The above result was also proved by Blinovskv [12] using a similar approach of connecting list-decoding for adversarial channels with the probability of error of list-decoding over AWGN channels. However, we use a different approach in connecting the Chebyshev radius of a code with the list-decoding error exponent for communication over AWGN channels. A more detailed discussion of the connections between these two works can be found in Section VIII-F.

It is a folklore (whose proof can be found in [7]) that as $L \to \infty$, $C_{L - 1}(P, N)$ converges to the following expression:

$$C_{LD}(P, N) = \frac{1}{2} \ln \frac{P}{N}.$$  

(4)

This quantity, and the bounds derived in [18] for $(P, N, L - 1)$-multiple packing are plotted in Figure 1 with $L = 5$. The horizontal axis is the noise-to-signal ratio $N/P$ and the vertical axis is the value of various bounds. Equation (3) turns out to be the largest lower bound for all $N, P \geq 0$ and $L \in \mathbb{Z}_{>2}$. Furthermore, it was shown in [18] via a completely different approach (Gallager’s bounding technique and large deviation principle) that the same bound also holds for

---

3It is an abuse of terminology to use “list-decoding capacity” here to refer to the large $L$ limit of the $(L - 1)$-list-decoding capacity.
B. Unbounded Packings

from below to Equation (4) as
on the capacity from [18] for
packing problem. In Theorem 12, the following lower bound
Elias-Bassalygo-type upper bound
We also plot in Figure 2 our lower bound together with an
expurgated spherical codes under average-radius list-decoding.

The lower bound using Gaussian codebooks and the upper bound (Equation (5))
are derived in [18].

C. List-Decoding Error Exponents

As alluded to above, our bounds on the multiple packing density (Equations (3), and (6)) are obtained via a curious connection to list-decoding error exponents of Additive White Gaussian Noise (AWGN) channels. Informally, the error exponent of a code \(C\) used over an AWGN channel is the asymptotic value of \(-\frac{1}{2} \ln(P_{\text{avg}}(C))\), where \(P_{\text{avg}}(C)\) is the average probability of error when the code is used to communicate over an AWGN channel. See Section VIII-A for formal definitions of the error exponents. Deriving tight bounds on the best achievable list-decoding error exponents is of independent interest in information theory. Another part of the contribution of this paper consists in the derivation

expurgated spherical codes under average-radius list-decoding.
We also plot in Figure 2 our lower bound together with an
Elias-Bassalygo-type upper bound

\[ C_{L-1}(P, N) \leq \frac{1}{2} \ln \frac{(L - 1)P}{LN} \]  

on the capacity from [18] for \(L = 3, 4, 5\). They both converge from below to Equation (4) as \(L\) increases.

We then derive various bounds for the \((N, L - 1)\)-multiple packing problem. In Theorem 12, the following lower bound
on \(C_{L-1}(N)\)

\[ C_{L-1}(N) \geq \frac{1}{2} \ln \frac{L - 1}{2\pi eNL} - \frac{\ln L}{2(L - 1)} \]  

is obtained via the connection with error exponents for the AWGN channel using a codebook generated using Poisson Point Processes (PPPs). In [23] it is shown that the same bound is in fact the exact asymptotics of a certain ensemble of infinite constellations under \((N, L - 1)\)-average-radius list-decoding (which is stronger than \((N, L - 1)\)-list-decoding).

It is known (see, e.g., [24]) that as \(L \to \infty\), \(C_{L-1}(N)\) converges to the following expression:

\[ C_{\text{LD}}(N) = \frac{1}{2} \ln \frac{1}{2\pi e N}. \]  

Therefore, our bound converges to \(C_{\text{LD}}(N)\) as \(L \to \infty\).

The bound in Equation (6) together with the Elias-Bassalygo-type upper bound [18]

\[ C_{L-1}(N) \leq \frac{1}{2} \ln \frac{L - 1}{2\pi e NL} \]  

are plotted in Figure 3 for \(L = 3, 4, 5\). The horizontal axis is \(N\) and the vertical axis is the value of various bounds. Equation (6) turns out to be the largest known lower bound for all \(N \geq 0\) and \(L \in \mathbb{Z}_{\geq 2}\). Equations (6) and (8) both converge from below to Equation (7) as \(L\) increases.

C. List-Decoding Error Exponents

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
of explicit lower bounds on the maximal error exponents for AWGN channels under list-decoding.

Let $\sigma > 0$ and $L \in \mathbb{Z}_{\geq 2}$. Consider a channel which takes as input an $\mathbb{R}^n$-valued vector and adds to it an $n$-dimensional independent Gaussian noise vector each entry i.i.d. with mean 0 and variance $\sigma^2$. The problem of deriving tight upper and lower bounds on the error exponent for unique decoding is a classical problem that dates back to Shannon [66]. The error exponent for random codes is known to be tight only above a certain rate, called the critical rate. Elias [2] observed that the critical rate can be decreased by considering larger list sizes $L > 2$. Closed form expressions for the random coding and sphere packing exponents appeared in [27] and [67]. The random coding and expurgated error exponents for general discrete memoryless channels and arbitrary constant list sizes can be obtained in parametric form [27] as the solution to an optimization problem. The behavior of the list decoding expurgated exponent at zero rate was studied by Blinovsky in [68].

1) Input Constrained Case: In the input (power) constrained case, the channel input $x$ is subject to a power constraint $\|x\|^2 \leq \sqrt{n}P$ for some $P > 0$. Merhav [28] analyzed the random coding and expurgated error exponents for discrete memoryless channels, and obtained nonexplicit expressions for the special case of the power-constrained AWGN channel.

Let $\text{snr} := P/\sigma^2$ denote the signal-to-noise ratio (SNR). It is well known that the capacity of the power-constrained AWGN channel is equal to [69] $\frac{1}{2} \ln(1 + \text{snr})$, and the error exponent for power-constrained AWGN channels can be bounded from below as follows [27]:

$$E_{\text{L}^{-1}}(R, \text{snr}) := \overline{E}_{\text{L}^{-1}}(R, \text{snr}),$$

$$R_{\text{crit}, L^{-1}}(\text{snr}) \leq R \leq \frac{1}{2} \ln(1 + \text{snr})$$

$$E_{\eta, L^{-1}}(R, \text{snr}), \quad 0 \leq R \leq R_{\text{crit}, L^{-1}}(\text{snr}),$$

where $E_{\text{L}^{-1}}$, $E_{\eta, L^{-1}}$, $E_{\text{ex}, L^{-1}}$ denote the random coding exponent, the straight line bound and the expurgated exponent, respectively. We derive the following expressions for these:

$$E_{\text{L}^{-1}}(R, \text{snr}) := \frac{1}{2} \left[ e^{2R} - \frac{\text{snr}(e^{2R} - 1)}{2} \left( 1 + \frac{4e^{2R}}{\text{snr}(e^{2R} - 1)} - 1 \right) \right] + \frac{\text{snr}}{4e^{2R}} \left( e^{2R} + 1 - (e^{2R} - 1) \right) \left( 1 + \frac{4e^{2R}}{\text{snr}(e^{2R} - 1)} \right),$$

$$E_{\eta, L^{-1}}(R, \text{snr}) := -R(L - 1) + \frac{L - 1}{2} \ln \left( L + \sqrt{(L - \text{snr})^2 + 4L\text{snr}} \right) + \frac{1}{2} \ln (L - \text{snr} + \sqrt{(L - \text{snr})^2 + 4L\text{snr}}) + \frac{1}{4} \left( L + \text{snr} - \sqrt{(L - \text{snr})^2 + 4L\text{snr}} \right) - \frac{L}{2} \ln(2L).$$

Although the behavior of the error exponents are well understood, we are not aware of a reference that provides the above explicit expressions, and this could be of independent interest. Merhav [28] derived an alternate expression for $R_{\text{crit}, L^{-1}}$, and this numerically evaluates to the expression that we obtain. The expression above gives a different parameterization of the expurgated exponent.

When specialized to $L = 1 = 1$, the above bounds recover the Gallager’s exponents [70], [27], Theorem 7.4.4 for unique-decoding. The above bounds are plotted in Figure 4 for $L = 1 = 1$ and $L = 1 = 2$, both with $\text{snr} = 1$ fixed.

2) Input Unconstrained Case: Many of these results can be generalized to AWGN channels without power constraints (a.k.a. infinite constellations), and the error exponents for the $L = 2$ case were derived by Poltyrev [71] by first constructing a finite codebook supported within a hypercube, and then tiling this codebook across $\mathbb{R}^n$. These ideas can be easily extended to derive lower bounds on the optimal error exponent for $L > 2$.

In this work, we derive closed-form expressions for the random coding and expurgated exponents using codebooks derived from Poisson point processes and Matérn point processes, and this is a novel feature of our work. In order to obtain a lower bound on the (adversarial) list decoding capacity, we will require an additional condition that the minimum pairwise distance between codewords be $\Theta(\sqrt{n})$. Unlike the coding scheme and expurgation process described in [71] for $L = 2$, our approach of deriving the expurgated exponent using Matérn point processes will naturally guarantee that the minimum distance is $\Theta(\sqrt{n})$.

In the input unconstrained case, the capacity of an AWGN channel with noise variance $\sigma^2$ was shown by Poltyrev [71] to be $\frac{1}{2} \ln \frac{1}{2\pi e \sigma^2}$. To the best of our knowledge, the extension of these results to $L > 2$ were not known previously, and this is a novel feature of our work. In Theorems 13 and 15, we prove that there exist codes of rate (as per Equation (2)) $R = \frac{1}{2} \ln \frac{1}{2\pi e \sigma^2}$ for some $\alpha \geq 1$ that under maximum likelihood $(L - 1)$-list-decoding attain an error exponent $E_{\text{L}^{-1}}(\alpha)$ defined
list-decoding remains the same for sufficiently large rate, i.e., in Figures 4b and 4c, respectively. Interestingly, the error exponent under the exponents. The list-decoding error exponents and the unique-decoding exponent, the straight line bound and the expurgated exponent, where \( R_{LD} = 1 \) and where \( R_{LD} = 1 \), list-decoding does increase the error exponent. Moreover, the critical rates \( R_{LD} \) and \( R_{LD} \) (see Equations (12) and (13)) become smaller than \( R_{LD} \) and \( R_{LD} \), respectively, under \((L-1)\)-list-decoding.

as follows:

\[
E_{L-1}(\alpha) \geq \begin{cases} 
E_{L-1}(\alpha), & 1 \leq \alpha \leq \sqrt{L} \\
\sqrt{L} \leq \alpha \leq \sqrt{2L}, & \alpha \geq \sqrt{2L} 
\end{cases}, 
\] (14)

where \( E_{L-1}, E_{L-1}, E_{L-1}, E_{L-1} \) denote the random coding exponent, the straight line bound and the expurgated exponent, respectively. These bounds read as follows:

\[
E_{L-1}(\alpha) := \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}, \\
E_{L-1}(\alpha) := \frac{L - 1}{2} - \frac{L}{2} \ln L + (L - 1) \ln \alpha, \\
E_{L-1}(\alpha) := \frac{\alpha^2}{16} + \frac{1}{16} \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} - \frac{L - 1}{2} \ln \left( \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} - \alpha^2 + 4 \right) + \frac{L - 2}{2} \ln \left( \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} + \alpha^2 + 4 \right) + \frac{3}{2} \ln 2 - \frac{1}{4}.
\]

When specialized to \( L = 1 \), the above bounds recover Poltyrev’s exponents [71, Theorem 3] for unique-decoding. The above bounds are plotted in Figure 5 for \( L = 1 \) and \( L = 2 \).

III. LIST-DECODING CAPACITY FOR LARGE L

All bounds in this paper hold for any fixed \( L \). In this section, we discuss the impact of our finite-\( L \) bounds on the understanding of the limiting values of the largest multiple packing density as \( L \to \infty \). Some of these results were known previously and others follow from the bounds in the current paper.

Characterizing \( C_{L-1}(P, N) \) or \( C_{L-1}(N) \) is a difficult task that is out of reach given the current techniques. However, if the list-size \( L \) is allowed to grow, we can actually characterize

\[
C_{LD}(P, N) := \lim_{L \to \infty} C_{L-1}(P, N), \\
C_{LD}(N) := \lim_{L \to \infty} C_{L-1}(N),
\]

where the subscript LD denotes List-Decoding.

It is well-known that \( C_{LD}(P, N) = \frac{1}{2} \ln \frac{1}{N} \). Specifically, the following theorem appears to be a folklore in the literature and a complete proof can be found in [7].

Theorem 1 (Folklore, [7]): Let \( 0 < N \leq P \). Then for any \( \varepsilon > 0 \),
1) There exist $(P, N, L - 1)$-multiple packings of rate $\frac{1}{2} \ln \frac{P}{N} - \varepsilon$ for some $L = \mathcal{O}(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$;  
2) Any $(P, N, L - 1)$-multiple packing of rate $\frac{1}{2} \ln \frac{P}{N} + \varepsilon$ must satisfy $L = e^{\Omega(n\varepsilon)}$. 
Therefore, $C_{LD}(P, N) = \frac{1}{2} \ln \frac{P}{N} - \varepsilon$.

A simple calculation reveals that Equation (3) equals $C_{LD}(P, N) - \Theta(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$ for large $L$. This implies that we can construct $(P, N, L - 1)$ multiple packings of rate $C_{LD}(P, N) - \varepsilon$ and $L = \Theta(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$, thereby recovering the above result. It is an interesting open question to resolve whether this is indeed the right scaling.

The unbounded version $C_{LD}(N)$ is characterized in [24] which equals $\frac{1}{2} \ln \frac{1}{2\pi eN}$.

**Theorem 2 [24]:** Let $N > 0$. Then for any $\varepsilon > 0$,

1) There exist $(N, L - 1)$-multiple packings of rate $\frac{1}{2} \ln \frac{1}{2\pi eN} - \varepsilon$ for some $L = \mathcal{O}(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$;
2) Any $(N, L - 1)$-multiple packing of rate $\frac{1}{2} \ln \frac{1}{2\pi eN} + \varepsilon$ must satisfy $L = e^{\Omega(n\varepsilon)}$.

Therefore, $C_{LD}(N) = \frac{1}{2} \ln \frac{1}{2\pi eN}$. For large $L$, our lower bound in Equation (6) reduces to $C_{LD}(N) - \Theta(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$. Once again, we get that for rates that are $\varepsilon$-close to capacity, the list size scales as $\Omega(\frac{1}{\varepsilon} \ln \frac{1}{\varepsilon})$ thereby recovering the above result.

**IV. OUR TECHNIQUES**

To derive lower bounds on list-decoding capacity, the most popular strategy is random coding with expurgation [18], a standard tool from information theory. To show the existence of a list-decodable code of rate $R$, we can simply randomly sample $e^{nR}$ points independently each according to a certain distribution. We then throw away (a.k.a. expurgate) one point from each of the bad lists. By carefully analyzing the error event and choosing a proper rate, we can guarantee that the remaining code has essentially the same rate after the removal process. We then get a list-decodable code of rate $R$ by noting that the remaining code contains no bad lists. This strategy is sufficient to derive the list decoding capacity when the list size is sufficiently large.

In this paper, we derive lower bounds on the list decoding capacity by analyzing the Chebyshev radius of the code. The challenge is that analyzing the error event involving the Chebyshev radius is a tricky task. In this paper, we take a different approach by deriving a lower bound on the Chebyshev radius of a (possibly expurgated) code as a function of the error exponent of the code when used over an AWGN channel. The lower bound on the list decoding capacity using a similar high-level connection to error exponents originally appeared in [12] (see Section VIII-F for a discussion). We use different ideas to achieve this, and provide a complete alternate proof in Section VIII, which is a major contribution of this work. Towards this end, we provide geometric understanding of the higher-order Voronoi partition induced by $L$-lists which naturally arises as the error regions under maximum likelihood list-decoding. We obtain sharp estimates on the Gaussian measure of the higher-order Voronoi region associated with a list which relates the error probability to the Chebyshev radius of the list. See Theorem 4 for the precise statement.

Inequality bridges two quantities of fundamentally different natures. The Chebyshev radius is a combinatorial characteristic of a code against worst-case errors, whereas the error exponent is a probabilistic characteristic of a code against average-case errors. The multiple packing problem then reduces to bounding the error exponent.

Our results on list-decoding error exponents of Gaussian channels are of independent interest beyond the study of multiple packing. Specifically, in the bounded case, the error exponents are well understood [27], [28], [70] but we derive explicit expressions for the exponents in terms of the input and noise powers; in the unbounded case, we borrow ideas from [72] and [73] and analyze PPPs and their expurgated versions (known as Matérn processes) using tools from stochastic geometry, e.g., the Slivnyak’s theorem and the Campbell’s theorem.

**V. ORGANIZATION OF THE PAPER**

The rest of the paper is organized as follows. Notational conventions are listed in Section VI, and some useful facts/lemmas are listed in Section A. After that, we present in Section VII the formal definitions of multiple packing and pertaining notions.

In Section VIII, we prove the inequality that relates the Chebyshev radius to error exponent and combine it with bounds on error exponent to obtain lower bounds on the largest multiple packing density. We also derive expressions for the list decoding random coding and expurgated exponents for infinite constellations in Section IX. We also derive closed-form expressions for the list decoding random coding and expurgated exponents for power constrained AWGN channels in the appendix. We end the paper with several open questions in Section X.

**VI. NOTATION**

**A. Conventions**

Sets are denoted by capital letters in calligraphic typeface, e.g., $\mathcal{C}, \mathcal{B}$, etc. Random variables are denoted by lower case letters in boldface or capital letters in plain typeface, e.g., $x, s$, etc. Their realizations are denoted by corresponding lower case letters in plain typeface, e.g., $x, s$, etc. Vectors (random or fixed) of length $n$, where $n$ is the blocklength without further specification, are denoted by lower case letters with underlines, e.g., $\underline{x}, \underline{g}, \underline{x} - \underline{g}$, etc. Vectors of length different from $n$ are denoted by an arrow on top and the length will be specified whenever used, e.g., $\underline{t}, \underline{\alpha}$, etc. The $i$-th entry of a vector $\underline{z} \in \mathbb{R}^n$ is denoted by $\underline{z}(i)$ since we can alternatively think of $\underline{z}$ as a function from $[n]$ to $\mathbb{R}$. Same for a random vector $\underline{X}$. Matrices are denoted by capital letters, e.g., $A, \Sigma$, etc. Similarly, the $(i, j)$-th entry of a matrix $G \in \mathbb{F}^{n \times m}$ is denoted by $G(i, j)$. We sometimes write $G_{n \times m}$ to explicitly specify its dimension. For square matrices, we write $G_n$ for short. Letter $I$ is reserved for identity matrix.

**B. Functions**

We use the standard Bachmann–Landau (Big-Oh) notation for asymptotics of real-valued functions in positive integers.
For two real-valued functions \( f(n), g(n) \) of positive integers, we say that \( f(n) \) asymptotically equals \( g(n) \), denoted \( f(n) \asymp g(n) \), if
\[
\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.
\]
For instance, \( 2^{n + \log n} \asymp 2^n \).\( 2^{n + \log n} \neq 2^n \).
We write \( f(n) \asymp g(n) \) (read \( f(n) \) dot equals \( g(n) \)) if the coefficients of the dominant terms in the exponents of \( f(n) \) and \( g(n) \) match,
\[
\lim_{n \to \infty} \frac{\log f(n)}{\log g(n)} = 1.
\]
For instance, \( 2^{3n} \asymp 2^{n + \frac{n}{14}} \), \( 2^{n} \neq 2^{n + \log n} \). Note that \( f(n) \asymp g(n) \) implies \( f(n) \asymp g(n) \), but the converse is not true.

For any \( q \in \mathbb{R}_{>0} \), we write \( \log_q(\cdot) \) for the logarithm to the base \( q \). In particular, let \( \log(\cdot) \) and \( \ln(\cdot) \) denote logarithms to the base 2 and \( e \), respectively.

For any \( A \subseteq \Omega \), the indicator function of \( A \) is defined as, for any \( x \in \Omega \),
\[
\mathbb{I}_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}
\]
At times, we will slightly abuse notation by saying that \( \mathbb{I}_A \) is 1 when event \( A \) happens and 0 otherwise. Note that \( \mathbb{I}_A(\cdot) = \mathbb{I}\{\cdot \in A\} \).

### C. Sets

For any two nonempty sets \( A \) and \( B \) with addition and multiplication by a real scalar, let \( A + B \) denote the Minkowski sum of them which is defined as \( A + B := \{a + b : a \in A, b \in B\} \). If \( A = \{x\} \) is a singleton set, we write \( x + B \) and for \( \{x\} + B \).

For any \( r \in \mathbb{R} \), the \( r \)-dilation of \( A \) is defined as \( rA := \{ra : a \in A\} \). In particular, \( \cdots \).\( A := (-1)A \).

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

### D. Geometry

Let \( \|\cdot\|_2 \) denote the Euclidean/\( l_2 \)-norm. Specifically, for any \( \mathbf{z} \in \mathbb{R}^n \),
\[
\|\mathbf{z}\|_2 := \left( \sum_{i=1}^{n} z(i)^2 \right)^{1/2}.
\]
With slight abuse of notation, we let \( |\cdot| \) denote the “volume” of a set w.r.t. a measure that is obvious from the context. If \( A \) is a finite set, then \( |A| \) denotes the cardinality of \( A \) w.r.t. the counting measure. For a set \( A \subseteq \mathbb{R}^n \), let
\[
\text{aff}(A) := \left\{ \sum_{i=1}^{k} \lambda_i \mathbf{g}_i : k \in \mathbb{Z}_{\geq 1}; \right. \\
\left. \forall i \in [k], \mathbf{g}_i \in A, \lambda_i \in \mathbb{R}, \sum_{i=1}^{k} \lambda_i = 1 \right\}
\]
denote the affine hull of \( A \), i.e., the smallest affine subspace containing \( A \). If \( A \) is a connected compact set in \( \mathbb{R}^n \) with nonempty interior and \( \text{aff}(A) = \mathbb{R}^n \), then \( |A| \) denotes the volume of \( A \) w.r.t. the \( n \)-dimensional Lebesgue measure. If \( \text{aff}(A) \) is a \( k \)-dimensional affine subspace for \( 1 \leq k < n \), then \( |A| \) denotes the \( k \)-dimensional Lebesgue volume of \( A \).

The closed \( n \)-dimensional Euclidean unit ball is defined as
\[
B^n := \{ y \in \mathbb{R}^n : \|y\|_2 \leq 1 \}.
\]
The \( (n-1) \)-dimensional Euclidean unit sphere is defined as
\[
S^{n-1} := \{ y \in \mathbb{R}^n : \|y\|_2 = 1 \}.
\]
For any \( \mathbf{z} \in \mathbb{R}^n \) and \( r \in \mathbb{R}_{>0} \), let \( B^n(\mathbf{z}, r) := rB^n \), \( S^{n-1}(\mathbf{z}, r) := rS^{n-1} \).
Let \( V_n := |B^n(\cdot)| \).

### VII. Basic Definitions and Facts

Given the intimate connection between packing and error-correcting codes, we will interchangeably use the terms “multiple packing” and “list-decodable code”. The parameter \( L \in \mathbb{Z}_{\geq 2} \) is called the multiplicity of overlap or the list-size. The parameters \( N \) and \( P \) (in the case of bounded packing) are called the noise power constraint and input power constraint, respectively. Elements of a packing are called either points or codewords. We will call a size-\( L \) subset of a packing an \( L \)-list.

This paper is only concerned with the fundamental limits of multiple packing for asymptotically large dimension \( n \to \infty \). When we say “\( a \)” code \( C \), we always mean an infinite sequence of codes \( \{C_i\}_{i \geq 1} \) where \( C_i \subseteq \mathbb{R}^n \) and \( \{n_i\}_{i \geq 1} \) is an increasing sequence of positive integers. We call \( C \) a spherical code if \( C \subseteq S^{n-1}(\sqrt{nP}) \) and we call it a ball code if \( C \subseteq B^n(\sqrt{nP}) \).

In the rest of this section, we list a sequence of formal definitions and some facts associated with these definitions.

**Definition 1 (Bounded Multiple Packing):** Let \( N, P > 0 \) and \( L \in \mathbb{Z}_{>2} \). A subset \( C \subseteq B^n(\sqrt{nP}) \) is called a \( (N, P, L-1) \)-list-decodable code (a.k.a. a \( (P, N, L-1) \)-multiple packing) if for every \( y \in \mathbb{R}^n \),
\[
|C \cap B^n(y, \sqrt{nN})| \leq L - 1. \tag{15}
\]
The rate of \( C \) is defined as
\[
R(C) := \frac{1}{n} \ln |C|. \tag{16}
\]

**Definition 2 (Unbounded Multiple Packing):** Let \( N > 0 \) and \( L \in \mathbb{Z}_{>2} \). A subset \( C \subseteq \mathbb{R}^n \) is called a \( (N, L-1) \)-list-decodable code (a.k.a. an \( (N, L-1) \)-multiple packing) if for every \( y \in \mathbb{R}^n \),
\[
|C \cap B^n(y, \sqrt{nN})| \leq L - 1. \tag{17}
\]
The rate (a.k.a. density) of \( C \) is defined as
\[
R(C) := \liminf_{K \to \infty} \frac{1}{n} \ln \frac{|C \cap [-K, K]^n|}{|[-K, K]^n|}. \tag{18}
\]

**Remark 1:** Strictly speaking, Equation (18) defines the lower density since \( \liminf \) is used. The limit may not exist for a general \( C \subseteq \mathbb{R}^n \). However, the limit exists for \( C \) considered in this paper. Indeed, \( C \) in this paper is either periodic.
(i.e., having points in a finite union of the translations of a certain lattice) or given by certain point processes whose density is known to exist on average.

Remark 2: With a slight abuse of notation, we will use $R(C)$ to refer to the rate of either a bounded packing or an unbounded packing. The meaning of $R(C)$ will be clear from the context. The rate of an unbounded packing (as per Equation (18)) is also called the normalized logarithmic density in the literature. It measures the rate (w.r.t. Equation (16)) per unit volume.

Note that the condition given by Equations (15) and (17) is equivalent to saying that for every $(x_1, \cdots, x_L) \in \left(\mathbb{E}_L\right)$,

$$\bigcap_{i=1}^L B^\infty(x_i, \sqrt{nN}) = \emptyset.$$  \hfill (19)

Definition 3 (Chebyshev Radius of a List): Let $x_1, \cdots, x_L$ be $L$ points in $\mathbb{R}^n$. Then the squared Chebyshev radius $\text{rad}^2(x_1, \cdots, x_L)$ of $x_1, \cdots, x_L$ is defined as the (squared) radius of the smallest ball containing $x_1, \cdots, x_L$, i.e.,

$$\text{rad}^2(x_1, \cdots, x_L) := \min \{ \max_{y \in \mathbb{R}^n, i \in [L]} \|x_i - y\|^2 \}.$$ \hfill (20)

Remark 3: Since $L$ is finite, the minimization needs to be performed only over $y$ in a compact convex set (e.g., the convex hull of $x_1, \cdots, x_L$), and hence the minimum is achieved.

Remark 4: One should note that for an $L$-list of points, the smallest ball containing $L$ is not necessarily the same as the circumscribed ball, i.e., the ball that contains all points in $L$ live on the boundary of the ball. The circumscribed ball of the polytope $\text{conv}\{L\}$ spanned by the points in $L$ may not exist. If it does exist, it is not necessarily the smallest one containing $L$. However, whenever it exists, the smallest ball containing $L$ must be the circumscribed ball of a certain subset of $L$.

Definition 4 (Chebyshev Radius of a Code): Given a code $C \subset \mathbb{R}^n$ of rate $R$, the squared $(L - 1)$-list-decoding radius of $C$ is defined as

$$\text{rad}^2_L(C) := \inf_{C \in \mathbb{E}_L} \text{rad}^2(C).$$ \hfill (21)

Note that $(L - 1)$-list-decodability defined by Equation (15) or Equation (19) is equivalent to $\text{rad}^2_L(C) > nN$. We also define the $(P, N, L - 1)$-list-decoding capacity (a.k.a. $(P, N, L - 1)$-multiple packing density)

$$C_{L-1}(P, N) := \lim_{n \to \infty} \sup_{C \subset \mathbb{R}^n: \text{rad}^2(C) > nN} R(C),$$

and the squared $(L - 1)$-list-decoding radius at rate $R$ with input constraint $P$

$$\text{rad}^2_L(P, R) := \lim_{n \to \infty} \sup_{C \subset \mathbb{R}^n: \text{rad}^2(C) > nN} \text{rad}^2_L(C),$$

and their unbounded analogues $(N, L - 1)$-list-decoding capacity (a.k.a. $(N, L - 1)$-multiple packing density) $C_{L-1}(N)$ and the squared $(L - 1)$-list-decoding radius $\text{rad}^2_L(R)$ at rate $R$:

$$C_{L-1}(N) := \lim_{n \to \infty} \sup_{C \subset \mathbb{E}_L: \text{rad}^2(C) > nN} R(C),$$

$$\text{rad}^2_L(R) := \lim_{n \to \infty} \sup_{C \subset \mathbb{E}_L: \text{rad}^2(C) > R} \text{rad}^2_L(C).$$

VIII. LOWER BOUNDS ON LIST-DECODING CAPACITY VIA ERROR EXPONENTS

In this section, we will show the following lower bound on $C_{L-1}(P, N)$.

Theorem 3: For any $P, N > 0$ such that $N \leq \frac{L-1}{L}P$ and any $L \in \mathbb{Z}_{\geq 2}$, the $(P, N, L - 1)$-list-decoding capacity $C_{L-1}(P, N)$ is at least

$$C_{L-1}(P, N) \geq \frac{1}{2} \left[ \ln \frac{(L-1)P}{LN} + \frac{1}{L-1} \ln \frac{P}{L(P-N)} \right].$$ \hfill (22)

Remark 5: When $L \to \infty$, the above bound (Equation (22)) converges to the list-decoding capacity $\frac{1}{2} \ln \frac{P}{N}$ for $L \to \infty$ (see Section III). For $L = 2$, it recovers the best known bound $\frac{1}{2} \ln \frac{P}{N(P-N)}$ (see, e.g., [18]). Furthermore, it is tight at $N/P = 0$ where the optimal density is $\infty$ and $N/P = \frac{L-1}{L}$ where the optimal density is 0 (see [18] for the Plotkin point).

To handle the Chebyshev radius, we follow an indirect approach which relates the Chebyshev radius to a quantity called error exponent. To this end, we take a detour by first introducing the notion of error exponent and then presenting bounds on it. We find it curious that the $(P, N, L - 1)$-list-decodability against worst-case errors can be related to the error exponent of a Gaussian channel that only inflicts average-case errors.

A. Basic Definitions Regarding List-Decoding Error Exponents

We first introduce maximum likelihood list-decoding and error exponents in the context of transmission over AWGN channels.

Consider a Gaussian channel $y = x + g$ where the input $x$ satisfies $\|x\|^2 \leq \sqrt{nP}$ and $g \sim N(0, \sigma^2 I_n)$ is an additive white Gaussian noise with mean zero and variance $\sigma^2$. Let $C = \{x_i\}_{i=1}^M$ be a codebook for the above Gaussian channel, that is, $\|x_i\|^2 \leq \sqrt{nP}$ for all $1 \leq i \leq M$.

We are interested in the probability of $(L - 1)$-list-decoding error of $C$ under the maximum likelihood (ML) $(L - 1)$-list-decoder. Formally, let $\text{Dec}_{L-1, \text{ML}}^C : \mathbb{R}^n \to \mathbb{E}_L$ denote the ML $(L - 1)$-list-decoder. Given $y$, the ML list-decoder outputs the list of the nearest $L - 1$ codewords in $C$ to $y$.

We say that an $(L - 1)$-list-decoding error occurs if the transmitted codeword $x_i$ does not lie within the list $\text{Dec}_{L-1, \text{ML}}^C(x_i + g)$. Let us define $\nu_{e_{L-1}}^\text{ML}(i, C)$ to be the conditional probability of a decoding error when the $i$-th codeword is transmitted, i.e., the probability that the decoder outputs a list of codewords that does not contain $x_i$, conditioned on the event that $x_i$ was sent:

$$P_{e_{L-1}}^{\text{ML}}(i, C) := \Pr [\text{Dec}_{L-1, \text{ML}}^C(x_i + g) \not\contains x_i]$$

$$= \Pr \left[ \exists \{i_1, \cdots, i_{L-1}\} \in \binom{[M]}{L-1}, \forall j \in [L-1], \\|x_{i_j} - (x_i + g)\|^2 < \|g\|^2 \right].$$
Occasionally, we also write \( P_{c,L-1}(x, C) \) to denote the same quantity above. Then, the \textit{average (over codewords) probability of \((L - 1)\)-list-decoding error of} \( C \) under \( \text{Dec}_{L-1,C} \) is defined as

\[
P_{c,\text{avg},L-1}(C) := \frac{1}{M} \sum_{i=1}^{M} P_{c,L-1}(i, C).
\]

\textbf{B. Connection Between List-Decoding Error Exponents and Chebyshev Radius}

In this subsection, we present a connection between list-decoding error exponents of a code used over an AWGN channel to the Chebyshev radius of the same code. We show that the Chebyshev radius of a code can be bounded by a quantity that depends on the probability of error of the code for transmission over a suitable AWGN channel.

\textit{Theorem 4:} For any code \( C = \{x_i\}_{i=1}^{M} \) with minimum pairwise distance between codewords being \( \Theta(\sqrt{n}) \), there exists a subcode \( C' \subset C \) with \( |C'| \geq M/2 \) with the property that for all \( \mathcal{L} \subset (C')_L \), we have

\[
P_{c,\text{avg},L-1}(C) \geq \exp\left(-\frac{\text{rad}^2(\mathcal{L})}{2\sigma^2} - o(n)\right).
\]

The above theorem follows from Theorem 5 and Theorem 6 below.

\textit{Lemma 5:} For any code \( C = \{x_i\}_{i=1}^{M} \), there exists a subcode \( C' \subset C \) of size \( M' \geq M/2 \) such that for all \( \mathcal{L} \subset (C')_L \),

\[
P_{c,\text{avg},L-1}(\mathcal{L}) \leq 2P_{c,\text{avg},L-1}(C),
\]

where

\[
P_{c,\text{avg},L-1}(C) := \frac{1}{L} \sum_{x \in C} P_{c,L-1}(x, \mathcal{L}),
\]

and

\[
P_{c,L-1}(x, \mathcal{L}) := \Pr[\text{Dec}_{L-1,C}(x + g) \neq x, \forall x' \in \mathcal{L} \setminus \{x\}, \|x' - (x + g)\|_2 < \|g\|_2].
\]

\textit{Proof:} Without loss of generality, assume that the codewords in \( C \) are listed according to ascending order of \( P_{c,L-1}(i, C) \), that is,

\[
P_{c,L-1}(1, C) \leq P_{c,L-1}(2, C) \leq \cdots \leq P_{c,L-1}(M, C).
\]

By Markov’s inequality (Lemma 17), each of the first (at least) \( M/2 \) codewords has probability of error at most \( 2P_{c,\text{avg},L-1}(C) \). Let \( C' := \{x_i\}_{i=1}^{M} \subset C \). Take any \( \mathcal{L} \in (C')_L \) and any \( x \in \mathcal{L} \).

\[
P_{c,L-1}(x, \mathcal{L}) := \Pr[\forall x' \in \mathcal{L} \setminus \{x\}, \|x' - (x + g)\|_2 < \|g\|_2] \leq \Pr[\bigcup_{\mathcal{L}' \in (C')_L} \|x' - (x + g)\|_2 < \|g\|_2] = P_{c,L-1}(x, C') \leq 2P_{c,\text{avg},L-1}(C).
\]

Therefore

\[
P_{c,\text{avg},L-1}(\mathcal{L}) \leq 2P_{c,\text{avg},L-1}(C),
\]

which completes the proof.

\textit{Theorem 6:} Let \( \mathcal{L} = \{x_1, \ldots, x_L\} \subset \mathbb{R}^n \) be an arbitrary set of \( L \) (where \( L \geq 2 \)) points in \( \mathbb{R}^n \) satisfying that there exist constants \( c, C > 0 \) independent of \( n \) such that for every \( 1 \leq i \neq j \leq L \),

\[
\sqrt{nc} \leq \|x_i - x_j\|_2 \leq \sqrt{nc}.
\]

Then,

\[
P_{c,\text{avg},L-1}(\mathcal{L}) \geq \exp\left(-\frac{\text{rad}^2(\mathcal{L})}{2\sigma^2} - o(n)\right).
\]

Note that the case where \( L - 1 = 1 \) is trivial which corresponds to unique-decoding. Indeed, suppose \( \mathcal{L} = \{x_1, x_2\} \). Without loss of generality, assume \( x_1 = 0 \in \mathbb{R}^n \) and \( x_2 = [a, 0, \ldots, 0] \in \mathbb{R}^n \) for some \( a = \|x_1 - x_2\| \). It is not hard to see that

\[
P_{c,\text{avg},L-1}(x_1, \mathcal{L}) = \Pr[\|x_2 - (x_1 + g)\|_2 < \|g\|_2] = \Pr[\|x_2 - g\|_2 < \|g\|_2] = \Pr(a - (g(1))^2 < g(1)^2] = \Pr(g(1) > a/2] = \exp\left(-\frac{(a/2)^2}{2\sigma^2} - o(n)\right).
\]

The last equality is by Lemma 18. By symmetry, \( P_{c,\text{avg},L-1}(x_1, \mathcal{L}) = P_{c,\text{avg},L-1}(x_2, \mathcal{L}) \) both of which are equal to \( P_{c,\text{avg},L-1}(\mathcal{L}) \). Since \( \sqrt{\text{rad}^2(\{x_1, x_2\})} = 1/2 \|x_1 - x_2\| = a/2 \), we see that Theorem 6 holds for \( L - 1 = 1 \).

We prove Theorem 6 in two subsequent subsections. The special case of \( L - 1 = 2 \) is easier to handle as it exhibits a simpler geometric structure and admits more explicit calculations. In fact we will prove a stronger statement:

\[
P_{c,\text{avg},L-1}(x_1, x_2, \mathcal{L}) = \exp\left(-\frac{\text{rad}^2(x_1, x_2, x_3)}{2\sigma^2} - o(n)\right).
\]

We prove this in Section VIII-C. We then prove Theorem 6 in Section VIII-D for general \( L - 1 \geq 2 \) using Laplace’s method (Theorem 21).

\textbf{C. Proof of Theorem 6 When \( L = 2 \)}

1) \textit{Voronoi Partition and Higher-Order Voronoi Partition:} We first introduce the notion of a \textit{Voronoi partition} induced by a point set and its \textit{higher-order} generalization.

Let \( C \subset \mathbb{R}^n \) be a discrete set of points. The \textit{Voronoi region} \( \mathcal{V}_C(x) \) associated with \( x \in C \) is defined as the region in which any point is closer to \( x \) than to any other points in \( C \), i.e.,

\[
\mathcal{V}_C(x) := \{y \in \mathbb{R}^n : \forall y' \in C \setminus \{x\}, \|y - y'\|_2 > \|y - x\|_2\}.
\]

When the underlying point set \( C \) is clear from the context, we write \( \mathcal{V}(x) \) for \( \mathcal{V}_C(x) \). Clearly, \( \mathcal{V}_C(x) \cap \mathcal{V}_C(x') = \emptyset \) for \( x \neq x' \in C \) and \( \bigcup_{x \in C} \mathcal{V}_C(x) \) is different from \( \mathbb{R}^n \) by a set of zero Lebesgue measure. The collection of Voronoi regions induced by \( C \) is called the \textit{Voronoi partition} induced by \( C \). It is not hard to see that for any \( C \subset \mathbb{R}^n \) and any \( x \in \mathbb{R}^n \),
the Voronoi region $V_C(z)$ contains exactly one point from $C$, which is $z$ itself.

Every Voronoi region can be written as an intersection of halfspaces. To compute $V(z)$ for any $z \in C$, one can draw a hyperplane bisecting and perpendicular to the segment connecting $z$ and $z'$ for each $z' \in C \setminus \{z\}$. Let $H_z(x)$ be the halfspace induced by the hyperplane that contains $z$, i.e.,

$$H_z(x) := \left\{ y \in \mathbb{R}^n : \langle y, z - z' \rangle \geq \frac{||z||^2 - ||z'||^2}{2} \right\}.$$  

Then $V(z)$ is nothing but the intersection of all such halfspaces, i.e.,

$$V_C(z) = \bigcap_{z' \in C \setminus \{z\}} H_z(x).$$

More generally, one can define Voronoi regions associated with subsets of points in $C$. Let $L \in \mathbb{Z}_{\geq 1}$. The order-$L$ Voronoi region $V_{C,L}(L)$ associated with $L \in \binom{C}{L}$ is defined as the region such that the set of the nearest $L$ points from $C$ to any point in the region is $L$, i.e.,

$$V_{C,L}(L) := \left\{ y \in \mathbb{R}^n : \forall z' \in C \setminus L, \left\| y - z' \right\|_2 > \max_{z \in L} \left\| y - z \right\|_2 \right\}. \quad (25)$$

Again, we will ignore the subscripts if they are clear. If $L = \{z\}$ is a singleton set, $V_{C,L}(\{z\}) = V_C(z)$. Clearly, $V_{C,L}(L) \cap V_{C,L}(L') = \emptyset$ for $L \neq L' \in \binom{C}{L}$ and $\bigcup_{L \in \binom{C}{L}} V_{C,L}(L) = \mathbb{R}^n$ (up to a set of measure zero). The collection of order-$L$ Voronoi regions induced by all $L$-subsets of $C$ is called the order-$L$ Voronoi partition induced by $C$.

Computing the order-$L$ Voronoi partition of a point set $C \subset \mathbb{R}^n$ is in general not easy for $L > 1$. Even when $n = 2$, i.e., all points in $C$ are on a plane, the problem is not trivial and the resulting order-$L$ Voronoi partition may exhibit significantly different behaviours from the $L = 1$ case [74, Fig. 2.5].

However, if one is given the order-$\binom{L}{L-1}$ Voronoi partition of $C$ and the (first order) Voronoi partition for all sets $C \setminus L'$ (where $L' \in \binom{C}{L-1}$), then the order-$L$ Voronoi partition of $C$ can be computed in the following way. For $L \in \binom{C}{L}$, to compute $V_{C,L}(L)$, for each $z \in L$, compute the following set $V_{C,L-1}(L \setminus \{z\}) \cap V_{C,(L \setminus \{z\})}(z)$. Then $V_{C,L}(L)$ is nothing but its unions, i.e.,

$$V_{C,L}(L) = \bigcup_{z \in L} V_{C,L-1}(L \setminus \{z\}) \cap V_{C,(L \setminus \{z\})}(z).$$

2) Connection to List-Decoding Error Probability for AWGN Channels: Let us return to the task of estimating the probability of $(L - 1)$-list-decoding error of an $L$-list $L \subset \mathbb{R}^n$. Given the order-$(L - 1)$ Voronoi partition of $L$, the error probability of any $z \in L$ can be written as

$$P_{e,L-1}(x, L) = Pr[z + g \in V_{L,L-1}(L \setminus \{z\})], \quad (26)$$

i.e., the probability that $z$ is the furthest point to $z + g$ among $C$.

Let $z_1, z_2, z_3$ be three distinct points in $\mathbb{R}^n$. In the proceeding two subsections, we divide the analysis of Equation (26) into two cases according to the largest angle of the triangle spanned by $z_1, z_2, z_3$.

3) Case 1: The Largest Angle of the Triangle Spanned by $z_1, z_2, z_3$ is Acute or Right: As shown in Figure 6, in this case, the smallest ball containing $z_1, z_2, z_3$ coincides with the circumscribed ball. As explained in Section VIII-C1, the Voronoi partition induced by $\{z_1, z_2, z_3\}$ can be easily computed and is depicted in the first figure of Figure 6. The second order Voronoi partition can be computed given the (first order) Voronoi partition. For example, $V(z_1, z_2)$ is comprised of the subregion in $V(z_1)$ whose points are closer to $z_2$ (such a subregion can be computed by computing the Voronoi partition with $z_2$ removed) and the subregion in $V(z_2)$ whose points are closer to $z_1$ (such a subregion can be computed by computing the Voronoi partition with $z_1$ removed). One observes that each of the resulting second order Voronoi regions may contain no (see $V(z_1, z_2)$), one (see $V(z_1, z_3)$) or two points (see $V(z_1, z_2, z_3)$) from the point set. This is in contrast with the (first order) Voronoi regions which only contain one point from the point set. In general, points can also be on the boundary of the higher-order Voronoi regions. This happens when, e.g., $z_1, z_2, z_3$ span an equilateral triangle.

To show Theorem 6 in this case, we need to estimate $P_{e,L-2}(\{z_1, z_2, z_3\})$. Consider the plane containing $z_1, z_2, z_3$. As depicted in Figure 7, let the center of the smallest ball containing $z_1, z_2, z_3$ be the origin, denoted by $O$. Let the ray going from $z_1$ to $O$ be the $x_1$ axis and the line perpendicular to it be the $x_2$ axis. Under this parameterization, $\|z_1\|^2 = \|z_2\|^2 = \|z_3\|^2 = r^2 = \sqrt{\rho^2(z_1, z_2, z_3)}$ and $(\ell_1 = i = \ell_2(\text{if } 3 \leq i \leq n) = 0$ for every $3 \leq i \leq n$.

Let us first estimate $P_{e,L-2}(\{z_1, z_2, z_3\})$. Suppose that in the plane spanned by $z_1, z_2, z_3$, the boundaries of $V(z_2, z_3)$ are given by two rays $L_1$ and $L_2$ as depicted in Figure 7.

It is not hard to check that if the largest angle of the triangle spanned by the three points is acute or right, then $V(z_2, z_3)$ belongs to the halfspace $\{x_1 \geq 0\}$ whereas $z_1$ belongs to the other halfspace $\{x_1 \leq 0\}$. Suppose $L_1$ and $L_2$ are parameterized by $x_1 = a_1 x_1$ and $x_2 = a_2 x_2$ for some constants $a_1$, $a_2 > 0$, respectively. Let $V := \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0, -a_1 x_1 \leq x_2 \leq a_1 x_1, r := \sqrt{\rho^2(z_1, z_2, z_3)}\}$ and $a := \max\{a_1, a_2\} > 0$. We are now ready to estimate $P_{e,L-2}(\{z_1, z_2, z_3\})$.

$$P_{e,L-2}(\{z_1, z_2, z_3\}) = Pr[\|z + g\| \in V(z_2, z_3)]$$

where

$$P_{e,L-2}(\{z_1, z_2, z_3\}) = Pr[\|z + g\| \in V(z_2, z_3)]$$

6We explain below why the slopes $a_1 > 0, a_2 > 0$ must be lower bounded by some constant independent of $n$. Let $\ell_1 := \|z_2 - z_3\|_2, \ell_2 := \|z_1 - z_2\|_2, \ell_3 := \|z_1 - z_1\|_2$. Under the assumptions in Theorem 6, it is guaranteed that $\ell_1 \leq \ell_2, \ell_3 \leq \Theta(\sqrt{n})$. It is a well-known fact that the circumradius of a triangle with side lengths $\ell_1, \ell_2, \ell_3$ is equal to $r = \frac{\ell_1 \ell_2 \ell_3}{4 \sqrt{\ell_1^2 + \ell_2^2}} \ell_3$ where $s = \frac{\ell_1 + \ell_2 + \ell_3}{2}$. Under the assumptions in Theorem 6, $r = \Theta(\sqrt{n})$. Let $\alpha_1, \alpha_2$ denote the angles between the $x_1$ axis and the rays $L_1, L_2$, respectively. Then $\sin \alpha_1 = \frac{\ell_2}{r}, \sin \alpha_2 = \frac{\ell_3}{r}$ where $r = \frac{\ell_1 \ell_2 \ell_3}{4 \sqrt{\ell_1^2 + \ell_2^2}} \ell_3$. We therefore get the relations

$$\frac{\alpha_1}{\sqrt{1 + \alpha_1}} = \frac{\ell_2}{r}, \frac{\alpha_2}{\sqrt{1 + \alpha_2}} = \frac{\ell_3}{r}$$

are the RHSS of which are on the order of $\Theta(1)$. Hence $\alpha_1 = \frac{\ell_2}{r}, \alpha_2 = \frac{\ell_3}{r}$, the RHSs of which are on the order of $\Theta(1)$. Hence $\alpha_1 = \frac{\ell_2}{r}, \alpha_2 = \frac{\ell_3}{r}$, the RHSs of which are on the order of $\Theta(1)$.
axis. The circumradius coincides with the Chebyshev radius which equals the shorthand notation in this case, the smallest ball containing \(x\). Fig. 6. The Voronoi partition (left) and the second order Voronoi partition (right) of \(\mathcal{V}(x_1, x_2, x_3)\) when \(x_1, x_2, x_3\) span an acute/right triangle. Note that in this case, the smallest ball containing \(x_1, x_2, x_3\) coincides with the circumscribed ball. That is, all points lie on the boundary of the ball. We use the shorthand notation \(\mathcal{V}(x) = \mathcal{V}(x_1, x_2, x_3)(x)\) and \(\mathcal{V}(x_1, x_2) = \mathcal{V}(\{x_1, x_2\}, \{x_3\})\).

\[
\mathcal{V}(x_1, x_2, x_3) = \mathcal{V}(x_1, x_2, x_3)(x) \quad \text{and} \quad \mathcal{V}(x_1, x_2) = \mathcal{V}(\{x_1, x_2\}, \{x_3\})
\]

Fig. 7. Suppose that the largest angle of the triangle spanned by \(x_1, x_2, x_3\) is acute or right. The origin \(O\) is set to be the center of the smallest ball containing \(x_1, x_2, x_3\). The \(x_1\) axis is set to be the ray going from \(x_1\) to \(O\) and the \(x_2\) axis is the ray perpendicular to the \(x_1\) axis. The circumradius coincides with the Chebyshev radius which equals \(r = \sqrt{\sigma^2(x_1, x_2, x_3)} = \|x_1\|_2 = \|x_2\|_2 = \|x_3\|_2\). The (second order) Voronoi region \(\mathcal{V}(x_1, x_2, x_3)\) has two boundaries, denoted by the rays \(L_1\) and \(L_2\). The angle between the \(x_1\) axis and the rays \(L_1, L_2\) are denoted by \(\alpha_1, \alpha_2\), respectively. The pairwise distances of \(x_1, x_2, x_3\) are denoted by \(\ell_1, \ell_2, \ell_3\).

\[
\begin{align*}
\Pr[&[g_1, g_2] \in [r, 0] + \mathcal{V}] \\
&= \Pr[g_1 \geq r, -a_2(g_1 - r) \leq g_2 \leq a_1(g_1 - r)] \\
&= \int_{r}^{\infty} \int_{-a_2(x_1 - r)}^{a_1(x_1 - r)} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right) dx_2 dx_1 \\
&\geq \int_{r}^{\infty} \int_{0}^{a_1(x_1 - r)} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right) dx_2 dx_1 \\
&= \int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_1^2}{2\sigma^2}\right) dx_1 \\
&= \left[1 - \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_1^2}{2\sigma^2}\right) dx_1\right] \\
&\times \int_{0}^{a_1(x_1 - r)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_2^2}{2\sigma^2}\right) dx_2 \times \frac{1}{2} \int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_2^2}{2\sigma^2}\right) dx_2 \times \frac{1}{2} \int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_2^2}{2\sigma^2}\right) dx_2.
\end{align*}
\]

In Equation (27), \(g_1\) and \(g_2\) are two independent Gaussians with mean zero and variance \(\sigma^2\). In Equations (28) and (29), we use (twice) the bound on the \(Q\)-function (Lemma 18).

We then proceed to estimate the integral in Equation (29).

\[
\begin{align*}
\int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_1^2 + a^2(x_1 - r)^2}{2\sigma^2}\right) dx_1 \\
&= \int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_1^2}{2\sigma^2}\right) dx_1 \\
&\times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{a^2(x_1 - r)^2}{2\sigma^2}\right) dx_1 \\
&\approx \frac{1}{12} \int_{r}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_1^2 + a^2(x_1 - r)^2}{2\sigma^2}\right) dx_1.
\end{align*}
\]

In Equation (27), \(g_1\) and \(g_2\) are two independent Gaussians with mean zero and variance \(\sigma^2\). In Equations (28) and (29), we use (twice) the bound on the \(Q\)-function (Lemma 18).

Continuing with Equation (29), we have

\[
P_{\mathcal{V}}(x_1, x_2, x_3) = P_{\mathcal{V}}(\mathcal{V}(x_1, x_2, x_3))
\]

Equation (30) follows again from Lemma 18.
in the first subfigure of Figure 9, the distance equals to the distance from \( x_1 \) to the ML V oronoi partition in Figure 6, the same bound also holds for \( P_{e,2}^{ML} (x_1, \{x_2, x_3\}) \) and \( P_{e,2}^{ML} (x_2, \{x_1, x_3\}) \). Therefore Theorem 6 holds in this case.

4) Case 2: The Largest Angle of the Triangle Spanned by \( x_1, x_2, x_3 \) Is Obtuse or Flat: One can similarly compute the (first order) V oronoi region and the second order V oronoi region induced by \( x_1, x_2, x_3 \) as depicted in the first and second figures of Figure 8, respectively. Note that in this case, the smallest ball containing all three points is different from the circumscribed ball. In fact, the former one only touches two points among three whereas the latter one by definition touches all three points and is larger than the former one. Note that the Chebyshev radius of the triangle is now equal to half of the length of the longest edge. In the example depicted in Figure 8, \( \text{rad}^2 (x_1, x_2, x_3) = \left( \frac{1}{2} \| x_3 - x_1 \|_2 \right)^2 \).

Following similar calculations as done in Section VIII-C3, we can estimate \( P_{e,2}^{ML} (x_i, \{x_1, x_2, x_3\}) \) for each \( i = 1, 2, 3 \). Note that, as depicted in Figure 9, the distance from \( x_2 \) to \( V(x_1, x_3) \) and the distance from \( x_3 \) to \( V(x_1, x_2) \) are both equal to \( \text{rad}^2 (x_1, x_2, x_3) \), and both \( V(x_1, x_3) \) and \( V(x_1, x_2) \) contain a full quadrant. Therefore the same calculations as those in Section VIII-C3 yield

\[
\begin{align*}
P_{e,2}^{ML} (x_2, \{x_1, x_2, x_3\}) &= \exp \left( -\frac{r^2}{2\sigma^2} - o(n) \right), \\
P_{e,2}^{ML} (x_3, \{x_1, x_2, x_3\}) &= \exp \left( -\frac{r^2}{2\sigma^2} - o(n) \right),
\end{align*}
\]

where \( r = \sqrt{\text{rad}^2 (x_1, x_2, x_3)} \). However, the distance from \( x_1 \) to \( V(x_2, x_3) \) is strictly larger than \( r \). To see this, we note that in the first subfigure of Figure 9, the distance equals \( \| x_1 \|_2 \) and \( \| x_2 \|_2 = \| x_3 \|_2 = \| x_1 \|_2 \), the later two quantities of which are obviously larger than the radius of the ball. Hence

\[
P_{e,2}^{ML} (x_1, \{x_1, x_2, x_3\}) = \exp \left( -\frac{d^2}{2\sigma^2} - o(n) \right) \ll \exp \left( -\frac{r^2}{2\sigma^2} - o(n) \right),
\]

where \( d := d_{12} (x_1, V(x_2, x_3)) = \| x_1 \|_2 > \sqrt{\text{rad}^2 (x_1, x_2, x_3)} \). Overall, Theorem 6 still holds in this case.

D. Proof of Theorem 6 for General \( L - 1 \geq 2 \)

We now prove Theorem 6 in the general case where \( L - 1 \geq 2 \). Let \( \mathcal{L} \subset \mathbb{R}^n \) be an arbitrary set of \( L \) distinct points in \( \mathbb{R}^n \). We assume that \( \mathcal{L} \) satisfies the property that the pairwise distances between points are bounded between \( \sqrt{\gamma n} \) and \( \sqrt{nC} \) for some arbitrary constants \( 0 < c < C \) independent of \( n \).

Let \( B_\mathcal{L} \) be the smallest ball containing \( \mathcal{L} \). It is clear that there must be a point in \( \mathcal{L} \) that lies on the boundary of \( B_\mathcal{L} \), otherwise \( B_\mathcal{L} \) can be shrunk yet still contains \( \mathcal{L} \), which violates the minimality of \( B_\mathcal{L} \). Let \( z_0 \) denote any such point on the boundary of \( B_\mathcal{L} \), as depicted in the first subfigure of Figure 10.

Since there are only \( L \) points in \( \mathcal{L} \), \( \dim (\text{aff} \{ \mathcal{L} \}) \leq L - 1 \). By translating \( \mathcal{L} \) such that \( \text{aff} (\mathcal{L}) \) becomes a subspace, we can therefore parameterize \( \mathbb{R}^n \) using the orthonormal basis of \( \text{aff} (\mathcal{L}) \) (with its extension to \( \mathbb{R}^n \)). Under this parameterization, for any \( \mathbf{x} \in \mathcal{L} \), we have \( \mathbf{x}(i) = 0 \) for all \( L - 1 \leq i \leq n \). In the analysis we will only work with vectors in \( \mathbb{R}^{L-1} \) which are obtained by restricting vectors in \( \mathbb{R}^n \) to the first \( L - 1 \) coordinates and stick with the same notation.

As mentioned in Equation (26), for an \( L \)-list \( \mathcal{L} \), the complement of the ML \((L - 1)\)-list-decoding region of \( z_0 \) is given by the order-(\( L - 1 \)) Voronoi region \( V_{\mathcal{L}, L-1} (\mathcal{L} \setminus \{z_0\}) \) of \( \mathcal{L} \setminus \{z_0\} \). For \( L - 1 > 2 \), the shape of \( V_{\mathcal{L}, L-1} (\mathcal{L} \setminus \{z_0\}) \) seems delicate. However, we manage to prove the following lemma (Lemma 7) which helps us estimate the probability that the a Gaussian noise brings \( z_0 \) to the ML \((L - 1)\)-list-decoding error region \( V_{\mathcal{L}, L-1} (\mathcal{L} \setminus \{z_0\}) \).

To state the lemma, we need the following set of definitions. Let \( z_0 \) be a point in \( \mathcal{L} \) that lies on the boundary of \( B_\mathcal{L} \). As argued above, such an \( z_0 \) must exist. Let \( O \) be the center of \( B_\mathcal{L} \). We also set \( O \) to be the origin of our coordinate system. Let \( d_{\text{min}} \) denote the minimum pairwise distance between points in \( \mathcal{L} \), and \( \alpha \) be such that \( \sin \alpha = \frac{d_{\text{min}}}{\sqrt{\gamma n}} \) (see the third subfigure of Figure 10). Note that under the assumptions in Theorem 6, it is guaranteed that \( \alpha \) is a constant.
of $n$).\footnote{To see this, it suffices to show $\sqrt{\text{rad}^2(L)} = O(\sqrt{n})$. Apparently, $\sqrt{\text{rad}^2(L)} \leq \sqrt{nC}$ since $L \subset B^n(\sqrt{nC})$. Also, $\sqrt{\text{rad}^2(L)} \geq \frac{1}{2}\sqrt{nC}$ which is tight for $L = 2$. Therefore $\sqrt{\text{rad}^2(L)} = O(\sqrt{n})$ and $\alpha = \sin^{-1}\left(\frac{d_{min}/2}{\sqrt{\text{rad}^2(L)}}\right)$ is on the order of $\Theta(\sqrt{n})$.}

Let $D \subset \mathbb{R}^{L-1}$ be the cone of angular radius $\alpha$ with apex at $O$ and axis along the direction of $-x_0$. The cone $D$ is depicted in Figure 10. With these parameters/objects at hand, we claim that $D$ is a subset of $V_{L-1}(L \setminus \{x_0\})$ (the latter of which, by the notational convention of this section, is also a subset of $\mathbb{R}^{L-1}$ obtained by projecting the original $n$-dimensional (order-$L-1$) Voronoi region to its first $L-1$ coordinates).

Lemma 7: Let $L \subset \mathbb{R}^n(\sqrt{nC})$ be a set of $L$ points with minimum pairwise distance at least $d_{min}$. Let $B_L$ be the smallest ball containing $L$. Let $D \subset \mathbb{R}^{L-1}$ be the $\langle L-1 \rangle$-dimensional cone of angular radius $\alpha = \sin^{-1}\left(\frac{d_{min}/2}{\sqrt{\text{rad}^2(L)}}\right)$ depicted in Figure 10, and $x_0 \in L$ be any point on the boundary of $B_L$. Then $D \subset V_{L-1}(L \setminus \{x_0\})$.

Proof: We first note that all points on the ray shooting from $O$ along the direction of $-x_0$ are in $V_{L-1}(L \setminus \{x_0\})$. To see this, take any point $y$ on that ray and draw a ball of radius $\|x_0 - y\|_2$ around $y$ (see the second subfigure of Figure 10). Then $B^{L-1}(O, \sqrt{\text{rad}^2(L)}) \subset B^{L-1}(y, \|x_0 - y\|_2)$ and they are tangent at $x_0$. Therefore $x_0$ is the unique furthest point to $y$ in $B_L$. That is, given $y$ on the ray, the ML ($L-1$)-list-decoder will not output $x_0$.

The above argument for the ray can be extended to hold for the cone $D$ given the $d_{min}$-minimum distance guarantee. Clearly, to show that $D$ is a subset of $V_{L-1}(L \setminus \{x_0\})$, it suffices to consider points on the boundary of $D$. Now take any point $y \neq O$ on the boundary of $D$. (The case $y = O$ was already handled in the above paragraph.) Again, draw the ball $B^{L-1}(y, \|x_0 - y\|_2)$ (see the third subfigure of Figure 10). It is not hard to see that there is no point from $L$ other than $x_0$ that is in $B_L \setminus B^{L-1}(y, \|x_0 - y\|_2)$, since by the $d_{min}$-

Fig. 9. Suppose that $x_1, x_2, x_3$ span an obtuse/flat triangle and the length of the longest edge is given by $\|x_2 - x_3\|_2$. The radius of the smallest ball containing $x_1, x_2, x_3$ is equal to $r = \frac{1}{2}\|x_2 - x_3\|_2 = \sqrt{\text{rad}^2(x_1, x_2, x_3)}$. Then the distance from $x_2$ to $V(x_1, x_3)$ and the distance from $x_3$ to $V(x_1, x_2)$ are both equal to $r$. However, the distance $d$ from $x_1$ to $V(x_2, x_3)$ is strictly larger than $r$.

Fig. 10. Suppose $L \subset \mathbb{R}^n$ is a set of $L$ points each having norm $O(\sqrt{n})$ and the minimum pairwise distance is on the order of $\Theta(\sqrt{n})$. Let $B_L$ be the smallest ball containing $L$. Then there must exist a point $x_0 \in L$ on the boundary of $B_L$. We show that the $(L-1)$-list-decoding error of $x_0$ under ML list-decoder is large. We do so by lower bounding the Gaussian measure of the ML $(L-1)$-list-decoding error region of $x_0$ by that of a cone $D$ of angular radius $\alpha$ for some constant $\alpha > 0$. Indeed, from the geometry of the second and third subfigures, we show in Lemma 7 that any received vector $y$ in $D$ will result in a list-decoding error under ML $(L-1)$-list-decoder.
minimum distance guarantee, \( B^{L-1}(x_0, d_{\min}) \cap \mathcal{L} = \{ x_0 \} \). Therefore, \( x_0 \) is the furthest point in \( \mathcal{L} \) from \( y \), and given \( y \), the ML \((L - 1)\)-list-decoder will not output \( x_0 \). This finishes the proof of the lemma.

1) Proof of Theorem 6: Provided Lemma 7, we are finally ready to estimate the probability of ML \((L - 1)\)-list-decoding error (Equation (26)). As before, let \( r := \sqrt{\text{rad}^2(\mathcal{L})} \). We work with polar coordinates. Let the apex of the cone \( D \) be the origin \( O \). Parameterize \( x_0 \) as \( \{-r, x_0\} \in \mathbb{R} \). For some \( y_0 \in S^{L-2} \),

\[
P_{e,L-1}(x_0, \mathcal{L}) = \Pr[\{x_0 + \mathbf{g} \in \mathcal{V}_{L,L-1}(\mathcal{L} \setminus \{ x_0 \})] 
\geq \Pr[\{x_0 + \mathbf{g} \in D \}]
\]

\[
= \int_{-x_0 + D} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\|\mathbf{g}\|^2_2}{2\sigma^2}\right) d\mathbf{g}_2
\]

\[
= \int_{-x_0 + D} \int_{0}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho^{L-2} \cdot |S^{L-2}| \cdot \mathbf{1}_{-x_0 + D}(\rho) d\rho d\mu(\mathbf{u})
\]

\[
= \int_{S^{L-2}} \int_{r}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho^{L-2} \cdot |S^{L-2}| \cdot |\rho^{-1}(-x_0 + D)| d\rho
\]

\[
= \int_{S^{L-2}} \int_{r}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho^{L-2} \cdot |S^{L-2}| \cdot |\rho^{-1}(-x_0 + D)| d\rho
\]

\[
= \int_{S^{L-2}} \int_{r}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho^{L-2} \cdot |S^{L-2}| \cdot |\rho^{-1}(-x_0 + D)| d\rho
\]

In Equation (31), we switch to polar coordinates using Lemma 19 where \( \mu(\cdot) \) denotes the uniform probability measure on \( S^{L-2} \). Equation (32) follows since \( \|\mathbf{u}\|^2_2 = 1 \) for \( \mathbf{u} \in S^{L-2} \) and the inner integral vanishes for any \( \rho \) such that \( \rho = \|\rho\|_2 \leq \|\mathbf{u}\|_2 = r \). In Equation (33), we interchange the inner and outer integrations. Equation (34) follows by noting that the inner integral is nothing but the normalized surface area of the cap obtained by taking the intersection of \( S^{L-2} \) and the (shuffled and rescaled) cone \( \rho^{-1}(-x_0 + D) \). Equation (35) follows from the fact that the \((L-2)\)-dimensional volume scales like \( \|\mathbf{u}\|^{L-2}_2 = \rho^{L-2} |S^{L-2}| \).

To proceed, we bound the volume of the cap by first computing its radius \( s = s(\rho, \alpha, r) \) as a function of \( \rho \) (and \( \alpha, r \) as well). The geometry is depicted in Figure 11.

By Pythagorean theorem, it is not hard to see that

\[
\left(\frac{s}{\tan \alpha} + r\right)^2 + s^2 = \rho^2.
\]

Solving \( s \), we get

\[
s = s(\rho, \alpha, r) = \frac{(\tan \alpha)\left(\sqrt{(1 + \tan^2 \alpha)\rho^2 - (\tan^2 \alpha)r^2} - r\right)}{\tan^2 \alpha + 1}
\]

Fig. 11. In the above figure, \(-x_0 + D\) is a cone of angular radius \( \alpha \), the apex of which is \( r \) away from the origin \( O \). To integrate using polar coordinates, for each radius \( \rho \geq r \), we need to compute the surface measure of the cap obtained by taking the intersection of \( S^{L-2}(\rho) \) and \(-x_0 + D\). It suffices to compute the radius \( s \) of the cap. This can be done by examining the elementary geometry depicted above.

\[
= \left(\sin \alpha\right)\left(\sqrt{\rho^2 - r^2 \sin^2 \alpha - r \cos \alpha}\right)
\]

Since the volume of an \((L-2)\)-dimensional cap is lower bounded by that of an \((L-2)\)-dimensional ball of the same radius, continuing with Equation (35), we have

\[
P_{e,L-1}(x_0, \mathcal{L}) \geq \int_{r}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \cdot |B^{L-2}(s(\rho, \alpha, r))| d\rho
\]

\[
= \left(2\pi\sigma^2\right)^{-(L-1)/2} \int_{r}^{\infty} \frac{1}{(2\pi\sigma^2)^{(L-1)/2}} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \cdot |S^{L-2}(\rho) \cap (-x_0 + D)| d\rho
\]

Define the following two functions

\[
f(t) := \frac{t^2}{2\sigma^2}, \quad g(t) := \left(\sqrt{t^2 - \sin^2 \alpha - \cos \alpha}\right)^{L-2}
\]

We note that \( f'(t) = t/\sigma^2 \) and in the domain \([1, \infty) \), \( f(t) \) attains its unique minimum \( \frac{1}{2\sigma^2} \) at \( t^* = 1 \). Furthermore, \( g(t^*) = g^{(1)}(t^*) = g^{(2)}(t^*) = \cdots = g^{(L-3)}(t^*) = 0 \) where \( g^{(k)}(t) \) denotes the \( k \)-th derivative of \( g(t) \). However, the \((L-2)\)-st derivative of \( g(t) \) does not vanish at \( t^* \) and in fact one can check that it equals

\[
g^{(L-2)}(t^*) = \frac{(L-2)!}{(t^2 - \sin^2 \alpha - \cos \alpha)^{L-2}}_{t = t^*} = (L - 2)! \frac{1}{\cos^2 \alpha}.
\]

Now, we apply Laplace’s method (Theorem 21) to compute the integral above (Equation (37)). As \( n \to \infty \), we have \( r \to \infty \) and therefore

\[
\int_{1}^{\infty} \exp\left(-\frac{r^2 t^2}{2\sigma^2}\right) \left(\sqrt{t^2 - \sin^2 \alpha - \cos \alpha}\right)^{L-2} dt
\]

\[
= \int_{1}^{\infty} \exp(-r^2 f(t)) g(t) dt
\]
\[
\begin{align*}
\lim_{n \to \infty} & \exp(-r^2 f(t^*)) \cdot \frac{g(L-2)}{(r^2 f(1)(t^*))^{L-1}} \\
&= \frac{\sigma^2 (L-1) (L-2)!}{(\cos^{-2}(\alpha))^{2L-1}} \exp\left(-\frac{r^2}{2\sigma^2}\right).
\end{align*}
\]

Putting this back to Equation (37), we have, as \(n \to \infty\)
\[
P_{\text{e},L-1}(\mathcal{E}_0, \mathcal{L}) \geq \left(\frac{\sigma^2}{2\pi}\right)^{(L-1)/2} V_{L-2}(\tan^{-2}(\alpha)) \times \frac{(L-2)!}{P^{L-1}} \exp\left(-\frac{r^2}{2\sigma^2}\right).
\]

Since \(\sigma, L, \alpha\) are all constants independent of \(n\), we have shown
\[
P_{\text{e},L-1}(\mathcal{E}_0, \mathcal{L}) \geq \frac{1}{L} P_{\text{e},L-1}(\mathcal{E}_0, \mathcal{L}) = \exp\left(-\frac{r^2}{2\sigma^2} - o(n)\right),
\]
as desired. This completes the proof of the theorem.

E. Putting Things Together: Lower Bound on \(C_{L-1}(P, N)\)

Lemma 5 and Theorem 6 imply the following corollary which gives a lower bound on the error probability of a code \(C\) in terms of the Chebyshev radius.

**Corollary 8:** Let \(P, \sigma > 0\) and \(L \in \mathbb{Z}_{\geq 2}\). For any code \(C \subset \mathcal{B}^n(\sqrt{n\overline{P}})\) of size \(M\) and minimum pairwise distance at least \(\sqrt{nC}\) for some constant \(c > 0\), there exists a subcode \(C' \subset C\) of size at least \(M' \geq M/2\) such that for all \(L \in (\mathcal{C}_L)\),
\[
P_{\text{e},L-1}(C) \geq \exp\left(-\frac{\rad^2(L)}{2\sigma^2} - o(n)\right).
\]

It therefore suffices to prove the existence of codes \(C\) with \(e^{nR}\) codewords for which \(P_{\text{e},L-1}(C)\) is sufficiently small. The following is a classical result that can be derived using the approach in [27] (also see [28]).

**Theorem 9:** Let \(P, \sigma > 0\) and \(L \in \mathbb{Z}_{\geq 2}\). Define the expurgated error exponent
\[
E_{\text{ex},L-1}(R) := -\min_{\rho \geq 0, \rho_2 \geq 1} R(L-1)\rho - \rho \left[\frac{sLP}{2} + \frac{L-1}{2} \ln \left(1 - 2sP + \frac{P}{\sigma^2 L^2}\right)\right].
\]

For any \(0 < R < \frac{1}{2} \ln(1 + P/\sigma^2)\), there exist codes \(C \subset \mathcal{S}^{n-1}(\sqrt{n\overline{P}})\) of rate \(R\) with minimum pairwise distance at least \(\sqrt{n\overline{P}e^{-3R}}\), that attain the following probability of error under ML \((L-1)\)-list decoding over an AWGN channel with input constraint \(P\) and noise variance \(\sigma^2\).
\[
P_{\text{e},L-1}(C) \leq \exp(-nE_{\text{ex},L-1}(R) + o(n)).
\]

**Proof:**

The above result is obtained by choosing \(2e^{nR}\) codewords independently and uniformly at random from \(\mathcal{S}^{n-1}(\sqrt{n\overline{P}})\), and then expurgating \(e^{nR}\) “bad” codewords.

For any \(m \in [2e^{nR}]\), let \(P_{\text{e},L-1}(m, C)\) denote the conditional probability of ML \((L-1)\)-list decoding error given that the \(m\)th codeword was transmitted. Using the standard approach for deriving expurgated exponents [27, Chapter 5], we can show that for every \(m \in [2e^{nR}]\),
\[
\Pr_{\mathcal{C}}\left[P_{\text{e},L-1}(m, C) > \exp(-nE_{\text{ex},L-1}(R) + o(n))\right] \leq 0.49
\]

For a given realization of \(\mathcal{C}\), let us define the following cost function
\[
\chi(\mathcal{C}) = \sum_{m=1}^{2e^{nR}} 1\{P_{\text{e},L-1}(m, C) > \exp(-nE_{\text{ex},L-1}(R) + o(n))\}
\]
\[
+ 2e^{nR} \mathbb{1}\{d_{\min}(C) \leq \sqrt{n\overline{P}e^{-3R}}\},
\]

where \(d_{\min}(C)\) denotes the minimum pairwise Euclidean distance between codewords in \(C\). Using Equation (39) and Lemma 5, we have
\[
\mathbb{E}[\chi(\mathcal{C})] \leq e^{n(R(1 + o(1)))}.
\]

This implies that for sufficiently large \(n\), there must exist at least one code with minimum pairwise distance \(\sqrt{n\overline{P}e^{-3R}}\) having at least \(e^{nR}\) codewords that satisfy \(P_{\text{e},L-1}(m, C) \leq \exp(-nE_{\text{ex},L-1}(R) + o(n))\). Discarding the remaining \(e^{nR}\) codewords gives us a code having the desired minimum distance and error exponent. This completes the proof.

**Theorem 9** gives an upper bound on the error exponent of a code satisfying a minimum distance constraint. This additional property turns out to be a mild condition and can be easily met. This requires us to show that a random code satisfies a minimum distance constraint of \(\Theta(\sqrt{n})\) with high probability, as justified in the following lemma.

**Lemma 10:** Let \(C\) be a random spherical code obtained by choosing \(e^{nR}\) codewords independently and uniformly at random from an \(n\)-dimensional sphere of radius \(\sqrt{n\overline{P}}\). Let \(\delta > 0\) and let \(\eta(C)\) denote the number of pairs of codewords in \(C\) for which the pairwise distance is less than \(e^{-3R}\). Then,
\[
\Pr\left[\eta(C) \geq 1\right] \leq e^{-n(R(1 - o(1)))}.
\]

**Proof:** Fix \(c = P e^{-6R}\). For any \(i, j \in [e^{nR}]\) with \(i \neq j\), let \(Y_{i,j}\) be the indicator random variable which takes the value 1 if the distance between \(i\)-th and \(j\)-th codeword is less than \(\sqrt{nC}\), and 0 otherwise. Since the codewords are uniformly distributed over \(\mathcal{S}^{n-1}(\sqrt{n\overline{P}})\), we see that \(\Pr[Y_{i,j} = 1] \leq (c/P)n^{(1/2 - o(1))}\). Also,
\[
\mathbb{E}[\eta(C)] = \mathbb{E}\left[\sum_{\{i,j\}\in([e^{nR}]_2)} Y_{i,j}\right] \leq e^{2nR} (c/P)n^{(1/2 - o(1))} \leq e^{-nR(1 - o(1))}.
\]

Using Markov’s inequality allows us to complete the proof.

In the derivation of Theorem 9, one can therefore choose a random code that achieves the desired error exponent and satisfies the minimum distance criterion, and then expurgate this. We can therefore assume that the expurgated codebook achieving Theorem 9 satisfies the \(\sqrt{nC}\)-minimum distance condition for any \(0 \leq c \leq P e^{-6R}\).

Now, combining Corollary 8 and Theorem 9, we get a code \(C \subset \mathcal{S}^{n-1}(\sqrt{n\overline{P}})\) of size \(M = e^{nR}\) which contains a subcode \(C' \subset C\) of size at least \(M/2\) satisfying: for every \(L \in (\mathcal{C}_L)\),
\[
\exp\left(-\frac{\rad^2(L)}{2\sigma^2} - o(n)\right) \leq P_{\text{e},L-1}(C) \leq \exp(-nE_{\text{ex},L-1}(R) + o(n)).
\]
Let us choose \( N = \frac{1}{n} \text{rad}_L^2(C') = \frac{1}{n} \min \text{rad}^2(L) \). Then, \( C' \) is guaranteed to be \((P, N, L - 1)\) list decodable. Therefore, by Equation (40),
\[
\frac{N}{2\sigma^2} \geq E_{\text{ex},L-1}(R) - o(1).
\tag{41}
\]
We then ignore the \( o(1) \) factor and optimize out the ancillary parameters \( s \) and \( \rho \) to get an explicit bound on \( R \) in terms of \( P, N, L \). To this end, let
\[
F(s, \rho) := R(L - 1) - \rho - \frac{1}{2} \ln(1 - 2sP) + \frac{L - 1}{2} \ln \left(1 - 2sP + \frac{P}{\sigma^2 L \rho}\right).
\tag{42}
\]
The critical point \( s \) and \( \rho \) in the minimization of \( E_{\text{ex},L-1}(R) \) satisfies:
\[
\frac{\partial}{\partial s} F(s, \rho) = -P \rho \left( L - 1 - \frac{1}{2} \left(\frac{1}{2} + \frac{P}{(\sigma^2 L \rho)}\right) \right) = 0,
\tag{43}
\]
\[
\frac{\partial}{\partial \rho} F(s, \rho) = R(L - 1) - sLP - \frac{1}{2} \ln(1 - 2sP) - \frac{L - 1}{2} \ln \left(1 - 2sP + \frac{P}{\sigma^2 L \rho}\right) - \frac{(L - 1)P}{2(P + L(1 - 2sP) \rho \sigma^2)} = 0.
\tag{44}
\]
From Equation (43), we have
\[
\rho = \frac{(L - 1) - 2LPs}{2L^2 s(1 - 2sP) \sigma^2}.
\tag{45}
\]
Substitute \( \rho \) into Equation (44), we have
\[
- \ln(1 - 2sP) + (L - 1) \left[ 2R - \ln \frac{(L - 1)(1 - 2sP)}{L(1 - 2sP) - 1} \right] = 0.
\tag{46}
\]
Solving \( R \) from Equation (46), we get
\[
R = \frac{1}{2} \left[ \ln \frac{(L - 1)(1 - 2sP)}{L(1 - 2sP) - 1} + \frac{1}{L - 1} \ln(1 - 2sP) \right].
\tag{47}
\]
Note that for Equation (47) to be valid, we need one additional constraint on \( s \), i.e., \( s < \frac{1 - 1/L}{2P} \) which implies \( L(1 - 2sP) - 1 > 0 \). Now, putting the expressions of the critical \( \rho \) (Equation (45)) and the critical \( R \) (Equation (47)) into \( F(s, \rho) \) (Equation (42)), we have
\[
F(s, \rho) = -\frac{P(L(1 - 2sP) - 1)}{2L^2 \sigma^2 (1 - 2sP)}. \tag{48}
\]
Substitute Equation (48) back to the relation between \( N \) and \( E_{\text{ex},L-1}(R) \) (Equation (41)), we have
\[
N \geq \frac{P(L(1 - 2sP) - 1)}{L(1 - 2sP)}. \tag{49}
\]
Note that there is no \( \sigma^2 \) in the above relation as it is cancelled out. Since the RHS increases as \( s \) decreases, to maximize the list-decoding radius \( N \), we need to take the minimum \( s \). Therefore, we take \( s \) that saturates Equation (49):
\[
s = \frac{(L - 1)P - LN}{2L(P - N)P^2}. \tag{50}
\]
Finally, putting \( s \) (Equation (50)) to the expression of \( R \) (Equation (47)), we get the desired bound
\[
R = \frac{1}{2} \left[ \ln \frac{(L - 1)P}{LN} + \frac{1}{L - 1} \ln \frac{P}{L(P - N)} \right]. \tag{51}
\]
This completes the proof.

Remark 6: As a sanity check, the critical value \( s \) given by Equation (50) is indeed nonnegative since \( N \) is less than the Plotkin point \( \frac{L - 1}{2P} \). Also, it is not hard to check that \( s < \frac{1 - 1/L}{2P} \). Putting the critical value of \( s \) (Equation (50)) into the expression of \( \rho \) (Equation (45)), we get
\[
\rho = \frac{N(P - N)}{(L(P-N)-P)\sigma^2}.
\]
We note that \( \rho \) is nonnegative for the same reason. Moreover, since \( \sigma^2 \) does not show up in the final bound on \( R \), one can take a sufficiently small \( \sigma^2 \) to make \( \rho \geq 1 \). In particular, it suffices to take \( 0 < \sigma \leq \sqrt{\frac{N(P-N)/P}{L(P-N)-P}} \). Finally, to double check, we note that the expurgated exponent given by Equation (38) is achievable if \( R \geq R_{s,L-1}(\text{snr}) \) where \( R_{s,L-1}(\text{snr}) \) is defined by Equation (91). Since \( R_{s,L-1}(\text{snr}) \) is increasing as \( \text{snr} = P/\sigma^2 \) increases, that is, as \( \sigma^2 \) decreases, the condition \( R_{s,L-1}(\text{snr}) \geq R \) can be satisfied if we take \( \sigma^2 \) to be sufficiently small so that \( R_{s,L-1}(\text{snr}) \) becomes larger than Equation (51). The exact threshold is given by
\[
\text{snr} \geq \frac{P(L(P-N)-P)}{N(P-N)},
\]
or
\[
\sigma \leq \sqrt{\frac{N(P-N)}{L(P-N)-P}}, \tag{52}
\]
which is the same as what we obtained before.

Remark 7: At a first glance, it may appear that the rate in Equation (51) is achieved by any \( \sigma \) satisfying Equation (52) above. It turns out that this is not true. The reason why \( \sigma^2 \) does not appear in the final expression is because we chose \( \rho \) to maximize \( R \). In this process, \( \sigma^2 \) was conveniently canceled out. However, a numerical evaluation of \( 2\sigma^2 E_{s,L-1}(R) \) reveals that this is in fact decreasing in \( \sigma^2 \), and the maximum is in fact achieved by taking \( \sigma^2 \rightarrow 0 \).

Remark 8: Instead of using Theorem 9, we could potentially have used the (relatively) more explicit expressions in Theorem 34 or [28] as a lower bound on the error exponent. However, these expressions for the error exponent involve parameters \( t \) in the case of Theorem 34, and \( \rho \) in the case of [28]) that are not easy to manipulate when trying to obtain a closed form expression for \( R \) in terms of \( P, N, L \). We therefore opted to use the less explicit form of the expurgated exponent that we found easier to work with.
F. Connections to [12]

The paper [12] originally built the connection between list-decoding radius and error exponent (Equation (23)) and used it to obtain the same bound (Equation (22)) as ours. The proof presented in the current paper uses the same high-level idea as that presented in [12], but we deviate in our approach towards characterizing the order-$(L-1)$ Voronoi regions.

There were a few minor inaccuracies in [12], but they do not affect the overall results in the paper. It was claimed that the ML $(L-1)$-list-decoding region of a codeword $x_i$ (denoted by $n_i$ in [12]) satisfies

$$n_i = \left \{ y \in \mathbb{R}^n : \forall j \in [M] \setminus \{ i \}, \| y - x_i \|_2 \leq \| y - x_j \|_2 \right \},$$

which is nothing but $\mathcal{V}_L(x_i)$ (see Equation (24)). The true decoding region should be associated with a list $L$ (instead of a single codeword) and is in fact a higher-order Voronoi region $\mathcal{V}_{L,L}(L)$ (see Equation (25)) that cannot be bounded by a first-order Voronoi region.

A similar claim was made for the extended decoding region $\zeta_i$. It was incorrectly stated that

$$\zeta_i = \left \{ y \in \mathbb{R}^n : \exists x' \in L \setminus \{ x_i \}, \| y - x_i \|_2 \leq \| y - x' \|_2 \right \}.$$ 

However, the rest of the proof only requires that the error region $\mathbb{R}^n \setminus \zeta_i$ is nonempty and can be expressed as an intersection of finitely many halfspaces in $\mathbb{R}^n$. The following alternative definition should therefore suffice for the rest of the proof.

$$\zeta_i = \left \{ y \in \mathbb{R}^n : \exists x' \in L \setminus \{ x_i \}, \| y - x_i \|_2 \leq \| y - x' \|_2 \right \},$$

which gives

$$\mathbb{R}^n \setminus \zeta_i = \left \{ y \in \mathbb{R}^n : \forall x' \in L \setminus \{ x_i \}, \| y - x_i \|_2 > \| y - x' \|_2 \right \}.$$ 

Note that $\mathbb{R}^n \setminus \zeta_i$ is nothing but $\mathcal{V}_{L,L-1}(L \setminus \{ x_i \})$ (see Equation (26)).

In the current paper, we instead study Equation (25) which is really the higher-order Voronoi region. This enables us to get a relationship between the Chebyshev radius of the code and the error exponent of the code for any $\sigma > 0$. This connection may be of independent interest.

Also, [12] takes $\sigma \rightarrow 0$ in order to obtain Equation (23). In our alternate approach, this step seems to be avoided since the $\sigma^2$ term conveniently cancels out. However, as pointed out earlier, it so happens that $2\sigma^2 n E_{ML}$ is decreasing in $\sigma^2$ in the parameter regime of interest although explicit maximization is bypassed because we chose the $\rho$ to maximize $R$.

In essence, we provide a new proof of the connection between list decoding error exponents for AWGN channels and the Chebyshev radius of a code (perhaps, in a more explicit form), and also derive new expressions for the list decoding error exponents for AWGN channels with and without power constraints.

G. Unbounded Packings: Lower Bound on $C_{L-1}(N)$

We now adapt the techniques developed above for unbounded packings. The two key ingredients are: (i) a lower bound on the list-decoding error probability in terms of the Chebyshev radius; (ii) an upper bound on the list-decoding error probability. For (ii), we have bounds in Theorem 15 on the list-decoding error exponent of AWGN channels without input constraints. Unfortunately, (i) which was proved for finite codebooks cannot directly be generalized to the setting of infinite codebooks. While Theorem 6 is valid for arbitrary countable codebooks, Theorem 5 is true only for finite codebooks. One approach is to derive list decoding error exponents for infinite constellations under maximum probability of error.

An easier approach is to consider a finite codebook $C_a$ of sufficiently large size but restricted to lie within $[-a/2,a/2]^n$ for a sufficiently large $a$. We construct an infinite constellation by tiling the codebook

$$C = C_a + a(1 + o(1))Z^n.$$

We then lower bound the Chebyshev radius of this infinite constellation $C$ with the list decoding error exponent of $C_a$ under maximum probability of error.

1) From Infinite Constellations to Finite Codebooks and Back: Consider any infinite constellation $C_\infty$ of rate $R$. Recall that

$$R = \limsup_{a \rightarrow \infty} \frac{1}{n} \ln \frac{|C_\infty \cap [-a/2,a/2]^n|}{a^n}.$$ 

Fix $a = n^2$. Then, there exists $x \in \mathbb{R}^n$ such that $|\{(C_\infty + x) \cap [-a/2,a/2]^n| \geq a^n e^{nR}$. Let us define the finite codebook

$$C_a := (C_\infty + x) \cap [-a/2,a/2]^n,$$ 

and the infinite constellation

$$C := C_a + a(1 + n^{-1.4})Z^n.$$ 

The above infinite constellation has rate $R(1 - \delta_n)$ where $\lim_{n \rightarrow \infty} \delta_n = 0$. Any two distinct shifts $C_a + \xi_1$ and $C_a + \xi_2$, where $\xi_1 \neq \xi_2 \in a(1 + n^{-1.4})Z^n$, are separated by a distance of at least $n^{0.6}$. This immediately implies the following.

**Lemma 11:** Let $C_a$ and $C$ be as defined in Equations (53) and (54), respectively. If $\text{rad}_2(C_a) = \Theta(\sqrt{n})$ (as per Definition 4), then

$$\text{rad}_2(C) = \inf_{C \in \mathcal{C}} \text{rad}_2(L).$$

**Proof:** Clearly,

$$\text{rad}_2(C_a) = \min_{C \in \mathcal{C}} \text{rad}_2(L) \geq \inf_{C \in \mathcal{C}} \text{rad}_2(L).$$

Consider any $L \subset C$. If $L \subset C_a + \xi$ for some $\xi \in a(1 + n^{-1.4})Z^n$, then $\text{rad}_2(L) \geq \inf_{C \in \mathcal{C}} \text{rad}_2(L) = \text{rad}_2(C_a)$. If not, then there are at least two points $\zeta_1, \zeta_2$ in $L$ such that $\zeta_1 \in C_a + \zeta_1$ and $\zeta_2 \in C_a + \zeta_2$ where $\zeta_2 \neq \zeta_1 \in a(1 + n^{-1.4})Z^n$. But this implies that $|\zeta_2 - \zeta_1| \geq n^{0.6}$ and $\text{rad}_2(L) \geq n^{0.6}/2$. This completes the proof. □

Let $\alpha \geq 1$ and $R = \frac{1}{2} \ln \frac{1}{\pi e \alpha^{1/2}}$. Or equivalently, $\alpha = \sqrt{\frac{2}{\pi e}}$. In Theorem 15, we prove lower bounds on the achievable expurgated list decoding error exponents of infinite constellations. This is obtained by choosing the codebook to be a Matérn point process derived from a Poisson point
process. This means that the average probability of error is upper bounded by \(\exp(-nE_{\text{ex},L-1}(\alpha) + o(n))\). In the proof, we choose the exclusion radius of the Matérn point process to be \(\Theta(\sqrt{n})\), which ensures that the minimum pairwise distance between codewords is \(\Omega(\sqrt{n})\).

Let us take \(C_{\infty}\) to be the Matérn point process above, \(C_a = C_{\infty} \cap [-a/2, a/2]^n\) for \(a = n^2\). Using standard tail bounds for PPPs,

\[
\Pr[|C_a| < a^n(e^{nR} - n^3)] \leq \exp(-\Theta(n^2)),
\]
or

\[
\Pr[|C_a| < a^n e^{nR(1-o(1))}] \leq \exp(-\Theta(n^2)).
\]

Therefore, with probability \(1 - e^{-\Theta(n^2)}\), the rate of \(C\) is

\[
R(C) \geq \frac{a^n e^{nR(1-o(1))}}{a^n(1 + n^{-1.4})^n} = e^{nR(1-o(1))}. \tag{55}
\]

Combining Equation (55) above with Lemma 11, we get that for every \(L \in \binom{\mathbb{C}}{L}\) with \(\text{rad}^2(L) = \Theta(\sqrt{n})\),

\[
\exp\left(-\frac{\text{rad}^2(L)}{2\sigma^2} - o(n)\right) \leq \exp(-nE_{\text{ex},L-1}(\alpha) + o(n)).
\]

Since the minimum pairwise distance is \(\Omega(\sqrt{n})\), we can conclude that \(\text{rad}^2(L) = \Omega(\sqrt{n})\). Let us choose

\[
N = \frac{1}{n} \min_{L \in \binom{\mathbb{C}}{L}} \text{rad}^2(L).
\]

Then (omitting the \(o(1)\) term),

\[
N \geq 2\sigma^2 E_{\text{ex},L-1}(\alpha) = 2\sigma^2 E_{\text{ex},L-1}\left(\sqrt{\frac{e^{-2R}}{2\pi e \sigma^2}}\right). \tag{56}
\]

It can be verified that the RHS as a function of \(\sigma\) is maximized at

\[
\sigma = \sqrt{\exp\left(-\frac{L}{L-1} \ln L - \ln(2\pi e) - 2R\right)}, \tag{57}
\]

which corresponds to

\[
\alpha = \sqrt{\frac{e^{-2R}}{2\pi e \sigma^2}} = \sqrt{\frac{L}{\sqrt{\pi} \sigma}} \in \left[\sqrt{L}, \sqrt{2L}\right].
\]

Substituting the critical \(\sigma\) (Equation (57)) into Equation (56), we get the following inequality relating \(N\) to \(R\):

\[
N \geq 2 \exp\left(-\frac{L}{L-1} \ln L - \ln(2\pi e) - 2R\right) \cdot E_{\text{ex},L-1}\left(\sqrt{\frac{L}{\sqrt{\pi}}}\right) = 2 \exp\left(-\frac{L}{L-1} \ln L - \ln(2\pi e) - 2R\right) \cdot \frac{L-1}{2} = (L-1) \exp\left(-\frac{\ln L}{L-1} + \ln(2\pi e L) + 2R\right).
\]

Solving \(R\), we get the following lower bound on the \((N, L-1)\)-list-decoding capacity:

\[
R = \frac{1}{2} \ln \frac{L-1}{2\pi e NL} - \frac{\ln L}{2(L-1)}.
\]

We summarize our finding in the following theorem.

**Theorem 12:** Let \(N > 0\) and \(L \in \mathbb{Z}_{\geq 2}\). The \((N, L-1)\)-list-decoding capacity \(C_{L-1}(N)\) is at least

\[
C_{L-1}(N) \geq \frac{1}{2} \ln \frac{L-1}{2\pi e NL} - \frac{\ln L}{2(L-1)}.
\]

**H. Remark on the \(\sigma^2\) That Maximizes the Chebyshev Radius**

To prove Theorems 3 and 12 for the bounded and unbounded cases, respectively, we combine Theorem 6 with bounds on error exponents. This combination then gives rise to an inequality relating \(N\) to \(R\). See Inequalities (41) (56) for the bounded and unbounded cases. In Inequalities (41), the variance \(\sigma^2\) of the Gaussian noise happens to cancel on both sides. To maximize \(R\), one then needs to take the largest possible error exponent which occurs in the expurgated regime \(R \leq R_{x,L-1}\) (the latter quantity is defined in Equation (91)). However, in Inequalities (56), the Gaussian variance \(\sigma^2\) does not cancel and one should optimize it out. It turns out that the optimal \(\sigma^2\) does not lie in the expurgated regime. Instead, one should use the error exponent in the “straight line” regime \(\sqrt{L} < \alpha \leq \sqrt{2L}\) (under the parameterization of Theorem 15).

Unfortunately we do not have intuition of this phenomenon.

**IX. List-Decoding Error Exponents of AWGN Channels Without Input Constraints**

In this section, we obtain bounds on the \((L-1)\)-list-decoding error exponent of an AWGN channel with no input constraint and noise variance \(\sigma^2\). An unbounded code for such a channel contains codewords whose norm can be arbitrarily large. The rate of such a code is measured by Equation (18).

**A. Random Coding Exponent**

**Theorem 13:** For any \(\sigma > 0, \alpha \geq 1\) and \(L \in \mathbb{Z}_{\geq 2}\), there exists an unbounded code \(C \subset \mathbb{R}^n\) of rate \(R = \frac{1}{L} \ln \frac{1}{\pi e \sigma^2 \alpha^2}\) such that when used over an AWGN channel with noise variance \(\sigma^2\) and no input constraint, the exponent of the average probability of \((L-1)\)-list-decoding error of \(C\) (normalized by \(\lim_{n \to \infty} -\frac{1}{n} \ln(\cdot)\)) is at least \(E_{\text{L},L-1}(\alpha)\) defined as

\[
E_{\text{L},L-1}(\alpha) = \left\{ \frac{\alpha^2}{L-1} - \frac{\ln L}{2} + (L-1) \ln \alpha, \quad 1 \leq \alpha \leq \sqrt{L} \right\}. \tag{18}
\]

**Proof:** Let \(\alpha \geq 1\) and \(R = \frac{1}{L} \ln \frac{1}{\pi e \sigma^2 \alpha^2}\). Let \(C \subset \mathbb{R}^n\) be a Poisson Point Process with intensity \(\lambda = e^{nR} = \langle 2\pi e \sigma^2 \alpha^2 \rangle^{-n/2}\). By translating \(C\), we assume without loss of generality that \(0 \in C\). By Item 1 of Fact 28, the distribution of the translated process remains the same.

Let \(e_{\text{L},L-1}^{\text{ML}}(C)\) denote the error event of \(C\) under ML \((L-1)\)-list-decoding given \(0\) is transmitted.

\[
e_{\text{L},L-1}^{\text{ML}}(C) := \{ \exists \{z_1, \cdots, z_{L-1}\}, \forall i \in [L-1], \quad \|z_i - g\|_2 < \|g\|_2 \}. \tag{18}
\]
For any instantiated $C$, we can bound the probability of $\mathcal{E}_{L-1}^{ML}(C)$ as follows.

$$
\Pr[\mathcal{E}_{L-1}^{ML}(C)] = \frac{1}{r} \Pr[\mathcal{E}_{L-1}^{ML}(C) \mid \|g\|_2 = r] = \int_0^\infty f_{\|g\|_2}(r) \Pr[\mathcal{E}_{L-1}^{ML}(C) \mid \|g\|_2 = r] dr
\leq \int_0^r f_{\|g\|_2}(r) \Pr[\mathcal{E}_{L-1}^{ML}(C) \mid \|g\|_2 = r] dr + \int_r^\infty f_{\|g\|_2}(r) dr.
$$

(58)

The function $f_{\|g\|_2}$ denotes the p.d.f. of the $\ell_2$-norm of a Gaussian vector $g \sim \mathcal{N}(0,\sigma^2 I_n)$. The randomness of the above probability and expectation comes from the Gaussian noise $g$. In Equation (58), $r^* > 0$ is to be specified.

Conditioned on $\emptyset \in C$ being transmitted, the rest of $C$ follows the Palm distribution denoted by $\mathbb{E}^{\text{Palm}}$ and $\Pr^{\text{Palm}}$. We now average Equation (58) over the PPP $C$. The second term is independent of $C$ and remains the same under averaging. As for the first term, we note that

$$
\Pr[\mathcal{E}_{L-1}^{ML}(C) \mid \|g\|_2 = r] = P\left\{ \exists \{x_1, \ldots, x_{L-1}\} \in (C \setminus \emptyset)_{L-1}, \forall i \in [L-1], \|x_i - g\|_2 < r \mid \|g\|_2 = r \right\}
\leq \min \left\{ \sum_{\{x_1, \ldots, x_{L-1}\} \in (C_0)_{L-1}} \mathbb{P}[\forall i \in [L-1], \|x_i - g\|_2 < r], 1 \right\}.
$$

(59)

The first term in Equation (58) then is at most

$$
\sum_{\{x_1, \ldots, x_{L-1}\} \in (C_0)_{L-1}} \int_0^r f_{\|g\|_2}(r) \mathbb{P}[\forall i \in [L-1], \|x_i - g\|_2 = r] dr,
$$

(60)

where we only used the first term of the minimization in Equation (59). Now, the first term in Equation (58) averaged over $C$ can be bounded as follows, (61)–(64), shown at the bottom of the next page. Equation (61) is by Slivnyak’s theorem (Theorem 30). Equation (62) follows from Equation (60). Equation (63) is by Campbell’s theorem (Theorem 29).

We choose $r^*$ such that the sum of Equation (64) and the second term in Equation (58) is minimized. That is, $r^*$ is a zero of the derivative (w.r.t. $r^*$) of the sum. Recall the way one takes derivative w.r.t. the limit of an integral. If

$$
F(x) = \int_a^x f(t)dt,
$$

then

$$
\frac{d}{dx} F(x) = f(x).
$$

Therefore, $r^*$ satisfies

$$
\chi^{L-1} V_n^{L-1} f_{\|g\|_2}(r^*) (r^*)^{n(L-1)} - \frac{1}{2} \|g\|_2 (r^*) = 0
\implies r^* = \chi^{L-1} V_n^{1/(n-1)}.
$$

By the choice of $\lambda$, we further have

$$
r^* = e^{-R} \left( \frac{2\pi e}{n} \right)^{-1/2} (1 + o(1))
\leq \exp \left( -\frac{1}{2} \ln \frac{1}{2\pi e\sigma^2\alpha^2} \right) (2\pi e)^{-1/2} \sqrt{n} (1 + o(1))
= \alpha^2 \sqrt{n} (1 + o(1)).
$$

(65)

Next, we evaluate the bound we got for the error probability

$$
\chi^{L-1} V_n^{L-1} \int_0^r f_{\|g\|_2}(r) r^{n(L-1)} dr + \Pr[\|g\|_2 > r^*].
$$

(66)

The density of the $\ell_2$-norm of a Gaussian vector of variance $\sigma^2$ is

$$
f_{\|g\|_2}(r) = \sigma^{-1} f(r/\sigma),
$$

(67)

where $f(\cdot)$ is the density of the $\ell_2$-norm $\|g_0\|_2$ of a standard Gaussian vector $g_0 \sim \mathcal{N}(0, I_n)$. Neglecting the $o(1)$ factor in $r^*$ (Equation (65)), we get that the first term of Equation (66) (dot) equals

$$
(2\pi e\sigma^2\alpha^2)^{-1/2} (\frac{2\pi e}{n})^{1/2} \frac{1}{2} \frac{1}{n} 
\times \int_0^{\alpha \sqrt{n}} \sigma^{-1} f(r/\sigma) r^{n(L-1)} dr
= \frac{\sigma^2 \alpha^2}{n} \frac{1}{2} \frac{1}{n} \int_0^{\alpha \sqrt{n}} \sigma^{-1} f(s \sqrt{n}) (s \sqrt{n})^{n(L-1)} \sigma \sqrt{n} ds
$$

(68)

In Equation (68), we let $s = \frac{r}{\sigma \sqrt{n}}$.

The following asymptotics of $f(\cdot)$ was obtained in [73, Eqn. (129)].

**Lemma 14** [73]: The p.d.f. $f(\cdot)$ of the $\ell_2$-norm of an $n$-dimensional standard Gaussian vector satisfies the following pointwise estimate:

$$
f(s \sqrt{n}) = \exp \left( -n \left( \frac{s^2}{2} - \ln s - \frac{1}{2} \right) + o(n) \right),
$$

for any $s \geq 0$.

By Lemma 14, the first term of Equation (66) dot equals

$$
\int_0^{\alpha} \exp \left( -n \left( \frac{s^2}{2} - \ln s - \frac{1}{2} \right) \right) ds
- \left( L - 1 \right) \ln s + \left( L - 1 \right) \ln \alpha \right) \right) ds
= \int_0^{\alpha} \exp \left( -n \left( \frac{s^2}{2} - \ln s - \frac{1}{2} + (L - 1) \ln \alpha \right) \right) ds
$$

(69)

where we have suppressed the polynomial factor $\sqrt{n}$. To evaluate the integral in Equation (69), we will apply the Laplace’s
method (Theorem 20). It is easy to check that the function
\[ F(s) := s^2 - L \ln s - \frac{1}{2} + (L - 1) \ln \alpha \]
is decreasing in \( s \in [0, \sqrt{L}] \) and is increasing in \( s \in [\sqrt{L}, \infty] \).

If \( \alpha \leq \sqrt{L} \), the minimum value of \( F(s) \) in \([0, \alpha]\) is achieved at \( s = \alpha \). By Theorem 20, the integral in Equation (69) dot
\[
\exp\left(-n \left[ \frac{\alpha^2}{2} - L \ln \alpha - \frac{1}{2} + (L - 1) \ln \alpha \right] \right)
\]
equals
\[
\exp\left(-n \left[ \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2} \right] \right).
\]
\[
(70)
\]
If \( \alpha > \sqrt{L} \), the minimum value of \( F(s) \) in \([0, \alpha]\) is achieved at \( s = \sqrt{L} \). By Theorem 20, the integral in Equation (69) dot
\[
\exp\left(-n \left[ \frac{L}{2} - \frac{L}{2} \ln L - \frac{1}{2} + (L - 1) \ln \alpha \right] \right)
\]
equals
\[
\exp\left(-n \left[ \frac{L}{2} - \frac{L}{2} \ln L + (L - 1) \ln \alpha \right] \right).
\]
\[
(71)
\]
Let \( E_1(\alpha, L) \) and \( E_2(\alpha, L) \) be the normalized first-order exponent of the first and second term in Equation (66), respectively, i.e.,
\[
E_1(\alpha, L) := \lim_{n \to \infty} \frac{1}{n} \ln \left( \lambda^{L-1} V_n^{L-1} \int_0^{r^*} f(\|g\|_2) r^{n(L-1)} dr \right),
\]
\[
E_2(\alpha, L) := -\lim_{n \to \infty} \frac{1}{n} \ln \Pr[\|g\|_2 > r^*].
\]
By Equations (70) and (71), \( E_1(\alpha, L) \) is given by
\[
E_1(\alpha, L) = \begin{cases} \alpha - \frac{1}{2}, & \alpha \leq \sqrt{L} \\ \frac{L-1}{2} - \frac{1}{2} L + (L - 1) \ln \alpha, & \alpha > \sqrt{L}. \end{cases}
\]
Let \( C := \frac{1}{2} \ln \frac{1}{2 \pi e \sigma^2} \). Note that \( R = \frac{1}{2} \ln \frac{1}{2 \pi e \sigma^2} = C - \ln \alpha \).
The exponent \( E_2(\alpha, L) \) is the large deviation exponent of the tail of a chi-square random variable which is given by Fact 26.
In fact, it was shown in [72, Eqn. (29)] and [71] that, under the choice of \( r^* \) given by Equation (65), we have
\[
E_2(\alpha, L) = \frac{1}{2} e^{(C-R) - 1 - 2(R-C)}
\]
\[
= \frac{1}{2} e^{2\ln \alpha - 1 - 2 \ln \alpha} = \frac{1}{2} (\alpha^2 - 2 \ln \alpha)
\]
\[
= \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}
\]
Note that \( E_2(\alpha, L) \) coincides with \( E_1(\alpha, L) \) for \( 1 \leq \alpha \leq \sqrt{L} \) whereas it strictly dominates \( E_1(\alpha, L) \) when \( \alpha > \sqrt{L} \). Finally,
\[
-\lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}_{\mathcal{C}}[\Pr[\mathcal{C}_{L-1}(\mathcal{C})|\mathcal{C}]]
\]
\[
\geq -\lim_{n \to \infty} \frac{1}{n} \ln \text{Equation(66)}
\]
\[
\geq \min\{E_1(\alpha, L), E_2(\alpha, L)\}
\]
The bound on error exponent proved in the last section (Section IX-A) can be improved using the expurgation technique when the rate is sufficiently low. In this section, we prove the following theorem.

Theorem 15: For any $\sigma > 0$, $\alpha \geq 1$ and $L \in \mathbb{Z}_{\geq 2}$, there exists an unbounded code $C \subset \mathbb{R}^n$ of rate $R = \frac{1}{2} \ln \frac{2}{\pi e \sigma^2}$ such that when used over an AWGN channel with noise variance $\sigma^2$ and no input constraint, the exponent of the average probability of $(L - 1)$-list decoding error of $C$ (normalized by $\lim_{n \to \infty} -\frac{1}{n} \ln (\cdot)$) is at least $E_{ex,L-1}(\alpha)$ defined as

$$E_{ex,L-1}(\alpha) = \begin{cases} \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}, & 1 \leq \alpha \leq \sqrt{L} \\ \frac{L-1}{2} - \frac{L}{2} \ln L + (L - 1) \ln \alpha, & \alpha > \sqrt{L}. \end{cases}$$

where

$$F(\alpha, L) := \frac{\alpha^2}{16} + \frac{1}{16} \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} - \frac{L-1}{2} \ln \left( \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} - \alpha^2 + 4 \right) + \frac{L-1}{2} \ln \left( \sqrt{\alpha^4 + 8\alpha^2(2L - 3) + 16} + \alpha^2 + 4 \right) + \frac{3}{2} \ln \frac{\alpha}{2} - \frac{1}{2}.$$}

Furthermore, $\|z - z'\|_2 \geq \alpha \sigma (1 - o(1)) \sqrt{n}$ for any $z \neq z' \in C$.

Proof: Let $\alpha \geq 1$ and $R = \frac{1}{2} \ln \frac{1}{2 \pi e \sigma^2}$. Let $C \subset \mathbb{R}^n$ be a Matérn process obtained from a PPP with intensity $\lambda = e^{-\sigma^2 R} = (2\pi e \sigma^2)^{-n/2}$ and exclusion radius $\xi := \tilde{\alpha} \sigma \sqrt{n}$ where $\tilde{\alpha} := \alpha (1 - \varepsilon_n)$ for a proper choice of $\varepsilon_n \overset{\mathcal{D}}{\to} 0$ to be specified momentarily. The intensity $\lambda'$ of the Matérn process is $\lambda' = \lambda \exp(-\lambda |B^n(\xi)|) = \lambda \exp\left(-\lambda \nu(\alpha (1 - \varepsilon_n) \sigma \sqrt{n})^n\right)$

\[
\approx \lambda \exp\left(-\left(2\pi e \sigma^2 \alpha^2\right)^{-n/2}\frac{1}{\sqrt{n}}\right)^{n/2} \left((\alpha^2(1 - \varepsilon_n)^2 \sigma^2 n)^{n/2}\right) = \lambda \exp\left(-\frac{1}{\sqrt{n}}\right).
\]

Taking $\varepsilon_n = \ln n = o(1)$, we have

$$\lambda' \approx \lambda \exp\left(-\frac{\ln n}{\sqrt{n}}\right) = \lambda \exp\left(-\frac{1}{\sqrt{n}}\right) \approx \lambda.$$\]

In the following analysis, we will ignore the $o(1)$ factor $\varepsilon_n$ and assume for simplicity $\tilde{\alpha} = \alpha$.

Suppose $0 \in C$. Under the Palm distribution, the order-$\lfloor L - 1 \rfloor$ factorial moment $\lambda'((x_1, \ldots, x_{L-1}))$ of $C$ can be bounded as follows

$$\lambda'((x_1, \ldots, x_{L-1})) \leq \lambda^{L-1} \prod_{i=1}^{L-1} \mathbb{E}^C_{\mathbb{R}^n} \{x_i \in B^n(\xi) \cap C\}.$$

Following similar arguments to those in Section IX-A, we have

$$\Pr[|\mathcal{E}_{L-1}(C)| = \int_0^{\infty} f_{\|g\|_2}(r) \Pr\left[\mathcal{E}_{L-1}(C) \cap \|g\|_2 = r\right] dr.$$\]

The above identity holds for any instantiated $C \subset \mathbb{R}^n$ and the randomness in the probability comes from the channel noise $g$. Averaging the RHS of the above equation over the Matérn process $C$, we have, (74)–(76), shown at the bottom of the next page, where $\varepsilon = [1, 0, \cdots, 0] \in \mathbb{R}^n$. In Equation (74), we skipped several steps which are similar to Equation (61), Equation (62) and Equation (63). In particular, we used Slivnyak’s theorem (Theorem 30), the first bound of the minimum in Equation (59), Campbell’s theorem (Theorem 29) and the bound on the (Palm) intensity of Matérn processes (Equation (73)). In Equation (76), we take the direction of $g$ to be $\varepsilon$ since the integral in Equation (75) does not depend on the direction of $g$.

Incorporating the second term of the minimum in Equation (59), we get

$$\mathbb{E}\left[\int_0^{\infty} f_{\|g\|_2}(r) \Pr\left[\mathcal{E}_{L-1}(C) \cap \|g\|_2 = r\right] dr\right] \leq \int_0^{\infty} f_{\|g\|_2}(r) \min\left\{\lambda^{L-1} |B^n(\varepsilon_{\xi}, r) \cap B^n(\xi) |^{L-1}, 1\right\} dr.$$\]

We apply the relation Equation (67), change variable $s = \frac{r}{\sigma \sqrt{n}}$ and get

$$\mathbb{E}\left[\int_0^{\infty} f_{\|g\|_2}(r) \Pr\left[\mathcal{E}_{L-1}(C) \cap \|g\|_2 = r\right] dr\right] \leq \int_0^{\infty} \sigma^{-1} f(s) \min\left\{\lambda^{L-1} |B^n(\varepsilon_{\xi}, s) \cap B^n(\xi) |^{L-1}, 1\right\} ds.$$\]

The following upper bound on $|B^n(\varepsilon_{\xi}, r) \cap B^n(\alpha \sigma \sqrt{n}) |$ was shown in [73, Eqn. (106)].

Lemma 16 [73]: Let $\xi = [1, 0, \cdots, 0] \in \mathbb{R}^n$, $\alpha \geq 1$, $\sigma > 0$. Then for any $r > 0$,

$$|B^n(\varepsilon_{\xi}, r) \cap B^n(\alpha \sigma \sqrt{n}) | \leq |B^n(c(s) \sigma \sqrt{n}) | = V_n(c(s) \sigma \sqrt{n})^n,$$

where

$$c(s) = \begin{cases} 0, & 0 < s \leq \alpha/2 \\ \sqrt{s^2 - (s - \frac{\alpha}{2})^2}, & \alpha/2 < s \leq \alpha/\sqrt{2} \\ s, & s > \alpha/\sqrt{2}. \end{cases}$$
Using Lemma 16, continuing with Equation (77), we have

\[
\mathbb{E}_C[\Pr[\mathcal{E}^{ML}_L(C)|C]] \\
\leq \int_0^\infty f(s\sqrt{n}) \min \left\{ 2^{L-1} V_{n-1}^{-1}(c(s)\sigma \sqrt{n})^{n(L-1)}, 1 \right\} ds \\
= \int_0^\infty f(s\sqrt{n}) \min \left\{ (2\pi e\sigma^2 s^2)^{-\frac{1}{2}} n^{-\frac{1}{2}} (2\pi e)^{-1} \cdot (c(s)^2 \sigma^2 n)^{\frac{3}{2n(L-1)}} \right\} ds \\
= \int_0^\infty f(s\sqrt{n}) \min \left\{ \alpha^{-n(L-1)} - c(s)n^{-1}(1, 1) \right\} ds \\
\leq \int_0^\infty \exp \left( -n \left( \frac{s^2}{2} - \ln s - \frac{1}{2} \right) + (L-1) \ln \alpha - \ln(c(s))^+ \right) ds \\
= \int_0^{\alpha/2} \exp(-nF_1(s))ds + \int_{\alpha/2}^{\alpha/\sqrt{2}} \exp(-nF_2(s))ds + \int_{\alpha/\sqrt{2}}^\infty \exp(-nF_3(s))ds, \\
(78)
\]

where \( F_1(s), F_2(s), F_3(s) \) are defined as follows:

\[
F_1(s) := \frac{s^2}{2} - \ln s - \frac{1}{2} + (L-1)(\ln \alpha - \ln 0) \\
= \infty, \quad 0 \leq s \leq \alpha/2; \\
F_2(s) := \frac{s^2}{2} - \ln s - \frac{1}{2} \\
+ (L-1) \left[ \ln \alpha - \frac{1}{2} \ln \left( s^2 - \left( s - \frac{\alpha^2}{2s} \right)^2 \right) \right]^+ \\
= \frac{s^2}{2} - \ln s - \frac{1}{2} + (L-1) \left[ \ln \alpha \\
- \frac{1}{2} \ln \left( s^2 - \left( s - \frac{\alpha^2}{2s} \right)^2 \right) \right], \quad \alpha/2 < s \leq \alpha/\sqrt{2}; \\
F_3(s) := \frac{s^2}{2} - \ln s - \frac{1}{2} + (L-1)(\ln \alpha - \ln s)^+ \
\]

For \( F_2(s) \), we can remove the function \([\cdot]^+\) since the function \( f_2(s) := \sqrt{s^2 - (s - \frac{\alpha^2}{2s})^2} \) attains its maximum value \( \alpha/\sqrt{2} \) at \( s = \alpha/\sqrt{2} \). Therefore \( \ln \alpha - \ln f_2(s) \geq \ln \sqrt{2} > 0 \).

Define

\[
E_1(\alpha, L) := -\lim_{n \to \infty} \frac{1}{n} \ln \int_0^{\alpha/2} \exp(-nF_1(s))ds, \\
E_2(\alpha, L) := -\lim_{n \to \infty} \frac{1}{n} \ln \int_{\alpha/2}^{\alpha/\sqrt{2}} \exp(-nF_2(s))ds, \\
E_3(\alpha, L) := -\lim_{n \to \infty} \frac{1}{n} \ln \int_{\alpha/\sqrt{2}}^\infty \exp(-nF_3(s))ds.
\]

We compute \( E_1(\alpha, L), E_2(\alpha, L), E_3(\alpha, L) \) using Laplace’s method (Theorem 20).

For \( E_1(\alpha, L) \), we have

\[
E_1(\alpha, L) = \min_{s \in (0, \alpha/2]} F_1(s) = \infty.
\]

For \( E_2(\alpha, L), F_2(s) \) has a unique stationary point

\[
s_0 = \sqrt{\frac{\alpha^2 + 3\sqrt{\alpha^2 + 8\alpha^2(2L-3) + 16} + 4}{8}}.
\]

One can check that \( s_0 \geq \sqrt{2}/\alpha \) if \( \alpha \leq \sqrt{2L} \) and \( s_0 < \sqrt{2}/\alpha \) if \( \alpha > \sqrt{2L} \). Therefore

\[
E_2(\alpha, L) = \min_{s \in (\alpha/2, \alpha/\sqrt{2}]} F_2(s) = \begin{cases} 2(\sqrt{2}/\alpha), & \alpha \leq \sqrt{2L} \\ F_2(s_0), & \alpha > \sqrt{2L}, \end{cases}
\]

where

\[
F_2(\sqrt{2}/\alpha) = \frac{\alpha^2}{4} + \ln \alpha + \frac{L}{2} \ln 2 - \frac{1}{2}; \\
F_2(s_0) = \frac{\alpha^2}{16} + \frac{1}{16} \sqrt{\alpha^2 + 8\alpha^2(2L-3) + 16} - \frac{L-1}{2} \ln \left( \sqrt{\alpha^2 + 8\alpha^2(2L-3) + 16} - \alpha^2 + 4 \right).
\]
$$\lim_{L \to \infty} \frac{L-2}{2} \ln \left( \sqrt{\alpha^4 + 8\alpha^2(2L-3) + 16 + \alpha^2 + 4} \right) + \frac{3}{2} \ln 2 - \frac{1}{4}. \quad (79)$$

For $E_3(\alpha, L)$, we let

$$F_{3,1}(s) \triangleq \frac{s^2}{2} - \ln s - \frac{1}{2} + (L-1)(\ln \alpha - \ln s),$$
$$F_{3,2}(s) \triangleq \frac{s^2}{2} - \ln s - \frac{1}{2}.$$

The function $F_{3,1}(s)$ has a unique minimum point $s = \sqrt{L}$. Therefore, for $s \in (\alpha/\sqrt{2}, \alpha]$, the minimum value of $F_{3,1}(s)$ is, (80), shown at the bottom of the page.

The function $F_{3,2}(s)$ has a unique minimum point $s = 1 \leq \alpha$. Therefore, for $s \in (\alpha, \infty)$, the minimum value of $F_{3,2}(s)$ is

$$\min_{s \in (\alpha, \infty)} F_{3,2}(s) = F_{3,2}(1) = \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}. \quad (81)$$

One can easily check that Equation (81) is at least Equation (80) for any $\alpha \geq 1$. Therefore,

$$E_3(\alpha, L) = \min_{s \in (\alpha/\sqrt{2}, \infty)} F_3(s)$$
$$= \min \left\{ \min_{s \in (\alpha/\sqrt{2}, \alpha]} F_{3,1}(s), \min_{s \in (\alpha, \infty)} F_{3,2}(s) \right\}$$
$$= \min_{s \in (\alpha/\sqrt{2}, \alpha]} F_{3,1}(s).$$

Finally,

$$- \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[C_{\text{ML}}(L-1)(C|C)]$$
$$\geq \min \{ E_3(\alpha, L), E_2(\alpha, L), E_3(\alpha, L) \}$$
$$= \min \left\{ E_2(\alpha, L), E_3(\alpha, L) \right\}$$
$$= \min \left\{ \begin{array}{ll}
F_2(\alpha/\sqrt{2}), & \quad 1 \leq \alpha \leq \sqrt{L} \\
F_2(\alpha/\sqrt{2}), & \quad \sqrt{L} < \alpha \leq 2\sqrt{L} \\
F_2(s_0), & \quad \alpha > 2\sqrt{L}
\end{array} \right\}$$
$$\geq F_3(\alpha), \quad 1 \leq \alpha \leq \sqrt{L}$$
$$= F_3(\sqrt{L}), \quad \sqrt{L} < \alpha \leq 2\sqrt{L}$$
$$= F_2(s_0), \quad \alpha > 2\sqrt{L}$$
$$= \frac{\alpha^2}{2} - \frac{1}{2} - \ln \alpha - \frac{1}{2}, \quad 1 \leq \alpha \leq \sqrt{L}$$
$$= \frac{L-1}{2} - \frac{1}{2} \ln L + (L-1) \ln \alpha, \quad \sqrt{L} < \alpha \leq 2\sqrt{L}$$
$$= F_2(s_0), \quad \alpha > 2\sqrt{L}.$$  

Recall that the quantity $F_2(s_0)$ was defined in Equation (79). This finishes the proof.

**C. List-Decoding Error Exponents vs. Unique-Decoding Error Exponents**

Our results on list-decoding error exponents of AWGN channels without input constraints recover those by Polytyrev [71, Theorem 3] for unique-decoding. Indeed, when $L = 2$, Equation (72) specializes to

$$E_{ex,L=1}(\alpha) = \begin{cases} 
\frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}, & 1 \leq \alpha \leq \sqrt{2} \\
\frac{1}{2} - \ln 2 + \ln \alpha, & \sqrt{2} \leq \alpha \leq 2 \sqrt{2L} \\
\frac{\alpha^2}{2}, & \alpha > 2L.
\end{cases} \quad (82)$$

As in the case of power-constrained AWGN channels, list-decoding (with constant list sizes) for input unconstrained AWGN channels does not increase the capacity and moreover does not increase the error exponent for any $1 \leq \alpha \leq \sqrt{2}$. However, for any $\alpha > \sqrt{2}$, list-decoding does increase the error exponent. Furthermore, the critical values of $\alpha$ move from $\sqrt{2}$ and 2 to $\sqrt{2\alpha}$ and $\sqrt{2\alpha L}$, respectively, under list-decoding.

We plot Polytyrev’s exponents and our exponents (for $L = 3$) in Figure 5.

**X. OPEN QUESTIONS**

The problem of packing spheres in $\ell_p$ space was also addressed in the literatures [75], [76], [77], and [78]. Recently, there was an exponential improvement on the optimal packing density in $\ell_p$ space [79] relying on the Kabatiansky–Levenshtein bound [10]. It is worth exploring the $\ell_p$ version of the multiple packing problem.

Our lower bound is proved via a very interesting connection to error exponents. We do not know how to directly analyze the tail probability of the Chebyshev radius, even for Gaussian codes. One can view it as the tail of the maximum of a certain Gaussian process. This looks like a proper venue where the chaining method [80] is applicable. However, it seems unlikely that one can extract a meaningful exponent using the generic chaining machinery. Note that for the purpose of maximizing the rate, we do care about the exact exponent, not only an exponentially decaying bound.

For large $L$, our results imply that the list sizes must scale as $O(\frac{1}{2} \ln \frac{1}{2})$ for rates that are $\varepsilon$-close to capacity. The same can be obtained using different approaches [24]. An interesting open question is to resolve whether this is indeed the best possible scaling as a function of $\varepsilon$.

**APPENDIX A**

**COLLECTION OF USEFUL RESULTS**

In this section, we collect some known results that are used in various proofs.

**See also [73, Eqn. (108)] for a parameterization of Polytyrev’s bound using $\alpha$.**

$$\min_{s \in (\alpha/\sqrt{2}, \alpha]} F_{3,1}(s) = \begin{cases} 
F_3(\alpha) = \frac{\alpha^2}{2} - \ln \alpha - \frac{1}{2}, & 1 \leq \alpha \leq \sqrt{2} \\
F_3(\sqrt{L}) = \frac{L-1}{2} - \frac{1}{2} \ln L + (L-1) \ln \alpha, & \sqrt{L} \leq \alpha \leq \sqrt{2\alpha} \\
F_3(\alpha/\sqrt{2}) = \frac{\alpha^2}{4} - \ln \alpha + \frac{1}{2} \ln 2 - \frac{1}{2}, & \alpha > \sqrt{2\alpha L}.
\end{cases} \quad (80)$$
Definition 5 (Gamma function): For any \( z \in \mathbb{C} \) with \( \Re(z) > 0 \), the Gamma function \( \Gamma(z) \) is defined as
\[
\Gamma(z) := \int_0^\infty v^{z-1} e^{-v} \, dv.
\]

Lemma 17 (Markov): If \( x \) is a nonnegative random variable, then for any \( a > 0 \), \( \Pr[x \geq a] \leq E[x]/a \).

Definition 6 (Q-function): The Q-function is defined as
\[
Q(x) := \Pr[N(0, 1) > x] = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} \, dt.
\]

Lemma 18: For any \( x > 0 \),
\[
Q(x) = \frac{1}{12} e^{-x^2/2} (1 + e^{-\Omega(x)}).
\]
As a direct corollary, for any \( x > 0 \),
\[
\Pr[N(0, \sigma^2) > x] = Q(x/\sigma) = \frac{1}{12} e^{-\frac{x^2}{2\sigma^2}} (1 + e^{-\Omega(x)}).
\]

Lemma 19 (Integration in polar coordinates): For any integrable function \( f: \mathbb{R}^n \to \mathbb{R} \), we have
\[
\int_{\mathbb{R}^n} f(x) \, dx = \int_{S^{n-1}} \int_0^\infty f(r\theta) r^{n-1} |S^{n-1}| \, dr \, d\mu(\theta)
\]
where \( \mu \) is the uniform probability measure on \( S^{n-1} \), i.e., for \( A \subset S^{n-1} \), \( \mu(A) := \frac{|A|}{|S^{n-1}|} \).

Theorem 20 (Laplace’s Method): Let \( a < b \in \mathbb{R} \) and \( f, g: \mathbb{R} \to \mathbb{R} \).
1) If \( t^* \in (a, b) \) is the unique minimum point of \( f \) in \([a, b]\) such that \( f'(t^*) = 0 \), \( f''(t^*) > 0 \), \( g(t^*) \neq 0 \), then
\[
\int_a^b g(t) e^{-Mf(t)} \, dt \sim \frac{2\pi}{Mf''(t^*)}.
\]
2) If \( a \) is the unique minimum point of \( f \) in \([a, b]\) such that \( f'(a) = 0 \), \( f''(a) > 0 \), \( g(a) \neq 0 \), then
\[
\int_a^b g(t) e^{-Mf(t)} \, dt \sim \frac{\pi}{2Mf''(a)}.
\]
3) If \( a \) is the unique minimum point of \( f \) in \([a, b]\) such that \( f'(a) > 0 \), \( g(a) \neq 0 \), then
\[
\int_a^b g(t) e^{-Mf(t)} \, dt \sim \frac{2\pi}{Mf''(a)}.
\]

Theorem 21 (Laplace’s Method): Let \( a \in \mathbb{R} \) and \( L \geq 2 \) be an integer. Suppose \( g: \mathbb{R} \to \mathbb{R} \) satisfies \( g(a) = g'(1)(a) = g''(2)(a) = \cdots = g^{(L-1)}(a) = 0 \) and \( g^{(L)}(a) \neq 0 \) where \( g^{(i)} \) denotes the \( i \)-th derivative of \( g \). Suppose \( f: \mathbb{R} \to \mathbb{R} \) attains its unique minimum at \( a \) in the interval \([a, \infty)\) and \( f^{(1)}(a) > 0 \). Then we have
\[
\int_a^\infty g(t) e^{-Mf(t)} \, dt \sim e^{-Mf(a)} \frac{g^{(L-1)}(a)}{(Mf^{(1)}(a))^L}.
\]
Proof: The proof follows closely that of the standard Laplace’s formula and we only present a sketch of the former.\(^9\) The deviation is two-fold: (i) the function \( g \) is degenerate at a higher order; (ii) the extreme point \( a \) of \( f \) is on the boundary of the integration domain and is not a stationary point.

\[
\int_a^\infty g(t) e^{-Mf(t)} \, dt \\
\approx e^{-Mf(a)} \int_a^{a+\varepsilon} g(t) e^{-M(f(t)-f(a))} \, dt \\
\approx e^{-Mf(a)} \int_a^{a+\varepsilon} \left[ g(a) + g^{(1)}(a)(t-a) + \frac{g^{(2)}(a)}{2} (t-a)^2 \right] \, dt \\
+ \cdots + g^{(L-2)}(a) (L-2)! (t-a)^{L-2} + g^{(L-1)}(a) (L-1)! (t-a)^{L-1} \\
\times e^{-M\left[(f(a)+f^{(1)}(a)(t-a))-f(a)\right]} \, dt \\
(83)
\]
\[
eq e^{-Mf(a)} \int_a^{a+\varepsilon} g^{(L-1)}(a) (L-1)! (t-a)^{L-1} e^{-Mf^{(1)}(a)(t-a)} \, dt \\
(84)
\]

In Equation (83), we take the \((L-1)\)-st Taylor polynomial of \( g \) at \( a \) and the first Taylor polynomial of \( f \) at \( a \). Equation (84) follows since, by the assumption, the first \( L-1 \) terms of the Taylor polynomial of \( g \) vanish at \( a \). In Equation (85), we let \( u = Mf^{(1)}(a)(t-a) \). Equation (86) follows from the definition of Gamma function (Definition 5) and Equation (87) is because the Gamma function coincides with the factorial function at positive integer points.

Theorem 22 (Cramér): Let \( \{X_i\}_{i=1}^n \) be a sequence of i.i.d. real-valued random variables. Let \( s_n := \frac{1}{n} \sum_{i=1}^n x_i \). Then for any closed \( \mathcal{F} \subset \mathbb{R} \),
\[
\lim_{n \to \infty} \sup_{x \in \mathcal{F}} \Pr[s_n \in \mathcal{F}] \leq -\inf_{x \in \mathcal{F}} \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \ln E[e^{\lambda x}] \};
\]
and for any open \( \mathcal{G} \subset \mathbb{R} \),
\[
\lim_{n \to \infty} \sup_{x \in \mathcal{G}} \Pr[s_n \in \mathcal{G}] \geq -\inf_{x \in \mathcal{G}} \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \ln E[e^{\lambda x}] \}.
\]
Furthermore, when \( \mathcal{F} \) or \( \mathcal{G} \) corresponds to the upper (resp. lower) tail of \( s_n \), the maximizer \( \lambda \geq 0 \) (resp. \( \lambda \leq 0 \)).

Lemma 23 (Gaussian Integral): Let \( a > 0 \) and \( b, c \in \mathbb{R} \). We have
\[
\int_{\mathbb{R}} e^{-ax^2+bx+c} \, dx = \sqrt{\frac{\pi}{a}} \cdot e^{\frac{b^2}{4a}+c}.
\]
Lemma 24 (Gaussian Integral): Let \( A \in \mathbb{R}^{n \times n} \) be a positive-definite matrix. Then
\[
\int_{\mathbb{R}^n} \exp(-x^T A x) \, dx = \sqrt{\frac{\pi^n}{\det(A)}}.
\]

Definition 7: The chi-square distribution \( \chi^2(k) \) with degree of freedom \( k \) is defined as the distribution of \( \sum_{i=1}^k g_i^2 \) where \( g_i \sim N(0,1) \) for \( 1 \leq i \leq k \).

Fact 25: If \( x \sim \chi^2(k) \), then for \( \lambda < 1/2 \),
\[
\mathbb{E}[e^{\lambda x}] = \sqrt{1 - 2\lambda^k}.
\]

Plugging the formula in Fact 25 into Cramér's theorem (Theorem 22), we get the first order asymptotics of the tail of a chi-square random variable.

Lemma 26: If \( x \sim \chi^2(k) \), then for \( \delta > 0 \),
\[
\lim_{k \to \infty} \frac{1}{k} \ln \Pr[x > (1+\delta)k] = \frac{1}{2}(-\delta + \ln(1+\delta));
\]
for \( \delta \in (0,1) \),
\[
\lim_{k \to \infty} \frac{1}{k} \ln \Pr[x < (1-\delta)k] = \frac{1}{2}(\delta + \ln(1-\delta)).
\]

1) Poisson Point Processes: We use the following standard results on Poisson Point Processes. See [81] for a reference.

Definition 8 (PPP): A homogeneous Poisson Point Process (PPP) \( \mathcal{C} \) in \( \mathbb{R}^n \) with intensity \( \lambda > 0 \) is a point process satisfying the following two conditions.

1) For any bounded Borel set \( B \subset \mathbb{R}^n \), \( \mathcal{C} \cap B \sim \text{Pois}(\lambda |B|) \), that is,
\[
\Pr[|\mathcal{C} \cap B| = k] = e^{-\lambda |B|} \frac{(\lambda |B|)^k}{k!}
\]
for any \( k \in \mathbb{Z}_{\geq 0} \).

2) For any \( \ell \in \mathbb{Z}_{\geq 2} \) and any collection of \( \ell \) disjoint bounded Borel sets \( B_1, \ldots, B_\ell \subset \mathbb{R}^n \), the random variables \( |\mathcal{C} \cap B_1|, \ldots, |\mathcal{C} \cap B_\ell| \) are independent, that is,
\[
\Pr[|\mathcal{C} \cap B_i| = k_i] = \prod_{i=1}^\ell e^{-\lambda |B_i|} \frac{(\lambda |B_i|)^{k_i}}{k_i!}
\]
for any \( k_1, \ldots, k_\ell \in \mathbb{Z}_{\geq 0} \).

Remark 9: All PPPs in this paper will be homogeneous, that is, the intensity is a constant and does not depend on the location of a point.

Definition 9 (Intensity and Factorial Moment Measure): Let \( \mathcal{C} \) be a point process in \( \mathbb{R}^n \). The intensity measure \( \Lambda(\cdot) \) induced by \( \mathcal{C} \) is defined as the measure on \( \mathbb{R}^n \) satisfying \( \Lambda(B) = \mathbb{E}[|\mathcal{C} \cap B|] \) for any Borel set \( B \subset \mathbb{R}^n \). The intensity field (a.k.a. intensity for short) \( \lambda(\cdot) \) is the density of \( \Lambda(\cdot) \) (whenever exists), i.e.,
\[
\Lambda(B) = \int_B \lambda(z) \, dz.
\]

More generally, for any \( L \geq 1 \), the \( L \)-th factorial moment measure \( \Lambda^{(L)}(\cdot) \) induced by \( \mathcal{C} \) is defined as the measure on \( (\mathbb{R}^n)^L \) satisfying
\[
\Lambda^{(L)}(B_1 \times \cdots \times B_L) = \mathbb{E} \left[ \sum_{(x_1, \ldots, x_L) \in C^L} \prod_{i=1}^L \mathbb{1}_{B_i}(x_i) \right],
\]
for any \( L \)-tuple of Borel sets \( B_1, \ldots, B_L \) in \( \mathbb{R}^n \) (not necessarily disjoint). The \( L \)-th factorial moment density \( \lambda^{(L)}(\cdot, \cdot, \cdot) \) is the density of \( \Lambda^{(L)}(\cdot) \) (whenever exists):
\[
\Lambda^{(L)}(B_1 \times \cdots \times B_L) = \int_{B_1} \cdots \int_{B_L} \lambda^{(L)}(z_1, \ldots, z_L) \, dz_1 \cdots dz_L.
\]
Note that the first factorial moment measure/density coincides with the intensity measure/field.

Fact 27: For a homogeneous PPP with intensity \( \lambda \), the \( L \)-th factorial moment measure \( \Lambda^{(L)}(\cdot) \) is given by
\[
\Lambda^{(L)}(B_1 \times \cdots \times B_L) = \lambda^L \prod_{i=1}^n |B_i|
\]
and the \( L \)-th factorial moment density \( \lambda^{(L)}(\cdot, \cdot, \cdot) \) is given by \( \lambda^{(L)}(z_1, \ldots, z_L) = \lambda^L \).

Fact 28: A homogeneous PPP \( \mathcal{C} \) satisfies the following properties.

1) A homogeneous PPP is stationary, i.e., invariant to translation.

2) A homogeneous PPP is isotropic, i.e., invariant to rotation.

3) For any box \( Q := \prod_{i=1}^n (a_i, b_i) \) where \( a_i \leq b_i \) for all \( i \in [n] \), the points in \( \mathcal{C} \cap Q \) are independent and uniformly distributed in \( Q \), that is, the \( i \)-th \( (i \in [n]) \) coordinate of any vector in \( \mathcal{C} \cap Q \) is uniformly distributed in \( (a_i, b_i) \) and is independent of any other coordinates (in or not in the same vector).

Definition 10 (Matérn Process): A Matérn process \( C' \) in \( \mathbb{R}^n \) with exclusion radius \( r > 0 \) can be obtained from a PPP \( C \) in \( \mathbb{R}^n \) with intensity \( \lambda \) by removing all pairs of points in \( C \) with distance at most \( r \). The intensity \( \lambda' \) of the resulting Matérn process \( C' \) is given by \( \lambda' = \lambda e^{-\lambda |B^e(r)|} \).

Theorem 29 (Campbell): For any \( L \in \mathbb{Z}_{\geq 1} \), any point process \( \mathcal{C} \) on \( \mathbb{R}^n \) with \( L \)-th factorial moment measure \( \Lambda^{(L)}(\cdot) \) and any measurable function \( f : (\mathbb{R}^n)^L \to \mathbb{R} \), the following equation holds
\[
\mathbb{E} \left[ \sum_{(x_1, \ldots, x_L) \in C^L} f(x_1, \ldots, x_L) \right] = \int_{(\mathbb{R}^n)^L} f(z_1, \ldots, z_L) \Lambda^{(L)}(dz_1 \times \cdots \times dz_L).
\]
If \( \Lambda^{(L)}(\cdot) \) has a density \( \lambda^{(L)}(\cdot, \cdot, \cdot) \), then the equation becomes
\[
\mathbb{E} \left[ \sum_{(x_1, \ldots, x_L) \in C^L} f(x_1, \ldots, x_L) \right] = \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} f(z_1, \ldots, z_L) \lambda^{(L)}(dz_1 \times \cdots \times dz_L).
\]

Theorem 30 (Slivnyak): Conditioned on a point (WLOG the origin, by Item 1 of Fact 28) in a homogeneous PPP, the distribution of the rest of the PPP (which is called the Palm distribution) is equal to that of the original PPP.
APPENDIX B
LIST DECODING ERROR EXPONENTS FOR POWER-CONSTRAINED AWGN CHANNELS

In this section, we list/derive lower bounds on the achievable error exponents for power-constrained AWGN channels. The techniques used are fairly standard [27], hence we only provide sketch proofs in most cases.

Theorem 31: Let $L \in \mathbb{Z}_{\geq 2}$. Consider any memoryless channel $W_{Y|x} \in \Delta(\mathbb{R}^n)$ over the reals with input constraints

$$\mathcal{P}_f := \left\{ P_x \in \Delta(\mathbb{R}) : \int_{\mathbb{R}} P_x(x)f(x)dx \leq 0 \right\} \subset \Delta(\mathbb{R})$$

for some cost function $f: \mathbb{R} \rightarrow \mathbb{R}$. Then there exists a code $\mathcal{C} \subset \mathbb{R}^n$ of rate $R$, satisfying the input constraints $\mathcal{P}_f$ and $P_{e,avg,L-1}(\mathcal{C}) \leq$

$$\min_{P_x \in \mathcal{P}_f} \inf_{\delta > 0} \min_{s \geq 0, \rho \geq 1} e^{nR(L-1)\rho} \left( \frac{e^{s\delta}}{Z} \right)^{1+(L-1)\rho} \left[ \int_{\mathbb{R}} \left( \int_{\mathbb{R}} P_x(x)e^{s\delta}(x)W_{Y|x}(y|x)^{1+(L-1)\rho}dx \right) dy \right]^n,$$

where

$$Z := \int_{\mathbb{R}} P^\infty_x(x)1\left\{ \sum_{i=1}^n f(x(i)) \leq \delta, 0 \right\} dx. \quad (89)$$

For the same channel, there also exists a code $\mathcal{C} \subset \mathbb{R}^n$ of rate $R$, satisfying the input constraints $\mathcal{P}_f$ and

$$P_{e,avg,L-1}(\mathcal{C}) \leq \min_{P_x \in \mathcal{P}_f} \inf_{\delta > 0} \min_{s \geq 0, \rho \geq 1} 2^{L\rho}e^{nR(L-1)\rho} \left( \frac{e^{s\delta}}{Z} \right)^{L\rho} \left[ \sum_{(x_0, \ldots, x_{L-1}) \in \mathcal{X}^L} e^{s\delta \sum_{k=0}^{L-1} f(x_k)} \prod_{k=0}^{L-1} P_x(x_k) \right] \times \left[ \sum_{y \in \mathcal{Y}} \prod_{i=0}^{L-1} W_{Y|x}(y|x_i)^{1/L} \right]^{1/\rho \delta \rho^n}.$$

where $Z$ is defined in the same way as in Equation (89). Solving the optimization problem in Equation (88) and Equation (90) leads to the derivation of the random coding error exponent and the expurgated error exponent, respectively.

A. Random Coding Exponent for AWGN Channels With Input Constraints

Theorem 31 gives non-explicit upper bounds on the probability of error. In this section, we derive explicit lower bounds on Equation (89) in the case of AWGN channels with input constraint $P$ and noise variance $\sigma^2$ under $(L-1)$-list-decoding. We prove the following theorem.

Theorem 32: Let $P, \sigma > 0$ and $L \in \mathbb{Z}_{\geq 2}$. There exist codes of rate $R$ for the AWGN channel with input constraint $P$ and noise variance $\sigma^2$ such that the rate satisfies $0 \leq R \leq \frac{1}{2} \ln(1+P/\sigma^2)$ and the exponent $E_{L-1}(R, P/\sigma^2)$ of the probability of error (normalized by $\lim_{n \rightarrow \infty} -\frac{1}{n} \ln(\cdot)$) under $(L-1)$-list-decoding is bounded as follows.

Let $s := P/\sigma^2$ and

$$R_{e,L-1}(snr) := \frac{1}{2} \left( \ln \frac{L^2 + snr^2 - 2snr(L - 2) + L + snr}{2L} + \frac{1}{L-1} \ln \frac{L^2 + snr^2 - 2snr(L - 2) + L - snr}{2L} \right), \quad (91)$$

$$R_{crit,L-1}(snr) := \frac{1}{2} \ln \left( 1 + \frac{snr}{2L} \right) + \frac{1}{2} \sqrt{\frac{2(L - 2) - 2(L - 2)snr + snr^2}{L^2}}. \quad (92)$$

1) If $R_{crit,L-1}(snr) \leq R \leq \frac{1}{2} \ln(1+snr)$, then

$$E_{L-1}(R, snr) \geq \frac{1}{2} \ln \left[ e^{2R - \frac{snr(e^{2R} - 1)}{2}} \left( \frac{1+\frac{4e^{2R}}{snr(e^{2R} - 1)}}{1 - \frac{4e^{2R}}{snr(e^{2R} - 1)}} \right) \right]. \quad (93)$$

2) If $0 \leq R \leq R_{crit,L-1}(snr)$, then

$$E_{L-1}(R, snr) \geq -R(L-1) + \frac{L-1}{2} \ln \left( L + \ln \left[ \sqrt{L - snr^2} + 4snr \right] \right) + \frac{1}{2} \ln \left( L - snr + \sqrt{(L - snr)^2 + 4snr^2} \right) + \frac{L}{4} \left( L + snr - \sqrt{(L - snr)^2 + 4snr^2} \right) - \frac{L}{2} \ln(2L). \quad (94)$$

We would like to remark that although random coding and expurgated error exponents are well studied in the literature, we are not aware of a reference with the above closed form expression for the random coding exponent.

For an AWGN channel with input constraint $P$ and noise variance $\sigma^2$, the channel transition kernel is given by

$$W_{Y|x}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-x)^2}{2\sigma^2}}, \quad (95)$$

and the cost function is given by

$$f(x) = x^2 - P. \quad (96)$$

Let $P_x$ be the Gaussian density with variance $P$:

$$P_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}. \quad (97)$$

For a constant $\delta > 0$, we claim that the factor $e^{s\delta}/Z^{1+(L-1)\rho}$ scales like poly($n$) for asymptotically large $n$ and therefore does not effectively contribute to the exponent. Indeed, the following lemma holds.

Lemma 33: Let $P, R, \sigma, \delta > 0$ be constants. Let $P_x$ be the Gaussian density with variance $P$ as defined in Equation (97). Let $f(x) := x^2 - P$. Let $Z$ be defined by Equation (89). Then

$$Z \approx \frac{\delta}{2^P \sqrt{\pi e}}.$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Proof: The proof follows from the central limit theorem.

\[
Z = \int_{\mathbb{R}^n} P^{\otimes n} \chi \left\{ -\delta \leq \sum_{i=1}^{n} (x(i)^2 - P) \leq 0 \right\} \, dx
\]

\[
= \Pr \left[ -\delta \leq P \left( \sum_{i=1}^{n} N_i(0,1)^2 - n \right) \leq 0 \right]
\]

\[
= \Pr \left[ -\frac{\delta}{\sqrt{2n}} \leq \frac{\chi(n) - n}{\sqrt{2n}} \leq 0 \right]
\]

\[
\overset{n \to \infty}{\longrightarrow} \Pr \left[ -\frac{\delta}{P \sqrt{2n}} \leq N(0,1) \leq 0 \right]
\]

\[
\overset{n \to \infty}{\longrightarrow} \frac{1}{P \sqrt{2n}} \cdot \frac{1}{\sqrt{2\pi}}
\]

Equation (98) follows since \( \frac{\chi(n) - n}{\sqrt{2n}} \) converges to \( N(0,1) \) in distribution as \( n \to \infty \). Equation (99) follows since the Gaussian measure of a thin interval \( \left[ -\frac{\delta}{P \sqrt{2n}}, 0 \right] \) is essentially the area of a rectangle with width \( \frac{\delta}{P \sqrt{2n}} \) and height \( P \chi(n) = 1/\sqrt{2\pi} \) for asymptotically large \( n \).

We are now ready to evaluate the random coding bound on the probability of the \((L-1)\)-list-decoding error of AWGN channels with input constraint \( P \) and noise variance \( \sigma^2 \).

Proof: [Proof of Theorem 32] The exponent (i.e., the probability of error normalized by \( -\frac{1}{n} \ln(\cdot) \)) given by Equation (88) specializes to

\[-R(L-1)\rho - \ln \left( \int_{\mathbb{R}} \left( \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(y-x)^2}{2\sigma^2}} \right)^{1+(L-1)\rho} \, dy \right).
\]

For notational convenience, let \( \gamma := 1 + (L-1)\rho \). We first compute the inner integral

\[I(y) := \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi} \sigma^2} \frac{1}{\sqrt{2\pi} \sigma^2} \gamma \cdot e^{s(x^2-P)} \cdot e^{\frac{y-x}{\sigma^2}} \cdot e^{\frac{(y-x)^2}{2\sigma^2}} \, dx \cdot dy.
\]

By Lemma 23, the above integral \( I(y) \) equals

\[A \cdot \frac{1}{\sqrt{2\pi} \sigma^2} \frac{1}{\sqrt{2\pi} \sigma^2} \gamma \cdot \exp \left( \frac{y^2}{2\sigma^2} \right) \cdot \exp \left( \frac{-s}{\sigma^2} \right).
\]

With this, the random coding exponent becomes

\[E(s,\gamma) := -R(L-1)\rho - \ln \left( \int_{\mathbb{R}} \frac{2\pi P \sigma^2 \gamma}{\sigma^2(1-2sP) + P} \cdot \frac{1}{\sqrt{2\pi} \sigma^2} \cdot \frac{1}{\sqrt{2\pi} \sigma^2} \gamma \cdot e^{s(x^2-P)} \cdot e^{\frac{y-x}{\sigma^2}} \cdot e^{\frac{(y-x)^2}{2\sigma^2}} \, dx \cdot dy \right).
\]

For the above bound to be valid, we need \( s < \frac{1}{2\rho} \).

Recall that \( s \geq 0, \rho \in [0,1] \) and \( \gamma = 1 + (L-1)\rho \). We need to maximize \( E(s,\gamma) \) in the region \( s \in [0,1/(2P)], \gamma \in [1,L] \).

To this end, we compute the stationary \( s \) and \( \gamma \).

\[
\frac{\partial}{\partial s} E(s,\gamma) = 0
\]

and

\[
\frac{\partial}{\partial \gamma} E(s,\gamma) = 0.
\]
Let \( \text{snr} \coloneqq P/\sigma^2 \) denote the signal-to-noise ratio (SNR). Solving \( s \) from Equation (101), we get
\[
s = \frac{1}{4P} \left( 1 + \frac{\text{snr}}{\gamma} - \frac{1}{\gamma} \sqrt{(\gamma - \text{snr})^2 + 4\text{snr}} \right). \tag{103}
\]

One can easily check that \( s \geq 0 \) provided \( \gamma \geq 1 \). Furthermore, \( s < \frac{1}{2P} \).

Putting Equation (103) into Equation (102) and solving \( \gamma \) therein, we get
\[
\gamma = \frac{\text{snr}}{2e^2P} \left( 1 + \sqrt{1 + \frac{4\text{snr}}{\text{snr}(e^{2P} - 1)}} \right). \tag{104}
\]

It can be easily verified that \( \gamma \geq 1 \) for any \( R \leq \frac{1}{2} \ln(1 + \text{snr}) \).

Suppose \( \gamma \leq L \). Then the minimum value of \( E(s, \gamma) \) is indeed achieved at the above \( \gamma \) given by Equation (104). Note that the condition \( \gamma \leq L \) is equivalent to
\[
R \geq \frac{1}{2} \ln \left( \frac{1}{2} + \frac{\text{snr}}{2L} + \frac{1}{2} \sqrt{1 - \frac{2(L - 2)}{L^2} \text{snr} + \frac{\text{snr}^2}{L^2}} \right). \tag{105}
\]

Substituting the stationary \( \gamma \) (Equation (104)) back to Equation (103), we get the stationary \( s \) as a function of only \( \text{snr} \) and \( R \). Note that here \( s \) and \( \gamma \) do not depend on \( L \). Therefore the calculations in this case coincide with those for unique-decoding case as done in [27, Theorem 7.4.4] and we omit the details. Putting both \( s \) and \( \gamma \) into Equation (100), we finally get the random coding exponent
\[
\min_{s \in [0,1/(2P)], \gamma \in [1,L]} E(s, \gamma) = \frac{1}{2} \ln \left[ e^{2R} - \frac{\text{snr}(e^{2P} - 1)}{2} \left( 1 + \frac{4e^{2P}}{\text{snr}(e^{2P} - 1)} \right) \right] + \frac{\text{snr}}{4e^{2P}} \left( e^{2R} + 1 - (e^{2P} - 1) \right) \sqrt{1 + \frac{4e^{2P}}{\text{snr}(e^{2P} - 1)}}.
\]

This proves Item 1 in Theorem 32.

On the other hand, if the \( \gamma \) given by Equation (104) is larger than \( L \), i.e., Equation (105) holds in the reverse direction, then the minimum value of \( E(s, \gamma) \) is achieved at \( \gamma = L \). In this case, \( s \) given by Equation (103) becomes
\[
s = \frac{1}{4P} \left( 1 + \frac{\text{snr}}{L} - \frac{1}{L} \sqrt{(L - \text{snr})^2 + 4\text{snr}} \right), \tag{106}
\]

and the minimum value of \( E(s, \gamma) \) is achieved at \( \gamma = L \) and the \( s \) given by Equation (106):
\[
\min_{s \in [0,1/(2P)], \gamma \in [1,L]} E(s, \gamma) = -R(L - 1) + \frac{L - 1}{2} \ln \left( L + \text{snr} + \sqrt{(L - \text{snr})^2 + 4\text{snr}} \right) + \frac{1}{2} \ln \left( L + \text{snr} + \sqrt{(L - \text{snr})^2 + 4\text{snr}} \right) + \frac{1}{2} \ln(2L).
\]

This proves Item 2 in Theorem 32. \qed

### B. Expurgated Exponent for AWGN Channels With Input Constraints

We proceed to evaluate the expurgated exponent (\( \frac{1}{2} \ln \) of Equation (90)) in the case of AWGN channels with input constraint \( P \) and noise variance \( \sigma^2 \) under \((L-1)\)-list-decoding.

We prove the following theorem.

**Theorem 34**: Let \( P, \sigma > 0 \) and \( L \in \mathbb{Z}_{\geq 2} \). Consider an AWGN channel with input constraint \( P \) and noise variance \( \sigma^2 \). Let \( \text{snr} = P/\sigma^2 \). Let \( R_{L-1}(\text{snr}) \) be defined by Equation (91). Then there exist codes of rate \( 0 \leq R \leq R_{L-1}(\text{snr}) \) for the above channel such that the exponent \( E_{L-1}(R, \text{snr}) \) of the probability of error (normalized by \( \lim_{n \to \infty} -\frac{1}{n} \ln() \)) under \((L - 1)\)-list-decoding is bounded as follows:
\[
E_{L-1}(R, \text{snr}) \geq \frac{\text{snr}(L - 1)}{2L}, \tag{107}
\]
where \( t \) is the unique solution of \((L - 1)e^{2R} = (L - 1)t \frac{e^{2R}}{2L} \) in \( t \in [1/L, 1] \).

**Proof**: For the channel of interest, the channel transition kernel \( W_{y|x} \), the cost function \( f \) and the input distribution \( P_x \) are given by Equations (96) and (95), respectively. For a constant \( \delta > 0 \), by Lemma 33, the factor \( 2L^2 \left( \frac{e^{2R}}{2L} \right)^{LP} \) is subexponential in \( n \) and does not play a role in the exponent. Therefore, the exponent of Equation (90) specializes to
\[
R(L - 1)p + p \ln \left( \int \frac{1}{\sqrt{2\pi\sigma^2}} e^\frac{X_0 - \frac{L-1}{2p} \rho}{\sigma^2} dx \right)^{1/p} d(x_0, \ldots, x_{L-1}). \tag{108}
\]

The inner integral w.r.t. \( y \) in Equation (108) is a Gaussian integral and can be computed as follows using Lemma 23.
\[
\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=0}^{L-1} (y_i - x_i)^2 \right) dy = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp \left( -\frac{1}{2\sigma^2} y^2 + \frac{1}{2\sigma^2} \sum_{i=0}^{L-1} x_i - \frac{L-1}{2\sigma^2} \sum_{i=0}^{L-1} x_i^2 \right) dy = \exp \left( \frac{1}{2\sigma^2 L^2} \left( \sum_{i=0}^{L-1} x_i \right)^2 - \frac{L-1}{2} \sum_{i=0}^{L-1} x_i^2 \right).
\]

Now the \( L \)-dimensional integral inside the logarithm in Equation (108) equals
\[
\int L \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( s \sum_{i=0}^{L-1} x_i^2 - sLP - \frac{1}{2P} \sum_{i=0}^{L-1} x_i^2 \right) + \frac{1}{2\sigma^2 L^2} \rho \sum_{0 \leq i \neq j \leq L-1} x_i x_j.
\]
\[-(L - 1) \sum_{i=0}^{L-1} x_i^2 \right) d(x_0, \ldots, x_{L-1})
\]
\[= \frac{e^{-sLP}}{\sqrt{2\pi P^L}} \int_{\mathbb{R}^L} \exp \left( s - \frac{1}{2P} - \frac{L - 1}{2\sigma^2 L^2 \rho} \right) \sum_{i=0}^{L-1} x_i^2 \right)d(x_0, \ldots, x_{L-1})
\]
\[+ \frac{1}{2\sigma^2 L^2 \rho} \sum_{0 \leq i \leq j \leq L-1} x_i x_j \right)d(x_0, \ldots, x_{L-1})
\]
\[= \frac{e^{-sLP}}{\sqrt{2\pi P^L}} \int_{\mathbb{R}^L} \exp \left( -\frac{1}{2P} A\vec{x}^T \right) d\vec{x},
\]  
\[(109)\]

where \(\vec{x} = [x_0, \ldots, x_{L-1}] \in \mathbb{R}^L\) and \(A \in \mathbb{R}^{L \times L}\) is a matrix with all diagonal entries equal to
\[a := \frac{1}{2P} + \frac{L - 1}{2\sigma^2 L^2 \rho} - s\]
and all off-diagonal entries equal to
\[b := -\frac{1}{2\sigma^2 L^2 \rho}.
\]

By Lemma 24, the RHS of Equation (109) equals
\[\frac{e^{-sLP}}{\sqrt{2\pi r^L}} \det(A) = \frac{e^{-sLP}}{\sqrt{2\pi \det(A)}}.
\]  
\[(110)\]

To compute \(\det(A)\), we note that \(A = (a - b)I_L + \sqrt{-b} L\tau_1 (\sqrt{-b} I_L)^T\), where \(I_L\) denotes the all-one vector of length \(L\).

**Lemma 35 (Matrix determinant lemma):** Let \(A \in \mathbb{R}^{n \times n}\) be a non-singular matrix and let \(u, v \in \mathbb{R}^n\). Then
\[\det(A + uv^T) = (1 + v^T A^{-1} u) \det(A).
\]

By Lemma 35, we have
\[\det(A) = \left[ 1 + (\sqrt{-b} I_L)^T ((a - b)I_L)^{-1} (\sqrt{-b} I_L) \right] \times \det((a - b)I_L)
\]
\[= \left( 1 + \frac{b}{a - b} \right) (a - b)^L
\]
\[= (a - (L - 1)b)(a - b)^{L-1}
\]
\[= \left( \frac{1}{2P} - s \right) \left( \frac{1}{2P} + \frac{1}{2\sigma^2 L^2 \rho} - s \right)^{L-1}.
\]

Therefore, the (natural) logarithm of the RHS of Equation (110) equals
\[-sLP - \frac{L}{2} \ln(2P) - \frac{1}{2} \ln \left( \frac{1}{2P} - s \right)
\]
\[-\frac{L - 1}{2} \ln \left( \frac{1}{2P} + \frac{1}{2\sigma^2 L^2 \rho} - s \right)
\]
\[-\left( sLP + \frac{1}{2} \ln(1 - 2sP) + \frac{P}{\sigma^2 L^2 \rho} \right).
\]

Plugging the above expression back to Equation (108), we see that to get the largest error exponent, we need to minimize the following expression over \(s \geq 0\) and \(\rho \geq 1\).
\[R(L - 1) \rho - \rho \left[ sLP + \frac{1}{2} \ln(1 - 2sP) + \frac{L - 1}{2} \ln \left( 1 - 2sP + \frac{P}{\sigma^2 L^2 \rho} \right) \right].
\]  
\[(111)\]

From the calculations in Section VIII-E, one can obtain an expression of the solution to the above minimization problem. Specifically, negating Equation (111), by Equation (48), we know that the maximum value equals
\[P(L(1 - 2Ps) - 1) = \frac{\text{snr}(Lt - 1)}{2L^2},
\]  
\[(112)\]

where \(t := 1 - 2Ps\). Recall that \(0 \leq s \leq \frac{1 - 1/L}{2P}\) satisfies Equation (47) which can be rewritten in term of \(t\) as
\[R = \frac{1}{2} \ln \left( \frac{(L - 1) t}{Lt - 1} + \frac{1}{L - 1} \ln t \right).
\]  
\[(113)\]

Equivalently, \(t\) is the unique solution of the equation \(Lt - 1 = e^{2R}(L - 1) t^{\frac{1}{1 - \frac{2}{2P}}}\) in \(t \in [1/L, 1]\).

Equation (112) is valid whenever \(\rho \geq 1\). Recall the relation between \(\rho\) and \(s\) (Equation (45)). We rewrite it in terms of \(t\):
\[
\rho = \frac{(L - 1) + L(t - 1)}{2L^2 - \frac{1}{2P} \cdot t \cdot \sigma^2} = \frac{(Lt - 1) \text{snr}}{L^2(1 - t) t}.
\]

By the above relation between \(\rho\) and \(t\), the condition \(\rho \geq 1\) is equivalent to
\[t \geq \frac{L - \text{snr} + \sqrt{L^2 + \text{snr}^2 - 2\text{snr}L - 2}}{2L}.
\]  
\[(114)\]

Plugging the RHS of Equation (114) to Equation (113), the condition \(\rho \geq 1\) is further equivalent to
\[R \leq \frac{1}{2} \left( \ln \frac{\sqrt{L^2 + \text{snr}^2 - 2\text{snr}L - 2} + L + \text{snr}}{2L}
\]
\[+ \frac{1}{L - 1} \ln \frac{\sqrt{L^2 + \text{snr}^2 - 2\text{snr}L - 2} + L - \text{snr}}{2L} \right),
\]

the RHS of which is defined as \(R_{x,L-1}(\text{snr})\). We conclude that the error exponent given by the RHS of Equation (112) can be achieved for any \(R \leq R_{x,L-1}(\text{snr})\).

\[\square\]

**ACKNOWLEDGMENT**

The authors would like to thank the anonymous reviewers for valuable comments that helped improve the manuscript. They also thank the reviewer 2 for pointing out the fix for the inaccuracy in Blinovsky’s work [12]. Yihan Zhang would like to thank Nir Ailon and Ely Porat for several helpful conversations throughout this project and Alexander Barg for insightful comments on the manuscript.

**REFERENCES**

[1] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, vol. 290. New York, NY, USA: Springer, 2013.

[2] P. Elias, “List decoding for noisy channels,” Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 335, 1957.

[3] J. M. Wozencraft, “List decoding,” *Quart. Prog. Rep., Res. Lab. Electron.*, vol. 48, pp. 90–95, 1958.

[4] M. Langberg, “Private codes or succinct random codes that are (almost) perfect,” in *Proc. 45th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2004, pp. 325–334.

[5] A. D. Sarwate, “Robust and adaptive communication under uncertain interference,” Ph.D. dissertation, EECS Dept., Univ. California, Berkeley, CA, USA, Jul. 2008. [Online]. Available: [http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-86.html](http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-86.html)
Y. Zhang and S. Vatedka, “Lower bounds on list decoding capacity using error exponents,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2022, pp. 1324–1329.

C. E. Shannon, “Probability of error for optimal codes in a Gaussian channel,” Bell Syst. Tech. J., vol. 38, no. 3, pp. 611–656, May 1959.

C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I.” Inf. Control, vol. 10, no. 1, pp. 65–103, Jan. 1967.

V. M. Blinovsky, “Error probability exponent of list decoding at low rates,” Problems Inf. Transmiss., vol. 37, no. 4, pp. 277–287, Oct./Dec. 2001.

C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948.

R. Gallager, “A simple derivation of the coding theorem and some applications,” IEEE Trans. Inf. Theory, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.

G. Poltyrev, “On coding without restrictions for the AWGN channel,” IEEE Trans. Inf. Theory, vol. 40, no. 2, pp. 409–417, Mar. 1994.

A. Ingber, R. Zamir, and M. Feder, “Finite-dimensional infinite constellations,” IEEE Trans. Inf. Theory, vol. 59, no. 3, pp. 1630–1656, Mar. 2013.

V. Anantharam and F. Baccelli, “Information-theoretic capacity and error exponents of stationary point processes under random additive displacements,” 2010, arXiv:1012.4924.

D.-T. Lee, “On k-nearest neighbor Voronoi diagrams in the plane,” IEEE Trans. Comput., vol. C-31, no. 6, pp. 478–487, Jun. 1982.

R. A. Rankin, “The closest packing of spherical caps in n dimensions,” Proc. Glasgow Math. Assoc., vol. 2, no. 3, pp. 139–144, Jul. 1955.

E. Spence, “Packing of spheres in LP,” Glasgow Math. J., vol. 11, no. 1, pp. 72–80, 1970.

K. Ball, “Inequalities and sphere-packing in $\ell_p$,” Isr. J. Math., vol. 58, no. 2, pp. 243–256, Jun. 1987.

A. Samorodnitsky. (2013). A Bound on $l_1$ Codes. [Online]. Available: https://www.cs.huji.ac.il/~saless/papers/L1_codes.pdf

R. van Handel, “Probability in high dimension,” Oper. Res. Financial Eng., Princeton Univ., Princeton, NJ, USA, Tech. Rep. ADA623999, 2014.

M. Haenggi, Stochastic Geometry for Wireless Networks. Cambridge, U.K.: Cambridge Univ. Press, 2012.

Yihan Zhang received the B.Eng. degree in computer science and technology from Northeastern University, Shenyang, China, in June 2016, and the Ph.D. degree from the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, in August 2020. He was a Post-Doctoral Researcher with the Henry and Marilyn Taub Faculty of Computer Science, Technion-Israel Institute of Technology, from October 2020 to October 2021. He has been a Post-Doctoral Researcher with the Institute of Science and Technology Austria since October 2021. His current research interests include coding theory, information theory, and statistics theory (in no particular order).

Shashank Vatedka (Senior Member, IEEE) received the M.Sc. (Engineering) and Ph.D. degrees from the Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru, India, in 2017. From 2016 to 2018, he was a Research Assistant and a Post-Doctoral Researcher with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong, SAR. From 2018 to 2019, he was a Post-Doctoral Fellow with Télécom Paris, France. Since October 2019, he has been an Assistant Professor with the Department of Electrical Engineering, IIT Hyderabad. His current research interests include information theory and coding, with applications to data compression, security, and statistical inference. He was a recipient of the TCS Fellowship during his Ph.D. studies; and the Seshagiri Kaikini Medal for Best Ph.D. Thesis from the Department of Electrical Communication Engineering, IISc Bengaluru. He also received the Best Paper Award-Honorable Mention from SPCOM 2020, the Best Poster Award from the 2021 Stanford Compression Workshop, and the Best Paper Award (Communications Track) from NCC 2023.