The Dispersion of Infinite Constellations

Amir Ingber† , Ram Zamir and Meir Feder
Dept. of EE-Systems, The Faculty of Engineering,
Tel Aviv University
Tel Aviv 69978, Israel
Email: {ingber, zamir, meir}@eng.tau.ac.il

***Abstract*—The setting of a Gaussian channel without power constraints is considered. In this setting, proposed by Poltyrev, the codewords are points in an n-dimensional Euclidean space (an infinite constellation). The channel coding analog of the number of codewords is the density of the constellation points, and the analog of the communication rate is the normalized log density (NLD). The highest achievable NLD with vanishing error probability is known (which can be thought of as the capacity), as well as error exponents for the setting. In this work we are interested in the optimal NLD for communication when a fixed, nonzero error probability is allowed. In classical channel coding the gap to capacity is characterized by the channel dispersion (and cannot be derived from error exponent theory). In the unconstrained setting, we show that as the codeword length (dimension) n grows, the gap to the highest achievable NLD is inversely proportional (to the first order) to the square root of the block length. We give an explicit expression for the proportion constant, which is given by the inverse Q-function of the allowed error probability, times the square root of $\frac{1}{2}$. In an analogy to a similar result in classical channel coding, it follows that the dispersion of infinite constellations is given by $\frac{1}{2}\mathsf{nat}^2$ per channel use. We show that this optimal convergence rate can be achieved using lattices, therefore the result holds for the maximal error probability as well. Connections to the error exponent of the power constrained Gaussian channel and to the volume-to-noise ratio as a figure of merit are discussed.**

I. Introduction

Coding schemes over the Gaussian channel are traditionally limited by the average/peak power of the transmitted signal [1]. Without the power restriction (or a similar restriction) the channel capacity becomes infinite, since one can space the codewords arbitrarily far apart from each other and achieve a vanishing error probability. However, many coded modulation schemes take an infinite constellation (IC) and restrict the usage to points of the IC that lie within some n-dimensional form in Euclidean space (a 'shaping' region). Probably the most important example for an IC is a lattice, and examples for the shaping regions include a hypersphere in n dimensions, and a Voronoi region of another lattice [2].

In 1994, Poltyrev [3] studied the model of a channel with Gaussian noise without power constraints. In this setting the codewords are simply points in the n-dimensional Euclidean space. The analog to the number of codewords is the density γ of the constellation points (the average number of points per unit volume). The analog of the communication rate is the normalized log density (NLD) $\delta \triangleq \frac{1}{n}\log\gamma$. The error probability in this setting can be thought of as the average error probability, where all the points of the IC have equal transmission probability (precise definitions follow later on in the paper). Poltyrev showed that the NLD δ is the analog of the rate in classical channel coding, and established the analog term to the capacity, the ultimate limit for the NLD, denoted δ^*. Random coding and sphere packing error exponent bounds were also derived, which are analogous to Gallager's error exponents in the classical channel coding setting [4], and to the error exponents of the power-constrained AWGN channel [5], [4].

In classical channel coding, the channel capacity gives the ultimate limit for the rate when arbitrarily small error probability is required, and the error exponent quantifies the (exponential) speed at which the error probability goes to zero when the rate is fixed (and below the channel capacity). Another question that is of interest is the following: for a fixed error probability ε, what is the optimal (maximal) rate that is achievable when the codeword length n is fixed. While the exact answer for this question for any finite n is still open (see [6] for the current state of the art), the speed at which the optimal rate converges to the capacity is known. By letting $R_\varepsilon(n)$ denote the maximal rate for which there exist communication schemes with codelength n and error probability at most ε, it is known that for a channel with capacity C [7]:

$$R_\varepsilon(n) = C - \sqrt{\frac{V}{n}}Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{1}$$

where $Q^{-1}(\cdot)$ is the inverse complementary standard Gaussian cumulative distribution function. The constant V, termed the *channel dispersion*, is the variance of the information spectrum $i(x;y) \triangleq \log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$ for a capacity-achieving distribution. More details and extensions can be found in [6].

In this paper we are interested in finding out whether the behavior demonstrated in (1) exists in the setting of a Gaussian channel without power constraints. We answer this question to the positive. The main result is the following: for a given, fixed, nonzero error probability ε, denote by $\delta_\varepsilon(n)$ be the maximal NLD for which there exists an IC with dimension n and error probability at most ε. Then

$$\delta_\varepsilon(n) = \delta^* - \sqrt{\frac{1}{2n}}Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{2}$$

† A. Ingber is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

where δ^* is the ultimate limit for the NLD with any dimension [3], given by $\frac{1}{2}\log\frac{1}{2\pi e\sigma^2}$ where σ^2 is the variance of the additive Gaussian noise (logarithms are taken w.r.t. to the natural base e).

In the achievability part we use lattices (and the Minkowski-Hlawka theorem [8]). Because of the regular structure of lattices, our achievability result holds for the maximal error probability. The proof technique used is somewhat different than that used by Poltyrev in [3]. Here we use a suboptimal 'typicality decoder', closer in spirit to that used in the standard achievability proofs [9] (rather than the the the technical ML decoder based proof [3]). In addition, a variant of the typicality decoder can be used to prove Poltyrev's random coding exponent (see [10] for details). In the converse part of the proof we consider the average error probability and any IC (not only lattices), therefore our result (2) holds for both average and maximal error probability, and for any IC (lattice or not).

Another figure of merit for lattices (that can be defined for general ICs as well) is the volume-to-noise ratio (VNR), which generalizes the SNR notion (see, e.g. [11]). The VNR quantifies how good a lattice is for channel coding over the unconstrained AWGN. It is known that the VNR of any lattice cannot be below $2\pi e$, and that there exist lattices that approach this value as the dimension grows. As a consequence of the paper's main result we show the asymptotical behavior of the optimal VNR.

In the next section we discuss the relations of our result to the error exponent theory and to the power constrained AWGN channel. In Section III we go over the necessary notations and state a key lemma that is required for our main result, which is given in Section IV. In Section V we obtain the optimal VNR as a consequence of the main result. Due to space limitations only proof outlines are provided. Detailed proofs and extensions can be found in [10].

II. Connections to Error Exponents and the Power Constrained AWGN

By the similarity of Equations (1) and (2) we can isolate the constant $\frac{1}{2}$ and identify it as the dispersion of the unconstrained AWGN setting. In this section we discuss this fact and its relation to classical channel coding and to the power-constrained AWGN channel.

One interesting property of the channel dispersion theorem (1) is the following connection to the error exponent. Under some mild regularity assumptions, the error exponent can be approximated near the capacity by

$$\mathbf{E}(R) \cong \frac{(C-R)^2}{2V}, \tag{3}$$

where V is the channel dispersion. This property, which is attributed to Shannon (see [6, Fig. 18]), holds for DMCs and for the power constrained AWGN channel and is conjectured to hold in more general cases. Note, however, that while the parabolic behavior of the exponent hints that the gap to the capacity should behave as $O\left(\frac{1}{\sqrt{n}}\right)$, the dispersion theorem (1) cannot be derived directly from the error exponent theory (even if the error probability was given by $e^{-n\mathbf{E}(R)}$ exactly).

Analogously to (3), we examine the error exponent for the unconstrained Gaussian setting. For NLD values above the critical NLD $\delta_{cr} \triangleq \frac{1}{2}\log\frac{1}{4\pi e\sigma^2}$ (but below δ^*), the error exponent is given by [3]:

$$\mathbf{E}(\delta,\sigma^2) = \frac{e^{-2\delta}}{4\pi e\sigma^2} + \delta + \frac{1}{2}\log 2\pi\sigma^2. \tag{4}$$

By straightforward differentiation we get that the second derivative (w.r.t. δ) of $\mathbf{E}(\delta,\sigma^2)$ at $\delta = \delta^*$ is given by 2, so according to (3), it is expected that the dispersion for the unconstrained AWGN channel will be $\frac{1}{2}$. This agrees with our main result and its similarity to (1), and extends the correctness of the conjecture (3) to the unconstrained AWGN setting as well. It should be noted, however, that our result provides more than just proving the conjecture: there exist examples where the error exponent is well defined (with second derivative), but a connection of the type (3) can only be achieved asymptotically with $\varepsilon \to 0$ (see, e.g. [12]). Our result (2) holds for any finite ε.

Another indication that the dispersion for the unconstrained setting should be $\frac{1}{2}$ comes the connections to the the power constrained AWGN. While the capacity $\frac{1}{2}\log(1+P)$, where P denotes the channel SNR, is clearly unbounded with P, the form of the error exponent curve does have a nontrivial limit as $P \to \infty$. In [2] it was noticed that this limit is the error exponent of the unconstrained AWGN channel (sometimes termed the 'Poltyrev exponent'), where the distance to the capacity is replaced by the NLD distance to δ^*. By this analogy we examine the dispersion of the power constrained AWGN channel at high SNR. In [6] the dispersion was found, given (in nat^2 per channel use) by

$$V_{AWGN} = \frac{P(P+2)}{2(P+1)^2}. \tag{5}$$

This term already appeared in Shannon's 1959 paper on the AWGN error exponent [5], where its inverse is exactly the second derivative of the error exponent at the capacity (i.e. (3) holds for the AWGN channel). It is therefore no surprise that by taking $P \to \infty$, we get the desired value of $\frac{1}{2}$, thus completing the analogy between the power constrained AWGN and its unconstrained version. This convergence is quite fast, and is tight for SNR as low as 10dB (see Fig. 1).

III. Preliminaries

A. Notation

We adopt most of the notations of Poltyrev's paper [3]: Let $\mathrm{Cb}(a)$ denote a hypercube in $\mathbb{R}^n$

$$\mathrm{Cb}(a) \triangleq \left\{\mathbf{x} \in \mathbb{R}^n \ s.t.\ \forall_i |x_i| < \frac{a}{2}\right\}. \tag{6}$$

Let $\mathrm{Ball}(r)$ denote a hypersphere in $\mathbb{R}^n$ and radius $r > 0$, centered at the origin

$$\mathrm{Ball}(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n \ s.t.\ \|\mathbf{x}\| < r\}, \tag{7}$$

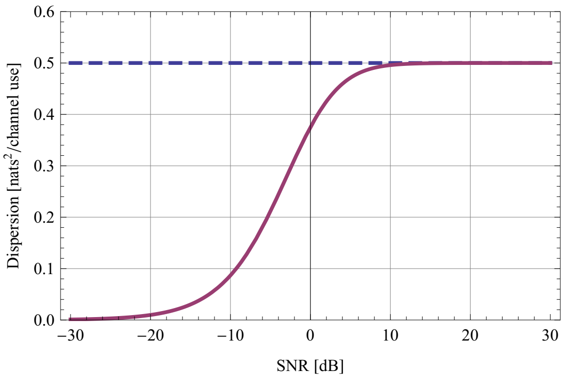

Fig. 1. The power-constrained AWGN dispersion (solid) vs. the unconstrained dispersion (dashed)

and let $\text{Ball}(\mathbf{y}, r)$ denote a hypersphere in $\mathbb{R}^n$ and radius $r > 0$, centered at $\mathbf{y} \in \mathbb{R}^n$

$$\text{Ball}(\mathbf{y}, r) \triangleq \{\mathbf{x} \in \mathbb{R}^n \ s.t. \ \|\mathbf{x} - \mathbf{y}\| < r\}. \tag{8}$$

Let $\mathcal{S}$ be an IC. We denote by $M(\mathcal{S}, a)$ the number of points in the intersection of $\text{Cb}(a)$ and the IC $\mathcal{S}$, i.e. $M(\mathcal{S}, a) \triangleq |\mathcal{S} \bigcap \text{Cb}(a)|$. The density of $\mathcal{S}$, denoted by $\gamma(\mathcal{S})$, or simply γ, measured in points per volume unit, is defined by

$$\gamma(\mathcal{S}) \triangleq \limsup_{a\to\infty} \frac{M(\mathcal{S}, a)}{a^n}. \tag{9}$$

The normalized log density (NLD) δ is defined by

$$\delta \triangleq \frac{1}{n} \log \gamma. \tag{10}$$

It will prove useful to define the following:

Definition 1 (Expectation over a hypercube): Let $f : \mathcal{S} \to \mathbb{R}$ be an arbitrary function. Let $\mathbb{E}_a[f(s)]$ denote the expectation of $f(s)$, where s is drawn uniformly from the code points that reside in the hypercube $\text{Cb}(a)$:

$$\mathbb{E}_a[f(s)] \triangleq \frac{1}{M(\mathcal{S}, a)} \sum_{s\in\mathcal{S}\cap\text{Cb}(a)} f(s). \tag{11}$$

Throughout the paper, an IC will be used for transmission of information through the unconstrained AWGN channel with noise variance σ^2 (per dimension). The additive noise shall be denoted by $\mathbf{Z} = [Z_1, ..., Z_n]^T$. An instantiation of the noise vector shall be denoted by $\mathbf{z} = [z_1, ..., z_n]^T$.

For $s \in \mathcal{S}$, let $P_e(s)$ denote the error probability when s was transmitted. When the maximum likelihood (ML) decoder is used, the error probability is given by $P_e(s) = \Pr\{s + \mathbf{Z} \notin W(s)\}$, where $W(s)$ is the *Voronoi region* of s, i.e. the convex polytope of the points that are closer to s than to any other point $s' \in \mathcal{S}$. The maximal error probability is defined by

$$P_e^{\max}(\mathcal{S}) \triangleq \sup_{s\in\mathcal{S}} P_e(s), \tag{12}$$

and the average error probability is defined by

$$P_e(\mathcal{S}) \triangleq \limsup_{a\to\infty} \mathbb{E}_a[P_e(s)]. \tag{13}$$

B. A Key Lemma

A key lemma that will be used throughout the paper is a lemma regarding the norm of a Gaussian vector.

Lemma 1: Let $\mathbf{Z} = [Z_1, ..., Z_n]^T$ be a vector of n zero-mean, independent Gaussian random variables, each with mean σ^2. Let $r > 0$ be a given arbitrary radius. Then the following holds for any dimension n:

$$\left| \Pr\{\|\mathbf{Z}\| > r\} - Q\left(\frac{r^2 - n\sigma^2}{\sigma^2\sqrt{2n}}\right) \right| \le \frac{6T}{\sqrt{n}}, \tag{14}$$

where $Q(\cdot)$ is the standard complementary cumulative distribution function, $\|\cdot\|$ is the usual ℓ_2 norm, and

$$T = \mathbf{E}\left[\left| \frac{X^2 - 1}{\sqrt{2}} \right|^3 \right] \approx 3.0785, \tag{15}$$

where X is a standard Gaussian RV.

Proof outline: The proof relies on the convergence of a sum of independent random variables to a Gaussian random variable, i.e. the central limit theorem. We first note that

$$\Pr\{\|\mathbf{Z}\| > r\} = \Pr\left\{ \sum_{i=1}^n Z_i^2 > r^2 \right\}. \tag{16}$$

Let $Y_i = \frac{Z_i^2 - \sigma^2}{\sigma^2\sqrt{2}}$ and let $S_n \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$. It is easy to verify that Y_i and S_n have zero mean and unit variance. It follows that

$$\Pr\left\{ \sum_{i=1}^n Z_i^2 \ge r^2 \right\} = \Pr\left\{ S_n \ge \frac{r^2 - n\sigma^2}{\sigma^2\sqrt{2n}} \right\}. \tag{17}$$

S_n is a normalized sum of i.i.d. variables, and by the central limit theorem converges to a standard Gaussian random variables. The Berry-Esseen theorem (see, e.g. [13, Ch. XVI.5]) quantifies the rate of convergence in the cumulative distribution function sense, and states that for any $\alpha > 0$

$$|\Pr\{S_n \ge \alpha\} - Q(\alpha)| \le \frac{6T}{\sqrt{n}}, \tag{18}$$

where $T = \mathbb{E}[|Y_i|^3]$. The proof of the lemma is completed by applying the Berry-Esseen theorem to the RHS of (17). ■

IV. Main Result

A. Direct

In the direct part we show that for any fixed, nonzero error probability $\varepsilon > 0$, there exist ICs with error probability at most ε, and NLD δ according to (2). These ICs will be *lattices*, and therefore the same result will hold for the *maximal* error probability. For lattices the average error probability is identical for all the code points, since all the Voronoi cells of a lattice are congruent. In addition, the volume of a voronoi cell is equal for all cells, and is equal to the determinant of the lattice $\det \Lambda$. The density (in code points per volume unit) is therefore $\gamma = (\det \Lambda)^{-1}$, and the NLD is $\delta = \frac{1}{n} \log \gamma = -\frac{1}{n} \log \det \Lambda$.

Theorem 1: Let $\varepsilon > 0$. There exists a lattice Λ with maximal error probability at most ε, and NLD

$$\delta = \delta^* - \sqrt{\frac{1}{2n}}Q^{-1}(\varepsilon) + O\left(\frac{1}{n}\right). \quad (19)$$

Proof outline: Let Λ be a lattice that is used as an IC for transmission over the unconstrained AWGN. We consider a suboptimal *typicality decoder*, which operates as follows. Suppose that $\lambda \in \Lambda$ is sent, and the point $\mathbf{y} = \lambda + \mathbf{z}$ is received, where $\mathbf{z}$ is the additive noise. Let $r > 0$ be an arbitrary radius. If there is only a single point in the ball $\text{Ball}(\mathbf{y}, r)$, then this will be the decoded word. If there are no codewords in the ball, or more than one codeword in the ball, an error is declared (one of the code points is chosen at random).

When analyzing the error probability we can assume w.l.o.g. that the zero codeword was sent. We note that the average error probability with the typicality decoder is bounded by

$$\begin{aligned} P_e(\Lambda) \leq &\Pr\{\mathbf{Z} \notin \text{Ball}(r)\} \quad (20)\\ &+ \sum_{\lambda \in \Lambda \setminus \{0\}} \Pr\{\mathbf{Z} \in \text{Ball}(\lambda, r) \cap \text{Ball}(r)\}, \end{aligned}$$

where $\mathbf{Z}$ denotes the noise vector. In order to bound the second term in (20) we use a version of the Minkowski-Hlawka (MH) theorem [8, Lemma 3, p. 65], and conclude that there exist a lattice Λ for which (for any density $\gamma > 0$ and any $\xi > 0$)

$$\begin{aligned} &\sum_{\lambda \in \Lambda \setminus \{0\}} \Pr\{\mathbf{Z} \in \text{Ball}(\lambda, r) \cap \text{Ball}(r)\} \\ &\qquad \leq \gamma \int_{\mathbb{R}^n} \Pr\{\mathbf{Z} \in \text{Ball}(\lambda, r) \cap \text{Ball}(r)\}\, d\lambda + \xi. \end{aligned}$$

The integral can be bounded by $r^n V_n$, the volume of $\text{Ball}(r)$ (where $V_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$ is the volume of the unit sphere).

Combined with (20) we get that there exist a lattice Λ with density γ, in which

$$P_e(\Lambda) \leq \Pr\{\|\mathbf{Z}\| > r\} + \gamma V_n r^n + \xi, \quad (21)$$

where $r > 0$, $\gamma > 0$ and $\xi > 0$ can be chosen arbitrarily.

It appears that the dominant term in (21) is $\Pr\{\|\mathbf{Z}\| > r\}$. The intuition follows from the the converse result (Theorem 2 below), where $\Pr\{\|\mathbf{Z}\| > r\}$ is the only term in the lower bound.

Let $\varepsilon > 0$ be the desired error probability. Determine r s.t. $\Pr(\|\mathbf{Z}\| > r) = \varepsilon\left[1 - \frac{2}{\sqrt{n}}\right]$, γ s.t. $\gamma V_n r^n = \frac{\varepsilon}{\sqrt{n}}$, and $\xi = \varepsilon \frac{1}{\sqrt{n}}$. This way it is assured that the error probability is not greater than $\varepsilon\left[1 - \frac{2}{\sqrt{n}}\right] + \frac{\varepsilon}{\sqrt{n}} + \frac{\varepsilon}{\sqrt{n}} = \varepsilon$.

Define α_n s.t. $r^2 = n\sigma^2(1 + \alpha_n)$ (recall that r implicitly depends on n as well). Using Lemma 1 and some algebra gives

$$\alpha_n = \sqrt{\frac{2}{n}}Q^{-1}(\varepsilon) + O\left(\frac{1}{n}\right). \quad (22)$$

So far, we have shown the existence of a lattice Λ with error probability at most ε. The NLD is given by $\delta = \frac{1}{n}\log\gamma = \frac{1}{n}\log\frac{\varepsilon}{V_n r^n \sqrt{n}}$. The required result follows using (22), the Stirling approximation for V_n and the Taylor approximation for $\log(1+x)$. ■

B. Converse

In the the direct part we have shown the existence of good ICs with NLD that approaches the NLD capacity δ^*. These ICs were lattices, and the convergence to δ^* was of the form $\sqrt{\frac{1}{2n}}Q^{-1}(\varepsilon)$. We now show that this is the *optimal* convergence rate, for *any* IC (not only for lattices).

The results in the converse part are concerned with the average error probability $P_e(\mathcal{S})$. A lower bound on the average error probability is clearly a lower bound on the maximal error probability as well.

Theorem 2: Let $\mathcal{S}$ be an IC with NLD $\delta = \frac{1}{n}\log\gamma$ and average error probability $P_e(\mathcal{S}) = \varepsilon$. Then the NLD δ is bounded by

$$\delta \leq \delta^* - \sqrt{\frac{1}{2n}}Q^{-1}(\varepsilon) + \frac{1}{2n}\log n + O\left(\frac{1}{n}\right). \quad (23)$$

Proof outline: The proof has three parts. First we prove the converse for ICs where all the Voronoi cells have an equal volume. Then we extend the proof to ICs with some mild regularity properties, and only then prove the converse for *any* IC.

Suppose the Voronoi regions of $\mathcal{S}$ have the same volume $\frac{1}{\gamma}$. Such ICs include the important class of Lattices, as well as many other constellation types. Suppose $s \in \mathcal{S}$ is sent. Let r be the radius of a sphere with the same volume as the Voronoi region $W(s)$:

$$|W(s)| = \frac{1}{\gamma} = e^{-n\delta} = r^n V_n, \quad (24)$$

therefore $r = e^{-\delta} V_n^{-\frac{1}{n}}$. By the *equivalent sphere* argument [3][14], the probability that the noise leaves $W(s)$ is lower bounded by the probability to leave a sphere of the same volume:

$$P_e(s) \geq \Pr\{\|\mathbf{Z}\| \geq e^{-\delta} V_n^{-\frac{1}{n}}\}. \quad (25)$$

By assumption, all the Voronoi regions have the same volume. Therefore the bound (25) holds for any $s \in \mathcal{S}$, and also for the average error probability ε.

The probability $\Pr\{\|\mathbf{Z}\| \geq r\}$, or $\Pr\{\|\mathbf{Z}\|^2 \geq r^2\}$, is equal to the CDF of a χ^2 random variable with n degrees of freedom. There is no closed-form expression for the CDF of this probability distribution. In [3], this probability is lower bounded by $\exp[-n(E_L - o(1))]$, where E_L is a function of δ and σ^2 only (and not n). This gives the sphere packing exponent for this setting. In [14], this probability was calculated as a sum of $n/2$ elements that gives the exact expression, but its asymptotic behavior is hard to characterize. Here we use the normal approximation in order to determine the behavior of the NLD δ with n, where the error probability ε remains fixed.

Combined with Lemma 1 we have

$$\varepsilon \geq Q\left(\frac{e^{-2\delta}V_n^{-\frac{2}{n}} - n\sigma^2}{\sigma^2\sqrt{2n}}\right) - \frac{6T}{\sqrt{n}}, \quad (26)$$

where T is a constant given in (15). The desired result (23) then follows from the Stirling and Taylor approximations.

We now extend the result to ICs with some mild regularity assumptions:

Definition 2 (Regular ICs): An IC $\mathcal{S}$ is called *regular*, if:

1) There exists a radius $r_0 > 0$, s.t. for all $s \in \mathcal{S}$, the Voronoi cell $W(s)$ is contained in $\text{Ball}(s, r_0)$.
2) The density $\gamma(\mathcal{S})$ is given by $\lim_{a\to\infty} \frac{M(\mathcal{S},a)}{a^n}$ (rather than $\limsup$ in the original definition).

Let $\mathcal{S}$ be a regular IC. For $s \in \mathcal{S}$, we denote by $v(s)$ the volume of the Voronoi cell of s, $|W(s)|$. We also define the average Voronoi cell volume of a regular IC by $v(\mathcal{S}) \triangleq \limsup_{a\to\infty} \mathbb{E}_a[v(s)]$. It can be shown that for a regular IC, the average volume is the inverse of the density, i.e. $\gamma(\mathcal{S}) = \frac{1}{v(\mathcal{S})}$.

Let $\text{SPB}(v)$ denote the probability that the noise leaves a sphere of volume v. By the equivalent sphere argument we have $P_e(s) \geq \text{SPB}(v(s))$ for all $s \in \mathcal{S}$. It can be shown that $\text{SPB}(v)$ is a convex function of the volume v. We now extend the equivalent sphere bound to the average volume as well:

$$\begin{aligned} P_e(\mathcal{S}) &= \limsup_{a\to\infty} \mathbb{E}_a[P_e(s)] \\ &\overset{(a)}{\geq} \limsup_{a\to\infty} \mathbb{E}_a[\text{SPB}(v(s))] \\ &\overset{(b)}{\geq} \limsup_{a\to\infty} \text{SPB}(\mathbb{E}_a[v(s)]) \\ &\overset{(c)}{=} \text{SPB}(\limsup_{a\to\infty} \mathbb{E}_a[v(s)]) \\ &= \text{SPB}(v(\mathcal{S})). \end{aligned} \tag{27}$$

(a) follows from the sphere bound for each individual point $s \in \mathcal{S}$, (b) follows from the Jensen inequality and the convexity of $\text{SPB}(\cdot)$, and (c) follows since $\text{SPB}(\cdot)$ is continuous. Following the same steps as in the constant Voronoi volume case extends (23) to the case of regular ICs as well.

The final step in the converse proof is to extend (23) to non-regular ICs. Such ICs include constellations which are semi-infinite (e.g. contains points only in half of the space), and also constellations in which the density oscillates with the cube size a (and the formal limit γ does not exist). This is done with the aid of a *regularization process* - for any IC $\mathcal{S}$ with NLD δ and error probability ε, there exists a *regular* IC $\mathcal{S}'$ with NLD δ' and error probability ε' which are close to δ and ε respectively. Then we apply (23) on the regular IC $\mathcal{S}'$ and get the desired result. The technical details of the proof and the regularization process can be found in [10]. ■

V. Volume-to-Noise Ratio

The volume-to-noise ratio (VNR) of a lattice Λ is defined

$$\mu(\Lambda, \varepsilon) \triangleq \frac{[\text{Vol. of Voronoi region}]^{2/n}}{[\text{noise var.}]} = \frac{\gamma^{-2/n}}{\sigma^2(\varepsilon)}, \tag{28}$$

where $\sigma^2(\varepsilon)$ is the noise variance s.t. the error probability is exactly ε. This dimensionless figure of merit offers another way to quantify the goodness of the lattice for coding over the unconstrained AWGN channel. Note that the VNR is invariant to scaling of the lattice, and that the definition can be extended to general infinite constellations.

The minimum possible value of $\mu(\Lambda, \varepsilon)$ over all lattices in $\mathbb{R}^n$ is denoted by $\mu_n(\varepsilon)$, and it is known that for any $1 > \varepsilon > 0$, $\lim_{n\to\infty} \mu_n(\varepsilon) = 2\pi e$. Using the main result of the paper we can show *how* $\mu_n(\varepsilon)$ approaches $2\pi e$:

Theorem 3: For a fixed error probability $\varepsilon > 0$, The optimal VNR $\mu_n(\varepsilon)$ is given by

$$\mu_n(\varepsilon) = 2\pi e + \sqrt{\frac{8\pi^2 e^2}{n}} Q^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{29}$$

Proof outline: By definition, the following relation holds for any σ^2:

$$\mu_n(\varepsilon) = \frac{e^{-2\delta_\varepsilon(n)}}{\sigma^2} \tag{30}$$

(note that $\delta_\varepsilon(n)$ implicitly depends on σ^2 as well). (29) follows from algebraic manipulations of the main result (2). ■

Note that the theorem can be slightly strengthened by using the more delicate bounds on $\delta_\varepsilon(n)$ in Theorems 1 and 2 rather than the loose term $O\left(\frac{\log n}{n}\right)$.

References

[1] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2384–2415, 1998.

[2] U. Erez and R. Zamir, "Achieving 1/2 log(1+snr) over the additive white Gaussian noise channel with lattice encoding and decoding," *IEEE Trans. on Information Theory*, vol. 50, pp. 2293–2314, Oct. 2004.

[3] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. on Information Theory*, vol. 40, no. 2, pp. 409–417, 1994.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[5] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *The Bell System technical journal*, vol. 38, pp. 611–656, 1959.

[6] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2307 –2359, May 2010.

[7] V. Strassen, "Asymptotische abschätzungen in shannons informationstheorie," *Trans. Third Prague Conf. Information Theory, 1962, Czechoslovak Academy of Sciences*, pp. 689–723.

[8] E. Hlawka, J. Shoißengeier, and R. Taschner, *Geometric and Analytic Numer Theory*. Springer-Verlang, 1991.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & sons, 1991.

[10] A. Ingber, R. Zamir, and M. Feder, "The optimal density of infinite constellations for the Gaussian channel," *Available on arxiv.org*.

[11] R. Zamir, "Lattices are everywhere," in *4th Annual Workshop on Information Theory and its Applications*, UCSD, (La Jolla, CA), 2009.

[12] A. Ingber and M. Feder, "Parallel bit-interleaved coded modulation," *Available on arxiv.org*.

[13] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 2 (2nd Edition)*. John Wiley & Sons, 1971.

[14] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the performance of lattice codes," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 670 –681, mar. 1999.