Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN Channel With Lattice Encoding and Decoding

Uri Erez, *Member, IEEE,* and Ram Zamir, *Senior Member, IEEE*

***Abstract*—We address an open question, regarding whether a lattice code with lattice decoding (as opposed to maximum-likelihood (ML) decoding) can achieve the additive white Gaussian noise (AWGN) channel capacity. We first demonstrate how minimum mean-square error (MMSE) scaling along with dithering (lattice randomization) techniques can transform the power-constrained AWGN channel into a modulo-lattice additive noise channel, whose effective noise is reduced by a factor of $\sqrt{\frac{1+\text{SNR}}{\text{SNR}}}$. For the resulting channel, a uniform input maximizes mutual information, which in the limit of large lattice dimension becomes $\frac{1}{2}\log(1+\text{SNR})$, i.e., the full capacity of the original power constrained AWGN channel. We then show that capacity may also be achieved using nested lattice codes, the coarse lattice serving for shaping via the modulo-lattice transformation, the fine lattice for channel coding. We show that such pairs exist for any desired nesting ratio, i.e., for any signal-to-noise ratio (SNR). Furthermore, for the modulo-lattice additive noise channel lattice decoding is optimal. Finally, we show that the error exponent of the proposed scheme is lower bounded by the Poltyrev exponent.**

***Index Terms*—Additive white Gaussian noise (AWGN) channel, dirty paper channel, dither, Euclidean distance, lattice decoding, minimum mean-square error (MMSE) estimation, nested codes, Poltyrev exponent, random lattice ensemble, shaping.**

I. Introduction

THE search for low-complexity, structured encoding and decoding for the additive white Gaussian noise (AWGN) channel

$$Y = X + N, \;\; N \sim \mathcal{N}(0, P_N) \tag{1}$$

inspired the minds of researchers and continues to challenge the communication community today [21], [4]. The goal is to find codes with rates approaching capacity

$$C = \frac{1}{2}\log(1+\text{SNR}), \tag{2}$$

which allow for decoding with low probability of error at affordable complexity. It is desired to accomplish that for any signal-to-noise ratio SNR $= P_X/P_N$, i.e., under any power constraint $P_X > 0$ on the transmitted signal $X_1, \ldots, X_n$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \le P_X. \tag{3}$$

There are two aspects to signal space codebook design for the power-constrained AWGN channel. The granular structure of the codebook corresponds to the inter-codeword Euclidean distances, hence, it determines the decoding error probability. The structure of the *shaping region* of the codebook determines the power–volume tradeoff; hence, the gap from capacity [17].

Several different approaches for using structured codes for the AWGN channel correspond to different ways of taking into account the power constraint. Shannon's theory suggests that the codewords of a good code should look like realizations of a zero-mean independent and identically distributed (i.i.d.) Gaussian source with power P_X. For large codebook dimension n, this is equivalent to a uniform distribution over a sphere of radius $\sqrt{nP_X}$. Slepian considered the use of group codes for the AWGN channel in [30], where the codewords lie on the surface of this sphere of radius $\sqrt{nP_X}$.

The central line of development in the application of lattices for the AWGN channel, and the most directly related to the problem we study, originated in the work of de Buda. De Buda's theorem [9] states that a spherical lattice code, i.e., a code with second moment P_X, which is the intersection of a lattice with a sphere, can approach arbitrarily closely (in the limit of high dimension) the AWGN channel capacity. To achieve the best error exponent of the AWGN channel (or at least the lower bounds to the error exponent [22, Sec. 7.3], which are tight above the critical rate), a "thin" spherical region is taken instead of a full sphere. This result has been corrected and refined by several authors [24], [31], [28], [26].

However, when a lattice code is defined in this manner, much of the structure and symmetry of the underlying lattice is lost. In addition, the optimality of this scheme relies on *maximum-likelihood* (ML) decoding, i.e., requires finding the lattice point *inside* the sphere which is closest to the received signal. The resulting decision regions are not fundamental regions of the lattice and are unbounded. In contrast, *lattice decoding* amounts to finding the closest lattice point, ignoring the boundary of the code. Such an unconstrained search preserves the lattice symmetry in the decoding process and saves complexity, and thus it attracted special attention [1], [28].

When restricted to lattice decoding, however, existing lattice coding schemes can transmit reliably only at rates up to $\frac{1}{2}\log(\text{SNR})$ [10], [28]. This loss of "one" in the rate formula

Manuscript received June 5, 2001; revised January 30, 2004. This work was supported in part by the Israel Academy of Science #65/01. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Washington, DC, June 2001 and the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

U. Erez was with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 89978, Israel. He is now with the Signals, Information and Algorithms Laboratory, The Massachusetts Institute of Technology, Cambridge, MA USA (e-mail uri@allegro.mit.edu).

R. Zamir is with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv 89978, Israel (e-mail: zamir@eng.tau.ac.il).

Communicated by R. Urbanke, Associate Editor for Coding Techniques.

Digital Object Identifier 10.1109/TIT.2004.834787

means significant degradation in performance at low SNR and zero rate for SNR < 1. In fact, it was conjectured [10], [28], [26] that with lattice decoding the rate $\frac{1}{2}\log(\text{SNR})$ cannot be surpassed. See also the discussion in [20], [31].

We show that with a slightly different definition of lattice transmission and lattice decoding, the full capacity $\frac{1}{2}\log(1+\text{SNR})$ of the channel may be achieved. Our approach is based on transforming the power-constrained channel into an *unconstrained* modulo-lattice additive noise channel, and enhancing the SNR by one using linear minimum mean-square error (MMSE) estimation principles. This improves upon previous lattice-based representations of the AWGN channel in that, for a "good" lattice Λ, the transformation is (asymptotically in the dimension n) information preserving at *any* SNR. The modulo-Λ channel allows to incorporate a coding lattice Λ_1, in a configuration called "nested lattice codes, " $\Lambda \subset \Lambda_1$. The latter is a slight generalization of the concept of *Voronoi constellations* [7], [16].

Conway and Sloane were the first to propose Voronoi constellations, where the Voronoi region of a "self-similar" sublattice replaces the sphere as the shaping region of the lattice code [7], [16]. As will be shown later, there indeed exist lattices with a quasi-spherical Voronoi region having good shaping properties [32], [11]. More general lattice constructions based on multilevel coset codes were proposed in [20]. In our framework, the shaping sublattice Λ is not necessarily self-similar to the coding lattice Λ_1. See also [35], [18], and the references therein for further discussion, links, and applications of this configuration.

A key ingredient in our lattice transmission scheme is common randomness in the form of a *dither* variable $\boldsymbol{U}$, where $\boldsymbol{U}$ is uniformly distributed over the shaping region $\mathcal{V}$, i.e., over the basic Voronoi region of Λ. We subtract the dither from the channel input and add it to the MMSE-estimated channel output $\tilde{X} = \alpha Y$, where addition and subtraction are modulo-Λ, and $\alpha = \frac{P_X}{P_X+P_N}$ is the "Wiener coefficient" which achieves

$$\text{MMSE} = \min_{\alpha} E(\tilde{X} - X)^2 = \frac{P_X P_N}{P_X + P_N}.$$

Dithering is a common randomization technique in *lattice quantization* for source coding, used to assure that the mean-squared quantization error exactly meets the distortion constraint for any source input, and to decorrelate the quantization error from the source [33]. Similarly, in our scheme, the dither assures that the input power exactly meets the power constraint P_X for every codeword, and decorrelates the estimation error from the channel input.[1] Due to that, the effective noise in the dithered modulo-Λ channel is *statistically independent* of the input (although it is slightly non-Gaussian). As a result, ML decoding of the nested lattice code is *equivalent* to lattice decoding. This further implies that for large dimensions, the rate R of the scheme with lattice decoding can approach the mutual information rate of the modulo-Λ channel, which for good shaping lattices approaches

$$\frac{1}{2}\log\Big(\frac{P_X}{\text{MMSE}}\Big) = \frac{1}{2}\log(1+\text{SNR}).$$

[1]Note that the orthogonality principle implies $\tilde{X}-X \perp \tilde{X}$ but not $\tilde{X}-X \perp X$.

We derive our results in several steps. Section II establishes the necessary background on lattice codes. Section III analyzes the Shannon capacity of the modulo-Λ channel which incorporates dithering and linear scaling. Theorem 1 shows that for a good shaping lattice Λ, the capacity of this channel approaches the capacity of the original power-constrained AWGN channel (2). Then, Section IV describes the proposed nested lattice encoding/decoding scheme. Theorems 2 and 3 state our main results, that this scheme can approach capacity for two types of lattice decoders: a *noise-matched* lattice decoder and a *Euclidean* lattice decoder, respectively. The difference between the two follows from the fact that the effective noise in the modulo-Λ channel is not exactly Gaussian, though it approaches Gaussianity for good shaping lattices.

Before turning to the more technical sections which establish this result, we illustrate in Section V the role of linear (biased) estimation in decoding, by a simple example of scalar (uncoded) transmission.

Section VI extends the discussion to random coding error exponents. Theorem 4 shows that the error exponent of the modulo-Λ channel at rate R is at least as good as the Poltyrev exponent for *un*-constrained channels, [28], calculated at a volume-to-noise ratio of $e^{2(C-R)}$. Note that the latter is inferior to the optimal exponent of the power-constrained AWGN channel for rates below capacity. Section VII provides a construction for a "good" random ensemble of nested lattice pairs (Λ, Λ_1). Finally, Theorem 5 in Section VIII makes the last step and proves that the Poltyrev exponent can be achieved by Euclidean lattice decoding of a good nested lattice code, from which Theorems 2 and 3 follow as corollaries. Most of the technical detail is relegated to the appendixes.

Throughout the paper, we use the notation $o_n(1)$ to specify any function of n such that $o_n(1) \to 0$ as $n \to \infty$. In a similar manner, we denote $o_k(1)$, $o_m(1)$, etc. All logarithms in this paper are natural logarithms and rates are in nats.

II. Preliminaries: Lattices, Quantization, Lattice Decoding

A lattice Λ is a discrete subgroup of the Euclidean space $\mathbb{R}^n$ with the ordinary vector addition operation. Thus, if λ_1 λ_2 are in Λ, it follows that their sum and difference are also in Λ. A lattice Λ may be specified in terms of a generating matrix. Thus, an $n \times n$ real-valued matrix G defines a lattice Λ by

$$\Lambda = \{\lambda = G\boldsymbol{x} : \boldsymbol{x} \in \mathbb{Z}^n\}. \tag{4}$$

That is, the lattice is generated by taking all *integer* linear combinations of the basis vectors.

A coset of Λ in $\mathbb{R}^n$ is any translated version of it, i.e., the set $\boldsymbol{x}+\Lambda$ is a coset of Λ for any $\boldsymbol{x} \in \mathbb{R}^n$. The fundamental Voronoi region of $\Lambda \subset \mathbb{R}^n$, denoted by $\mathcal{V}$, is a set of minimum Euclidean norm coset representatives of the cosets of Λ. Every $\boldsymbol{x} \in \mathbb{R}^n$ can be uniquely written as

$$\boldsymbol{x} = \lambda + \boldsymbol{r} \tag{5}$$

with $\lambda \in \Lambda$, $r \in \mathcal{V}$, where $\lambda = Q_{\mathcal{V}}(\boldsymbol{x})$ is a nearest neighbor of $\boldsymbol{x}$ in Λ, and $\boldsymbol{r} = \boldsymbol{x} \bmod \Lambda$ is the apparent error $\boldsymbol{x} - Q_{\mathcal{V}}(\boldsymbol{x})$. We may thus write

$$\mathbb{R}^n = \Lambda + \mathcal{V} = \bigcup_{\lambda \in \Lambda} (\lambda + \mathcal{V}) = \bigcup_{\boldsymbol{x} \in \mathcal{V}} (\Lambda + \boldsymbol{x}) \tag{6}$$

and $\mathcal{V} = \mathbb{R}^n \bmod \Lambda$. For a comprehensive introduction to lattices we refer the reader to [19].

It will prove useful in the sequel to consider more general fundamental regions and quantizers. Let Ω be *any* fundamental region of Λ, i.e., every $\boldsymbol{x} \in \mathbb{R}^n$ can be *uniquely* written as $\boldsymbol{x} = \lambda + \boldsymbol{e}$ where $\lambda \in \Lambda, \boldsymbol{e} \in \Omega$. and $\mathbb{R}^n = \Lambda + \Omega$. We correspondingly define the quantizer associated with Ω by

$$Q_{\Omega}(\boldsymbol{x}) = \lambda, \qquad \text{if } \boldsymbol{x} \in \lambda + \Omega. \tag{7}$$

This is a nearest neighbor quantizer (as above) if we choose Ω to be a fundamental Voronoi region $\mathcal{V}$. But in general there are many other choices for Ω (e.g., the basic parallelepiped [6]), all have the same volume, denoted $V(\Lambda)$, which is given by the inverse density of the lattice points in space. Define the modulo-Λ operation corresponding to Ω as follows:

$$\boldsymbol{x} \bmod_{\Omega} \Lambda = \boldsymbol{x} - Q_{\Omega}(\boldsymbol{x}). \tag{8}$$

Note that this implies that $[\boldsymbol{x} \bmod_{\Omega} \Lambda] \in \Omega$ for all $\boldsymbol{x} \in \mathbb{R}^n$. For a nearest neighbor quantizer, we omit the subscript $\mathcal{V}$, i.e., $\boldsymbol{x} \bmod \Lambda = \boldsymbol{x} - Q_{\mathcal{V}}(\boldsymbol{x})$.

The second moment per dimension associated with Ω is defined as

$$\sigma^2(\Omega) = \frac{1}{n} E\|\boldsymbol{U}\|^2 = \frac{1}{n} \frac{\int_{\Omega} \|\boldsymbol{x}\|^2 d\boldsymbol{x}}{V} \tag{9}$$

where $\boldsymbol{U}$ is a random vector uniformly distributed over Ω and $V \triangleq V(\Lambda) = |\Omega|$. For a fixed lattice, $\sigma^2(\Omega)$ is minimized if we choose Ω as the fundamental Voronoi region $\mathcal{V}$. The normalized second moment of Λ is defined as (see, e.g., [6])

$$G(\Lambda) \triangleq \frac{\sigma^2(\mathcal{V})}{V^{2/n}} = \frac{1}{n} \frac{\int_{\mathcal{V}} \|\boldsymbol{x}\|^2 d\boldsymbol{x}}{V^{1+2/n}}. \tag{10}$$

The normalized second moment $G(\Lambda)$ is always greater than $\frac{1}{2\pi e}$, the normalized second moment of an infinite-dimensional sphere. It is known that for sufficiently large dimension there exist lattices whose Voronoi region $\mathcal{V}$ approaches a sphere in the sense that $G(\Lambda)$ is as close to $\frac{1}{2\pi e}$ as desired [32]. This is equivalent to saying that a random vector $\boldsymbol{U}$ uniform over $\mathcal{V}$ is closer to white Gaussian noise in the sense of normalized entropy, that is $\frac{1}{n} h(\boldsymbol{U})$ is close to $\frac{1}{2}\log 2\pi e \sigma^2(\mathcal{V})$. We say that such lattices are "good for quantization" [35].

A *lattice decoder* is simply a Euclidean quantizer, or more generally, a quantizer with respect to a fundamental region Ω. That is, the decoder quantizes the received vector to obtain the hypothesized codeword. Since most practical decoding algorithms for lattice codes indeed attempt lattice decoding rather than ML decoding, it would be desirable if such lattice decoding were near optimal.

When considering the performance of lattice decoding, it is insightful to consider the similarity to *linear coding* for the binary symmetric channel (BSC). The problems of coding for the BSC and coding for the AWGN channel are widely regarded as analogous to some extent. Both are additive noise channels

$$Y = X + N \tag{11}$$

with addition understood to be modulo-two for the BSC channel and ordinary addition over the reals for the AWGN channel. The BSC coding problem leads to a code in Hamming space, the AWGN coding problem to a code in Euclidean space.

Linear codes are the counterpart of lattices for the case of a BSC, and a minimum Hamming distance decoder is the counterpart of lattice decoding. It is well known that linear codes can achieve not only the capacity of the BSC channel but also the best known exponential bounds on error probability, see, e.g., [2]. Furthermore, for the BSC channel, ML decoding amounts to minimum Hamming distance decoding. Thus, minimum Hamming distance decoding is optimal in the case of a BSC channel.

When trying to take the analogy farther, one is however confronted with a basic problem. In a typical communication scenario over the AWGN channel, the transmitter is usually subject to some constraint, the most common being an average power constraint as in (3). This feature is not present in the BSC/linear case. In the next section, we describe a method for transforming the AWGN into a modulo additive noise channel. This maintains the parallelism between the two channel models and eventually shows that the capacity of the AWGN channel may be achieved using lattice codes and Euclidean lattice decoding.

III. Modulo-Lattice Additive Noise Channel

We describe a technique derived in [12] to transform the power-constrained AWGN channel into a modulo-lattice additive noise (MLAN) channel. The transformation is not strictly information lossless in the sense that it does not preserve the mutual information. However, for a "good" lattice, the (information) loss goes to zero as the dimension of the lattice, n, goes to infinity. This suffices for achieving the channel's capacity, albeit may result in a suboptimal error exponent, as shown in Section VI. For related background, see the treatment of MLAN channels in [20].

Let $\boldsymbol{U}$ be a random variable uniformly distributed over Ω as defined above. We employ $\boldsymbol{U}$ as a dither signal. It is assumed that $\boldsymbol{U}$ is known to both transmitter and receiver (common randomness) and is independent of the channel. The following property is extensively used in the sequel.

Lemma 1: For any random variable $\boldsymbol{X} \in \Omega$, statistically independent of $\boldsymbol{U}$, we have that the sum $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{U} \bmod_{\Omega} \Lambda$ is uniformly distributed over Ω, and is statistically independent of $\boldsymbol{X}$.

A proof in the context of dithered quantization can be found in [33]. The following is a simpler proof by group-theoretic considerations, that was pointed out to the authors by G. D. Forney, Jr.

Proof: Since $\boldsymbol{y} - \boldsymbol{x} \bmod_{\Omega} \Lambda$ runs through Ω as $\boldsymbol{y}$ runs through Ω, and the density $f_{\boldsymbol{U}}(\boldsymbol{u})$ is constant over $\boldsymbol{u} \in \Omega$, the density $f_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}, \boldsymbol{x}) = f_{\boldsymbol{U}}(\boldsymbol{y} - \boldsymbol{x} \bmod_{\Omega} \Lambda)$ is constant over $\boldsymbol{y} \in \Omega$ for any $\boldsymbol{x} \in \Omega$. □

A. The MLAN Channel Transformation

We transform a block of n uses of the AWGN channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}$ into an n-dimensional MLAN channel. The input alphabet of this channel is a fundamental region Ω of the lattice Λ which we call the *shaping lattice*. We later restrict our attention to the fundamental Voronoi region $\mathcal{V}$.

Given $\boldsymbol{t} \in \Omega$ and the dither $\bar{\boldsymbol{U}}$, the output of the transmitter is given by a modulo lattice operation

$$\boldsymbol{X_t} = [\boldsymbol{t} - \boldsymbol{U}] \bmod_{\Omega} \Lambda. \tag{12}$$

Upon reception, $\boldsymbol{Y} = \boldsymbol{X_t} + \boldsymbol{N}$ is multiplied by some "attenuating factor" $0 \leq \alpha \leq 1$ to be specified later, and the dither $\boldsymbol{U}$ is then added. The result is reduced modulo-Λ, giving

$$\boldsymbol{Y'} = [\alpha \boldsymbol{Y} + \boldsymbol{U}] \bmod_{\Omega} \Lambda \tag{13}$$

$$= [\alpha(\boldsymbol{X_t} + \boldsymbol{N}) + \boldsymbol{U}] \bmod_{\Omega} \Lambda. \tag{14}$$

The resulting channel from $\boldsymbol{t}$ to $\boldsymbol{Y'}$ is a modulo-Λ *additive* noise channel described by the following lemma:

Lemma 2 ("Inflated Lattice Lemma" [12]): The channel from $\boldsymbol{t}$ to $\boldsymbol{Y'}$, defined by (1), (12), and (13), is equivalent in distribution to the channel

$$\boldsymbol{Y'} = [\boldsymbol{t} + \boldsymbol{N'}] \bmod_{\Omega} \Lambda \tag{15}$$

where $\boldsymbol{N'}$ is independent of $\boldsymbol{t}$ and is distributed as

$$\boldsymbol{N'} = \left[\alpha \boldsymbol{N} - (1-\alpha)\boldsymbol{U}\right] \bmod_{\Omega} \Lambda \tag{16}$$

where $\boldsymbol{U}$ is a random variable uniformly distributed over Ω and is statistically independent of $\boldsymbol{N}$.

We refer to the resulting channel as a Λ-MLAN channel. The component $-(1-\alpha)\bar{\boldsymbol{U}}$ will be termed "self-noise" in the sequel. We see that the equivalent noise is the weighted sum of a Gaussian vector and a uniform random vector, folded (aliased) into the fundamental region Ω. When $\alpha < 1$, the MLAN transformation amounts to effectively "inflating" the lattice and scaling the noise by different factors as explained in Section V.

Proof:

$$\boldsymbol{Y'} = [\alpha(\boldsymbol{X_t} + \boldsymbol{N}) + \boldsymbol{U}] \bmod_{\Omega} \Lambda \tag{17}$$

$$= [\boldsymbol{X_t} + \boldsymbol{U} + (\alpha - 1)\boldsymbol{X_t} + \alpha \boldsymbol{N}] \bmod_{\Omega} \Lambda \tag{18}$$

$$= [(\boldsymbol{t} - \bar{\boldsymbol{U}}) \bmod_{\Omega} \Lambda + \bar{\boldsymbol{U}} - (1-\alpha)\boldsymbol{X_t} + \alpha \boldsymbol{N}] \bmod_{\Omega} \Lambda \tag{19}$$

$$= [\boldsymbol{t} - (1-\alpha)\boldsymbol{X_t} + \alpha \boldsymbol{N}] \bmod_{\Omega} \Lambda \tag{20}$$

where (20) follows since the modulo operation is distributive so the dither cancels out. The lemma follows, since the distribution of $\boldsymbol{X_t}$ is independent of $\boldsymbol{t}$ by Lemma 1, and it has the same distribution as $\boldsymbol{U}$, i.e., it is uniform over Ω. □

For an input *power* constraint, the best choice for a shaping region Ω is a fundamental *Voronoi* region of the lattice relative to the Euclidean norm. We denote this choice by $\Omega = \mathcal{V}$. Note that $\mathcal{V} = -\mathcal{V}$ (up to a boundary set of measure zero) and therefore in this case

$$\boldsymbol{N'} = [(1-\alpha)\boldsymbol{U} + \alpha \boldsymbol{N}] \bmod \Lambda \tag{21}$$

where $\bmod \Lambda$ means $\bmod_{\mathcal{V}} \Lambda$.

The lattice is scaled so that the second moment of $\mathcal{V}$ is P_X, i.e.,

$$\frac{1}{nV} \int_{\mathcal{V}} ||\boldsymbol{x}||^2 d\boldsymbol{x} = \sigma^2(\mathcal{V}) = P_X. \tag{22}$$

By Lemma 1, due to the dither, for any $\boldsymbol{t}$, the average transmitted power is

$$\frac{1}{n} E||\boldsymbol{X_t}||^2 = \frac{1}{n} E||\boldsymbol{U}||^2 = P_X. \tag{23}$$

B. Capacity of the MLAN Channel

Since the equivalent channel (15) is additive modulo-Λ, taking the input to be uniform over the Voronoi region of Λ, i.e., $\boldsymbol{T} \sim \mathrm{Unif}(\mathcal{V})$, achieves its capacity. With this choice, the output $\boldsymbol{Y'}$ is also uniformly distributed over $\mathcal{V}$. The resulting information rate is

$$\frac{1}{n} I(\boldsymbol{T}; \boldsymbol{Y'}) = \frac{1}{n} h(\boldsymbol{Y'}) - \frac{1}{n} h(\boldsymbol{Y'}|\boldsymbol{T}) \tag{24}$$

$$= \frac{1}{n} \log V - \frac{1}{n} h(\boldsymbol{N'}) \tag{25}$$

$$= \frac{1}{2} \log \frac{P_X}{G(\Lambda)} - \frac{1}{n} h(\boldsymbol{N'}) \tag{26}$$

where (26) follows from the definition of the normalized second moment (10) and from (22).

We are still left with the freedom of choosing α. Choosing $\alpha = 1$ results in an effective noise $\boldsymbol{N'} = \boldsymbol{N} \bmod \Lambda$ in (21), i.e., $\boldsymbol{N'}$ does not have a self-noise component. When n is large and $P_X \gg P_N$, and if Λ is a "good" lattice for quantization, i.e., $G(\Lambda) \approx \frac{1}{2\pi e}$, it can be shown that the effect of the modulo operation on the noise entropy becomes negligible. We would therefore have $\frac{1}{n} h(\boldsymbol{N'}) \approx \frac{1}{n} h(\boldsymbol{N})$ and a resulting information rate[2] of $\frac{1}{2} \log \frac{P_X}{P_N}$. As mentioned in the Introduction, this rate was previously conjectured to be the greatest achievable with lattice decoding.

Nevertheless, we can do better by taking the MMSE coefficient

$$\alpha = \frac{P_X}{P_X + P_N} = \frac{\mathrm{SNR}}{1 + \mathrm{SNR}}.$$

With this choice, we have

$$\frac{1}{n} E||\boldsymbol{N'}||^2 \leq \frac{1}{n} E||(1-\alpha)\boldsymbol{U} + \alpha \boldsymbol{N}||^2 \tag{27}$$

$$= (1-\alpha)^2 P_X + \alpha^2 P_N \tag{28}$$

$$= \frac{P_X P_N}{P_X + P_N} \tag{29}$$

where the inequality follows since for a Voronoi region $||\boldsymbol{x} \bmod \Lambda|| \leq ||\boldsymbol{x}||$ for any $\boldsymbol{x}$. Therefore, the effective noise power is reduced by a factor of $\frac{P_X}{P_X + P_N}$ (as if the noise were attenuated by a factor of $\sqrt{\frac{1+\mathrm{SNR}}{\mathrm{SNR}}}$), so the effective SNR of the MLAN channel is at least

$$\frac{P_X}{P_N \cdot \frac{\mathrm{SNR}}{1+\mathrm{SNR}}} = 1 + \mathrm{SNR}$$

so that it is increased by one.

[2]It is interesting to note that the information rate of $\frac{1}{2} \log(2\pi e P_X) - h(N)$ is achievable with $\alpha = 1$ even when $N = Y - X$ is not independent of X.

Consider now a sequence of lattices $\Lambda^{(n)}$ which are good for quantization as defined above, that is, $\lim_{n\to\infty} G(\Lambda^{(n)}) = \frac{1}{2\pi e}$.

Theorem 1 (Capacity of MLAN Channel): For the MLAN channel, if we choose $\boldsymbol{T} \sim \mathrm{Unif}(\mathcal{V})$, $\alpha = \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$, and if the sequence of lattices $\Lambda^{(n)}$ satisfies $G(\Lambda^{(n)}) \to \frac{1}{2\pi e}$, then

$$\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{T};\boldsymbol{Y}') = \frac{1}{2}\log(1+\mathrm{SNR}).$$

Proof: Since the capacity of the original AWGN channel is $C \triangleq \frac{1}{2}\log(1+\mathrm{SNR})$, it follows from the data processing inequality that

$$\frac{1}{n} I(\boldsymbol{T};\boldsymbol{Y}') \le \frac{1}{2}\log(1+\mathrm{SNR}). \tag{30}$$

Since the entropy of $\boldsymbol{N}'$ is upper-bounded by the the entropy of a white Gaussian vector with the same second moment [22], we have from (27) that

$$\frac{1}{n} h(\boldsymbol{N}') \le \frac{1}{2}\log\left(2\pi e \frac{P_X P_N}{P_X + P_N}\right) \tag{31}$$

which implies from (26)

$$\begin{aligned}\frac{1}{n} I(\boldsymbol{T};\boldsymbol{Y}') &\ge \frac{1}{2}\log\left(\frac{1}{2\pi e G(\Lambda^{(n)})} \cdot \frac{P_X + P_N}{P_N}\right) \\ &= \frac{1}{2}\log(1+\mathrm{SNR}) - \frac{1}{2}\log\left(2\pi e G(\Lambda^{(n)})\right). \end{aligned} \tag{32}$$

By assumption $\log\left(2\pi e G(\Lambda^{(n)})\right) \to 0$. Thus combining (30) and (32), the theorem follows. □

Therefore, with $\alpha = \alpha_{\mathrm{MMSE}}$ and a proper choice of shaping lattice Λ, the capacity of the MLAN channel indeed approaches the capacity of the original power constrained AWGN channel[3]. This entails drawing a *random code* according to the distribution $\boldsymbol{T} \sim \mathrm{Unif}(\mathcal{V})$, and applying ML decoding relative to the effective modulo-noise $\boldsymbol{N}'$ [22]. In the next section, we show how to replace the uniform random code by a lattice code.

IV. Nested Lattices for Shaping and Coding

As in the case of a BSC channel, we shall see that it is possible to achieve capacity using *linear* codes instead of a code drawn at random. For the MLAN channel, this means using a nested lattice code, where the coarse lattice Λ is used for shaping so it is a good quantizer, and the fine lattice Λ_1 defines the codewords so it is a good channel code. Furthermore, for the MLAN channel *lattice decoding* is optimal, so that we will obtain a lattice encoding/decoding scheme to replace the random-code/ML-decoding scheme of Section III, having the same capacity. The scheme is described in Section IV-A below, and its optimality is stated in Theorem 2 in Section IV-B.

A delicate point is, however, that the effective noise in the MLAN channel is not precisely Gaussian for any finite dimension; hence, lattice decoding no longer means Euclidean decoding but rather decoding with a noise-matched "metric"[4]. Nevertheless, for a more restricted class of nested lattices (see Sections VII and VIII), Euclidean lattice decoding becomes asymptotically optimal as the dimension goes to infinity, hence it achieves capacity as well. This result is formally stated in Theorem 3 in Section IV.C.

A nested lattice code is a lattice code whose boundary region is the Voronoi region of a sublattice. This may be visualized as in Fig. 1. The use of nested lattices goes back to the works of Conway and Sloane [7] and Forney [16] (where they were called "Voronoi codes" or "Voronoi constellations").[5] More recently, such codes found application in Wyner–Ziv and dirty paper encoding [35].

The shaping sublattice (i.e., the coarse lattice) is Λ, the lattice defining the MLAN channel. We will choose Λ so that its average power per dimension is P_X and its normalized second moment approaches that of a sphere, namely, $\frac{1}{2\pi e}$. The fine lattice should be good for channel coding, i.e., it should achieve the Poltyrev exponent, as explained in Section VII.

Formally, we say that a lattice Λ (the coarse lattice) is nested in Λ_1 (the fine lattice) if $\Lambda \subseteq \Lambda_1$, i.e., if Λ is a sublattice of Λ_1.[6] The fundamental Voronoi regions of Λ_1 and Λ are denoted by $\mathcal{V}_1$ and $\mathcal{V}$, respectively; their corresponding volumes by V_1 and V, where V_1 divides V by construction. We call $\left(\frac{V}{V_1}\right)^{\frac{1}{n}}$ the *nesting ratio*. The points of the set

$$\mathcal{C} = \{\Lambda_1 \bmod \Lambda\} \triangleq \{\Lambda_1 \cap \mathcal{V}\} \tag{33}$$

are called the *coset leaders* of Λ relative to Λ_1; for each $\boldsymbol{c} \in \mathcal{C}$, the shifted lattice $\Lambda_{\boldsymbol{c}} = \boldsymbol{c} + \Lambda$ is called a *coset* of Λ relative to Λ_1. The set of all cosets, i.e., the quotient group of Λ_1 by Λ, is denoted by Λ_1/Λ. It follows that there are V/V_1 different cosets, whose union gives the fine lattice

$$\bigcup_{\boldsymbol{c}\in\mathcal{C}} \Lambda_{\boldsymbol{c}} = \Lambda_1. \tag{34}$$

The coding rate of the nested lattice code is defined as

$$R = \frac{1}{n}\log|\mathcal{C}| = \frac{1}{n}\log|\Lambda_1/\Lambda|.$$

It follows that

$$R = \frac{1}{n}\log\frac{V}{V_1} = \log(\text{nesting ratio}). \tag{35}$$

A. Encoding/Decoding Scheme

We now incorporate a lattice code into the modulo transformation scheme of Section III, with nested lattice codes replacing the random codebook, as shown in Fig. 2. Let (Λ_1, Λ) be a rate-R nested lattice code as defined in (35), with $\sigma^2(\mathcal{V}) = P_X$. Let $\bmod \Lambda$ denote modulo-lattice operation with respect to the *Voronoi* region $\mathcal{V}$ of the coarse lattice. Let Ω_1 denote some fundamental region of the fine lattice Λ_1 to be specified later, and let Q_{Ω_1} denote the corresponding lattice quantizer.

[3]Inspired by a preprint of our work, Forney suggested to view this as a canonical model which connects between Wiener theory and Shannon theory [18].

[4]We use here the (popular) term "decoding metric" although the distance measure induced by ML decoding is not necessarily a metric.

[5]Conway and Sloane's original definition [7] was limited to self-similar lattices. Forney's Voronoi codes, allow, by construction, any nesting relation. Here we prefer to use the name "nested codes," which links to the more general context of algebraic binning [35].

[6]In some publications, the coarse lattice is denoted Λ_s (for shaping) or Λ_q (for quantization), while the fine lattice is denoted Λ_c (for coding).

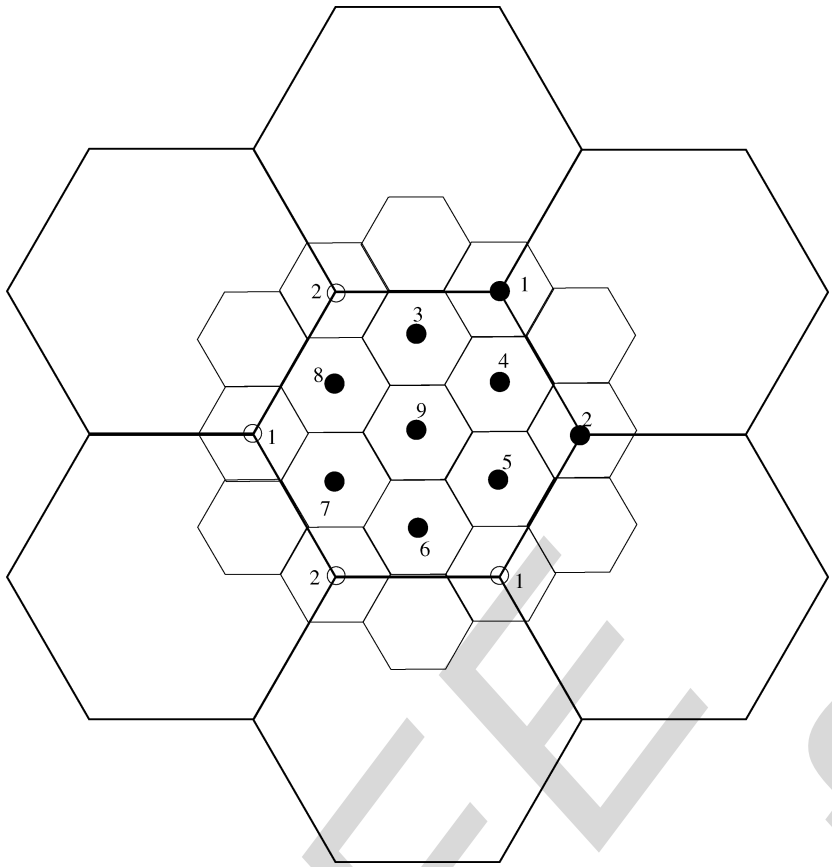

Fig. 1. Nested lattices of ratio three.

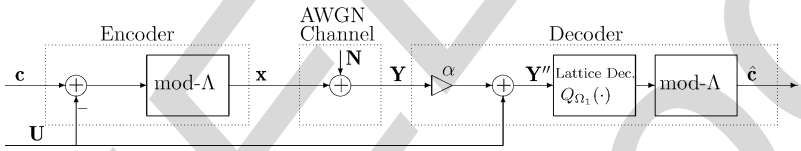

Fig. 2. Encoding/decoding scheme.

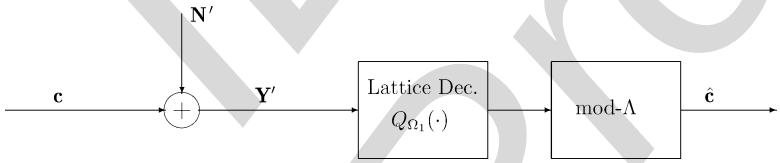

Fig. 3. Equivalent modulo-additive noise channel. Addition is modulo-Λ.

- *Message selection:* Associate a message with each member of the set of coset leaders $\mathcal{C} = \{\boldsymbol{c}\}$ as defined in (33).[7]
- *Encoding:* Let the dither $\boldsymbol{U}$ be defined by $\boldsymbol{U} \sim \mathrm{Unif}(\mathcal{V})$. Given the message $\boldsymbol{c} \in \mathcal{C}$, the encoder sends
$$\boldsymbol{X} = [\boldsymbol{c} - \boldsymbol{U}] \bmod \Lambda. \tag{36}$$
Consequently, by Lemma 1 and (23), $\boldsymbol{X}$ is uniform over $\mathcal{V}$ (independent of $\boldsymbol{c}$) and the average transmitted power is P_X.
- *Decoding:* Let $\alpha = \frac{P_X}{P_X + P_N}$ as in Section III. The decoder computes
$$\hat{\boldsymbol{c}} = Q_{\Omega_1}(\alpha \boldsymbol{Y} + \boldsymbol{U}) \bmod \Lambda. \tag{37}$$

[7]In fact, $\boldsymbol{c}$ may be replaced by any member of the coset $\Lambda_{\boldsymbol{c}}$.

This transmission scheme is depicted in Fig. 2, where $\alpha\boldsymbol{Y} + \boldsymbol{U}$ is denoted by $\boldsymbol{Y}''$. By the "distributive" property of the modulo operation, we can rewrite (37) as

$$\hat{\boldsymbol{c}} = Q_{\Omega_1}([\alpha \boldsymbol{Y} + \boldsymbol{U}] \bmod \Lambda) \bmod \Lambda \tag{38}$$
$$= Q_{\Omega_1}(\boldsymbol{Y}') \bmod \Lambda \tag{39}$$
$$= Q_{\Omega_1}\left([\boldsymbol{c} + \boldsymbol{N}'] \bmod \Lambda\right) \bmod \Lambda \tag{40}$$

where (39) follows since $\boldsymbol{Y}' = \boldsymbol{Y}'' \bmod \Lambda$ from (13) (with $\Omega = \mathcal{V}$), and (40) follows by the inflated lattice lemma (Lemma 2) where $\boldsymbol{N}' = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N} \bmod \Lambda$. The equivalent channel from $\boldsymbol{c}$ to $\hat{\boldsymbol{c}}$ is illustrated in Fig. 3.

Since the channel is modulo additive and Λ is nested in Λ_1, the decoding error probability for *any* codeword $\boldsymbol{c}$ is given by

$$P_e = \Pr(\boldsymbol{N}' \notin \Omega_1). \tag{41}$$

B. Noise-Matched Lattice Decoding

Since we use Λ_1 (or more precisely $\mathcal{C}$) as a channel code for the MLAN channel with noise $\boldsymbol{N}'$ which is not Gaussian (or spherically symmetric), the optimal decoding region Ω_1 is not the Voronoi region of Λ_1 with respect to Euclidean metric. Rather, we define Ω_1^*, a fundamental region of Λ_1, to be a ML decoding region with respect to $\boldsymbol{N}'$ of the zero codeword. Thus, Ω_1^* is a fundamental region satisfying[8]

$$\Omega_1^* = \left\{\boldsymbol{x} : f_{\boldsymbol{N}'}(\boldsymbol{x}) \geq f_{\boldsymbol{N}'}(\boldsymbol{x}-\boldsymbol{c} \bmod \Lambda)\ \forall \boldsymbol{c} \in \mathcal{C}\right\}. \quad (42)$$

A decoder using the quantizer $Q_{\Omega_1^*}(\cdot)$ will be called an ML lattice decoder or a noise-matched lattice decoder. Note that the decoder is a lattice decoder in the sense that the decoding regions are congruent but is not a (Euclidean) nearest neighbor decoder since $\boldsymbol{N}'$ is not quite spherically symmetric.

Theorem 2 (Capacity-Achieving Nested Lattices With ML Lattice Decoding): For any $\epsilon > 0$, there exists a sequence of n-dimensional nested lattice pairs $(\Lambda_1^{(n)}, \Lambda^{(n)})$ whose rate R as defined in (35) is greater than $C - \epsilon$ for sufficiently large n, and whose decoding error probability (41) vanishes as $n \to \infty$

$$P_e = \Pr(\boldsymbol{N}' \notin \Omega_1^{(n)*}) \to 0. \quad (43)$$

Theorem 2 can be deduced by a suitable modification (to nested lattices) of the analysis in [26]. Here we obtain it as a corollary to Theorem 5, which deals with the error probability of a *Euclidean* decoder. The latter is strictly inferior to the noise-matched decoder assumed here and thus Theorem 2 indeed follows from Theorem 5.

C. Euclidean Lattice Decoding

As we observed, a somewhat disagreeable aspect of the noise-matched lattice decoder is that the decoding "metric" is now coupled to the choice of shaping lattice Λ (via the probability density of the self-noise $\boldsymbol{U}$). Moreover, the decoding metric has memory. This is in contrast to the single-letter form of the Euclidean decoding metric, corresponding to white Gaussian noise.

Looking at the definition of $\boldsymbol{N}'$ (21), we see that there are two elements to this non-Euclidean nature of the decoder which we may separate. Define $\boldsymbol{N}'' = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N}$ so that $\boldsymbol{N}' = \boldsymbol{N}'' \bmod \Lambda$. The first element is that the self-noise is distributed uniformly over $\mathcal{V}$ rather than being Gaussian. The second is that the sum $\boldsymbol{N}'' = (1-\alpha)\boldsymbol{U} + (1-\alpha)\boldsymbol{N}$ is then reduced modulo-Λ. We may correspondingly depict the operation of a noise-matched decoder as follows. Upon receiving a vector $\boldsymbol{y}'' = \alpha\boldsymbol{y} + \boldsymbol{u}$, for every codeword $\boldsymbol{c} \in \mathcal{C}$ first compute the densities $f_{\boldsymbol{N}''}(\boldsymbol{y}'' - \boldsymbol{c} + \lambda)$ for all $\lambda \in \Lambda$, then sum them. That is, the metric associated with codeword $\boldsymbol{c}$ is the sum over all metrics of its coset $\{\boldsymbol{c} + \Lambda\}$, so that

$$\Pr(\boldsymbol{c}|\boldsymbol{y}'') \propto \sum_{\lambda\in\Lambda} f_{\boldsymbol{N}''}(\boldsymbol{y}'' - \boldsymbol{c} + \lambda) = \mu(\boldsymbol{y}'' - \boldsymbol{c}) \quad (44)$$

where

$$\mu(\boldsymbol{x}) \triangleq \sum_{\lambda\in\Lambda} f_{\boldsymbol{N}''}(\boldsymbol{x} + \lambda). \quad (45)$$

Accordingly, there are two natural simplified (suboptimal) decoders to be considered. First, we may approximate $\boldsymbol{N}''$ with a white Gaussian vector $\boldsymbol{Z}$ having the same second moment $\frac{P_X P_N}{P_X + P_N}$, and thus use the "folded Euclidean metric"

$$\mu'(\boldsymbol{x}) = -\log \sum_{\lambda\in\Lambda} \exp\left\{\frac{-\|\boldsymbol{x}-\lambda\|^2}{2\cdot\frac{P_X P_N}{P_X+P_N}}\right\}. \quad (46)$$

The decoder may further be simplified by dropping the sum, keeping only the largest term, resulting in the metric

$$\mu^*(\boldsymbol{x}) \triangleq \min_{\lambda\in\Lambda} \|\boldsymbol{x}-\lambda\|^2 = \|\boldsymbol{x} \bmod \Lambda\|^2. \quad (47)$$

The metric $\mu^*(\cdot)$ gives rise to a Euclidean quantization cell $\Omega_1 = \mathcal{V}_1$, so the decoding operation in (37) becomes $\hat{\boldsymbol{c}} = Q_{\mathcal{V}_1}(\alpha\boldsymbol{Y} + \boldsymbol{U}) \bmod \Lambda$, and the decoding error probability (41) becomes

$$P_e = \Pr\{\boldsymbol{N}' \notin \mathcal{V}_1\}. \quad (48)$$

Since decoding according to μ^* is suboptimal (i.e., mismatched decoding), P_e in (48) is in general larger than the decoding error probability in (43). Nevertheless, the following theorem shows that capacity can still be approached using appropriate nested lattice pairs.

Theorem 3 (Capacity-Achieving Nested Lattices With Euclidean Decoding): For any $\epsilon > 0$, there exists a sequence of n-dimensional nested lattice pairs $(\Lambda_1^{(n)}, \Lambda^{(n)})$ whose rate R as defined in (35) is greater than $C - \epsilon$ for sufficiently large n, and whose decoding error probability (48) satisfies as $n \to \infty$

$$P_e = \Pr(\boldsymbol{N}' \notin \mathcal{V}_1) \to 0. \quad (49)$$

Theorem 3 follows as a corollary to Theorem 5 in Section VIII, which goes further and bounds the *error exponent* of a nested lattice code with Euclidean lattice decoding. As a final remark, we note that, in practice, the folded Euclidean metric (46) may allow to approach capacity with a less demanding nested lattice construction and may be advantageous in practice, see [13], [14].

V. Linear Estimation, Bias and Inflated Lattice Decoding

In this section, we illustrate the effect of using an inflated lattice decoder[9] by considering a one-dimensional example, without the use of a dither. We consider uncoded pulse amplitude modulation (PAM) transmission and compare the average error probability of an inflated (or scaled) lattice decoder with that of a noninflated lattice decoder. The inflated lattice decoding approach relates to the issue of *biased* versus *unbiased* estimation in the context of detection. We attempt to shed some light on the merits of these two approaches.

[8]Note that Ω_1^* is not uniquely defined by (42) as ties may be broken in different ways. One possibility is to take Ω_1^* to be the union of all points either satisfying (42) with strict inequality, or in case of a tie, belonging to the Voronoi region $\mathcal{V}_1$.

[9]The term "inflated" will be explained later in this section.

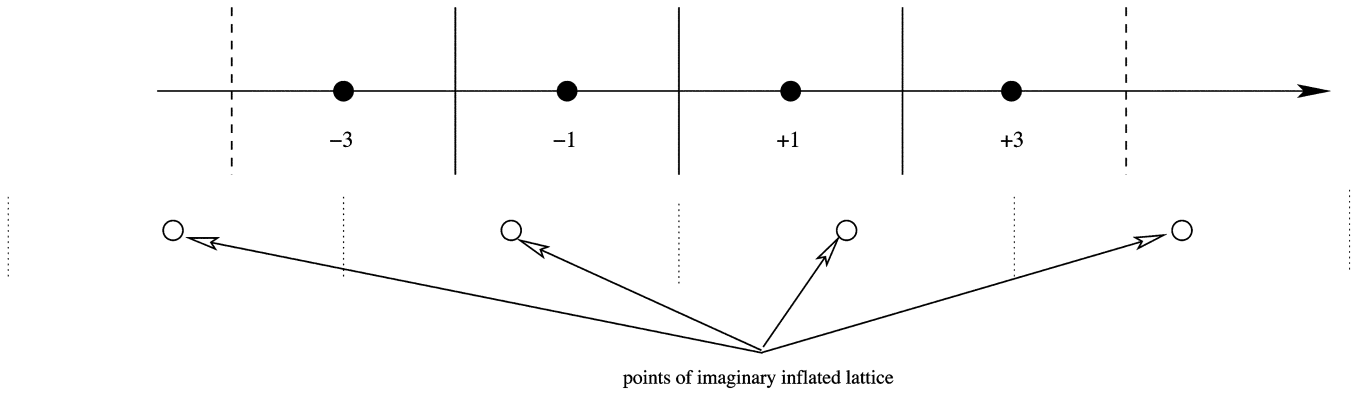

Fig. 4. Regular slicer (black circles and solid lines), lattice quantizer decoder (black circles, solid and dashed lines), and inflated lattice quantizer decoder (empty circles and dotted lines).

Assume an element of a 4-PAM constellation is sent over an additive noise channel with AWG noise as illustrated in Fig. 4. That is, let $X \in \{-3,-1,1,3\}$ and assume that the four symbols are equiprobable, i.e., $\Pr\{X = i\} = \frac{1}{4}$ for $i \in \{-3,-1,1,3\}$. The receiver observes $Y = X + N$ where $N \sim \mathcal{N}(0, P_N)$. A minimum-distance (ML) decoder would decode as follows:

$$\hat{X} = \begin{cases} -3, & Y < -2 \\ -1, & -2 \le Y < 0 \\ 1, & 0 \le Y < 2 \\ 3, & Y \ge 2. \end{cases}$$

These decision regions correspond to a standard slicer. The resulting average probability of error is

$$\bar{P}_e = \frac{1}{4} \sum_{i=-3,-1,1,3} \Pr\{\hat{X} \neq i | X = i\} \tag{50}$$

$$= 3/2 \cdot q(1/\sqrt{P_N}) \tag{51}$$

where

$$q(t) = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \tag{52}$$

denotes the standard Q function to avoid confusion with the quantizer function $Q_\Delta(\cdot)$ to be used next.

Suppose now that we replace the slicer with a one-dimensional midrise lattice quantizer $Q_\Delta(\cdot)$ of step size $\Delta = 2$, so that $\hat{X} = Q_\Delta(Y)$ where

$$Q_\Delta(Y) = i \qquad \text{iff } 2Y\Delta \in [i-1, i+1), \text{ for } i \in 2\mathbb{Z}+1.$$

Hence, the two outer decision boundaries in Fig. 4 come into play. The average probability of error of this system is $\bar{P}_e = 2 \cdot q(1/\sqrt{P_N})$. We obviously lose with respect to the ML decoder by bounding the decision regions of the two outer symbols -3 and 3.

Consider now a third decoder that uses a one-dimensional lattice quantizer but this time with step size Δ/α where $0 < \alpha < 1$, so that $\hat{X} = Q_{\Delta/\alpha}(Y)$, as illustrated in Fig. 4. We call such a decoder an *inflated lattice* decoder or simply an inflated quantizer. We may optimize the scaling coefficient α so as to minimize the averaged error probability

$$\bar{P}_e = \frac{1}{4} \sum_{i \in \{-3,-1,1,3\}} \Pr(Q_{\Delta/\alpha}(Y) \neq i | X = i). \tag{53}$$

We may alternatively view the inflated lattice decoder as using linear estimation prior to quantization. That is, we may keep the step size $\Delta = 2$ and decode as

$$\hat{X} = Q_\Delta(\tilde{X})$$

where $\tilde{X} \triangleq \alpha Y$ is a linear estimator of X given Y. Note that the estimator is *biased* and the estimation error $D = \tilde{X} - X$ is statistically dependent on the transmitted symbol X (as no dither is used). The optimizing scaling factor $\alpha_{\text{opt}}(\text{SNR})$ may be found numerically.

A suboptimal choice, at least for this one dimensional example, is to use MMSE scaling. The MMSE criterion chooses α so as to minimize the expected estimation error $E\{D^2\}$. The resulting (Wiener) coefficient is $\alpha_{\text{MMSE}}(\text{SNR}) = \frac{\text{SNR}}{1+\text{SNR}}$. Fig. 5 compares the average error probability as a function of the SNR of the a regular slicer, a lattice quantizer, and a scaled (inflated) lattice quantizer. The performance of the inflated lattice quantizer is depicted for various values of α as well as for the MMSE value $\alpha_{\text{MMSE}}(\text{SNR})$. The lower envelope of the dashed lines corresponds to $\alpha_{\text{opt}}(\text{SNR})$. It is seen that the inflated lattice quantizer has a substantial gain over the standard (unscaled) lattice quantizer at low SNR.

We conclude that when using a lattice quantizer at the decoder, we may gain by using a *biased* linear estimator prior to quantization. In contrast, when a regular slicer (with half open boundary decision regions) is used, the *unbiased* estimator is clearly superior as it performs ML detection.

In this one-dimensional example, the distinction between a minimum-distance decoder and a strict lattice quantizer decoder may seem of minor significance, affecting only the boundary points. However, in high dimensions, the boundary codewords are typical and the distinction becomes of central importance. This may be visualized by the multidimensional lattice transmission scheme with inflated lattice decoder depicted in Fig. 6. The noise is Gaussian and is depicted as a sphere. In high dimensions, almost all transmitted lattice points lie near the surface of a sphere of radius $\sqrt{nP_X}$. One such point is considered in the figure. With a nonscaled lattice decoder, correct decoding occurs when the noise falls within the Voronoi region. When the noise is large, this original Voronoi region is completely contained in and is strictly smaller than the noise sphere.

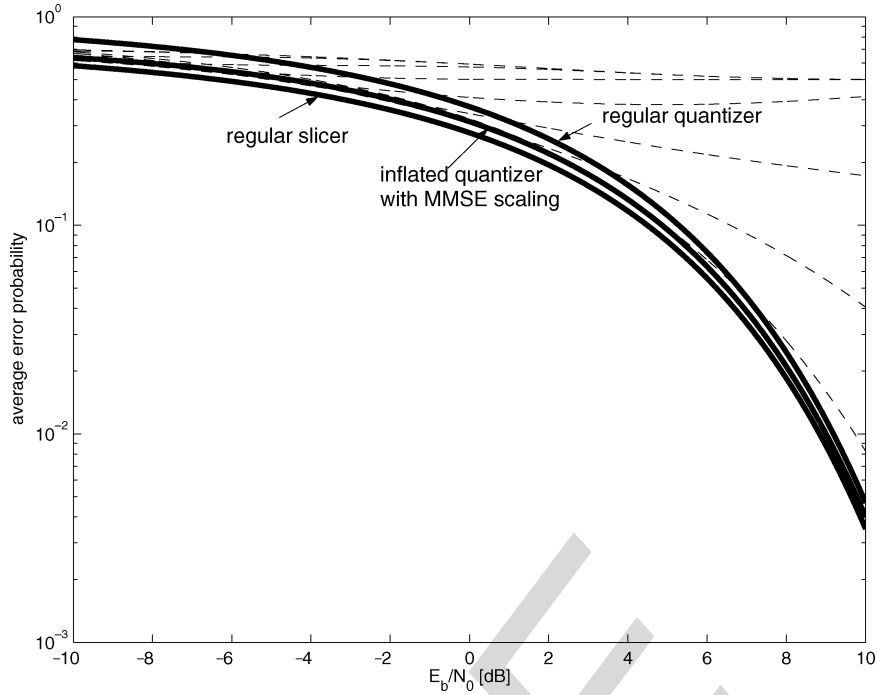

Fig. 5. Comparison of performance of regular slicer, regular (unscaled) lattice quantizer, and inflated lattice quantizer with MMSE scaling. The dashed lines correspond to inflated lattice quantizers with fixed values of $\alpha = 0.1, 0.2, \ldots, 1$.

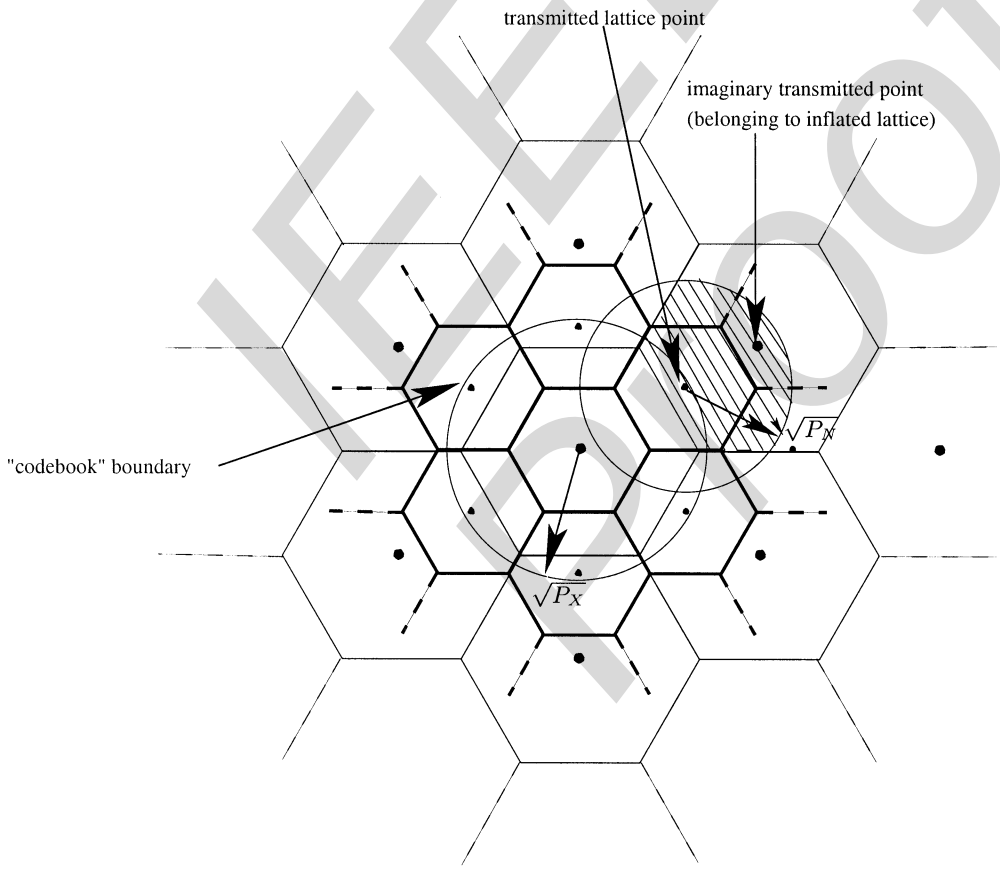

Fig. 6. Inflated lattice decoder. Bold lines = Voronoi regions of original lattice code; thin lines = decoding regions of the inflated lattice; small circles = actual transmitted codewords; bold circles = inflated (imaginary) codebook.

Thus, the probability of correct decoding is proportional to the relative portion of the noise sphere contained inside the original Voronoi region, which is strictly smaller than one. With an *inflated lattice* decoder, the probability of correct decoding

is proportional to the relative portion of the noise sphere (centered at the transmitted codeword) contained inside the *inflated* Voronoi region. The latter region is centered around the associated point of the inflated lattice, which we call "the imaginary transmitted point." We can see in the figure that the intersection of the noise sphere with the inflated Voronoi region has a larger volume than the volume of the original Voronoi region. Thus, the inflated lattice decoder has a smaller probability of error. Note also that if we were to increase the scaling ratio, at some point the noise sphere would cease to intersect the inflated Voronoi region. Thus, the optimal scaling ratio is finite. Furthermore, $\alpha_{\rm opt}(\text{SNR}) \to \alpha_{\rm MMSE}(\text{SNR})$ as the dimension goes to infinity.

VI. Random Coding Error Exponents of the MLAN Channel and the Poltyrev Exponent

We now show that the random coding error exponents of the MLAN channel are related to the Poltyrev exponent [28], [20]. We first give a heuristic explanation of the relation, arriving at an expression for the MLAN error exponent. A rigorous derivation is then given in Section VI-A.

Poltyrev studied the problem of coding for the unconstrained AWGN channel with the input alphabet being the whole space $\mathbb{R}^n$. In this setting, the notion of capacity becomes meaningless as infinite rates of transmission are possible. Instead, the error probability (of an ML decoder) is measured against the normalized density of the codewords. We now formalize these notions.

Let $\mathcal{C} \subset \mathbb{R}^n$ be an infinite constellation of points (codewords) and let $L\text{-CUBE} = [-L/2, L/2]^n$ be an n-dimensional cube of side length L centered at the origin. Denote by

$$\delta = \limsup_{L\to\infty} \frac{|\mathcal{C} \cap (L\text{-CUBE})|}{|L\text{-CUBE}|}$$

the density of the constellation. Note that $V_c \triangleq 1/\delta$ is the average volume of a Voronoi region of a codeword. Given an AWGN of variance P_N the (normalized per dimension) volume-to-noise ratio (VNR)[10] μ is defined as

$$\mu = \frac{V_c^{2/n}}{2\pi e} \Big/ P_N. \tag{54}$$

Note that $\frac{V_c^{2/n}}{2\pi e} \cdot n$ is the asymptotic (in dimension n) squared radius of a sphere of volume V_c. Thus, μ has the significance of the ratio of the squared "radius of a spherical Voronoi region" to the variance of the noise. When $\mu = 1$, a "spherical" Voronoi region has the same radius as the standard deviation of the noise; for smaller μ, an error is highly likely and reliable communication is not to be expected. Thus, $\mu = 1$ has the significance of capacity. See the discussion in [20, Sec. II-C] where this is referred to as the "sphere bound."

[10]The term VNR was coined in [20] where it is denoted by α^2. In [28] $2\pi e\mu$ is called the generalized signal to noise ratio and is denoted by μ, i.e, Poltyrev's μ differs from ours by a factor of $2\pi e$.

Define $\bar{P}_e(\mathcal{C})$ to be the limit supremum (over L) of the average probability of error of the codewords within L-CUBE, the size of the *cube* going to infinity. Denote the best possible average probability of error for a given μ by

$$\bar{P}_e(\mu) = \inf_{\mathcal{C}} \bar{P}_e(\mathcal{C})$$

where the infimum is over all codebooks with VNR μ. Poltyrev showed in [28] that

$$\bar{P}_e(\mu) \le e^{-nE_P(\mu)} \tag{55}$$

where $E_P(\mu)$, the "Poltyrev exponent," is given by

$$E_P(\mu) = \begin{cases} E_P^r(\mu) = \frac{1}{2}\left[(\mu-1) - \log\mu\right], & 1 < \mu \le 2 \\ E_P^r(\mu) = \frac{1}{2}\log\frac{e\mu}{4}, & 2 \le \mu \le 4 \\ E_P^x(\mu) = \frac{\mu}{8}, & \mu \ge 4 \end{cases} \tag{56}$$

and corresponds, as for finite capacity channels, to the random coding and expurgated bounds on the error exponent [22].

The problem of coding for the MLAN channel is rather similar to Poltyrev's problem of coding for the unconstrained AWGN channel. Whereas in the first problem the alphabet is compact, i.e., it is the Voronoi region $\mathcal{V}$ of a lattice, in the latter it is unbounded, i.e., the entire Euclidean space $\mathbb{R}^n$. Thus, we might suspect that the decoding error probability in the two problems may be related if we measure it in both cases against codeword density. A minor difference is that the noise in the MLAN channel is not strictly Gaussian but rather approaches a Gaussian distribution asymptotically as the dimension $n \to \infty$ (with a proper choice of a sequence of shaping lattices).

Consider a code of rate R for the MLAN channel with a fundamental Voronoi region of volume $V = |\mathcal{V}|$ and with

$$\alpha = \alpha_{\rm MMSE} = \frac{P_X}{P_X + P_N}.$$

The number of codewords is e^{nR} and thus the volume per codeword is $V_c = V/e^{nR}$ giving a codeword density

$$\delta = 1/V_c = \frac{e^{nR}}{V}. \tag{57}$$

As the effective noise has variance

$$\alpha P_N = \frac{P_X P_N}{P_X + P_N}$$

we may associate with the code and channel a corresponding *effective* VNR

$$\mu_{\rm eff} = \frac{V_c^{2/n}}{2\pi e \frac{P_X P_N}{P_X + P_N}} = \frac{1+\text{SNR}}{e^{2R}} = e^{2(C-R)} \tag{58}$$

using the fact that for a (high-dimensional) almost-spherical Voronoi region $\mathcal{V}$, we have $V \approx (2\pi e P_X)^{n/2}$. Note that $\mu_{\rm eff} > 1$ when $R < C$. We now show that for a lattice with $G(\Lambda)$ close to $\frac{1}{2\pi e}$ (i.e., with an approximately spherical shaping region in

a second moment sense) the error probability in ML decoding of an optimal code is indeed bounded by (55) with μ replaced by μ_{eff}.

A. Detailed Analysis

Denote by $E_\Lambda^r(R)$ and $E_\Lambda^x(R)$ the random coding and expurgated error exponents, respectively, of the Λ-MLAN channel $\bar{\boldsymbol{Y}}' = \boldsymbol{t} + \bar{\boldsymbol{N}}' \bmod \Lambda$, $\boldsymbol{t} \in \mathcal{V}$, as characterized in Lemma 2, relative to an input distribution uniform over the alphabet $\Omega = \mathcal{V}$. Note that for a modulo additive-noise channel, a uniform input indeed maximizes the random coding and expurgated bounds [22]. Let

$$E_\Lambda(R) = \max\left[E_\Lambda^r(R), E_\Lambda^x(R)\right] \tag{59}$$

so that $E_\Lambda(R)$ is a lower bound to the true error exponent of the Λ-MLAN channel.

As we have seen, choosing a shaping lattice Λ with normalized second moment $G(\Lambda)$ close enough to $\frac{1}{2\pi e}$ and $\alpha = \frac{P_X}{P_X+P_N}$, ensures a vanishing loss in capacity in the MLAN transformation. We now analyze the resulting error exponent.

The random coding error exponent of a modulo additive noise channel can be expressed conveniently in terms of Rényi entropies, see, e.g., [15]. For the Λ-MLAN channel, this gives

$$\begin{aligned} E_\Lambda^r(R) &= \max_{0\le\rho\le 1} \rho\left[\frac{1}{n}\log V - \frac{1}{n}h_{\bar\rho}(\boldsymbol{N}') - R\right] \\ &= \max_{0\le\rho\le 1} \rho\left[\frac{1}{n}\log\frac{V}{e^{nR}} - \frac{1}{n}h_{\bar\rho}(\boldsymbol{N}')\right] \\ &= \max_{0\le\rho\le 1} -\rho\left[\frac{1}{n}h_{\bar\rho}(\delta\bar{\boldsymbol{N}}')\right] \end{aligned} \tag{60}$$

where $\bar\rho = \frac{1}{1+\rho}$ and δ is the density of the codewords as defined in (57), and where the Rényi entropy of order β is defined by

$$h_\beta(\boldsymbol{N}') = \frac{\beta}{1-\beta}\log\left(\int_{\boldsymbol{x}} f_{\boldsymbol{N}'}(\boldsymbol{x})^\beta d\boldsymbol{x}\right)^{\frac{1}{\beta}} \tag{61}$$

where $f_{\boldsymbol{N}'}(\cdot)$ denotes the probability density of $\boldsymbol{N}'$. Taking into account that $P_X = G(\Lambda)\cdot V^{2/n}$, we have

$$E_\Lambda^r(R) = \max_{0\le\rho\le 1}\rho\left[\frac{1}{2}\log\frac{P_X}{G(\Lambda)} - \frac{1}{n}h_{\bar\rho}(\boldsymbol{N}') - R\right] \tag{62}$$

$$\ge \max_{0<\rho\le 1}\rho\left[\frac{1}{2}\log 2\pi e P_X - \frac{1}{n}h_{\bar\rho}(\boldsymbol{N}') - R\right] - \log 2\pi e G(\Lambda). \tag{63}$$

Let $Z\sim\mathcal{N}(0,\frac{P_XP_N}{P_X+P_N})$. In Lemma 5 in Appendix A, we show that for $C - \frac{\log 2}{2} < R < C$

$$\max_{0<\rho\le 1}\rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar\rho}(Z) - R\right] = E_P^r(\mu) \tag{64}$$

with $\mu = e^{2(C-R)}$ and $C = \frac{1}{2}\log(1+\frac{P_X}{P_N})$. Thus, Lemma 5 expresses Poltyrev's random coding exponent in terms of Gallager's random coding exponent of a mod-Λ channel. Recall from (27) that for $\alpha = \frac{P_X}{P_X+P_N}$, the variance per dimension of $\boldsymbol{N}'$ is $\frac{P_XP_N}{P_X+P_N}$. Thus, comparing (63) with (64) we see that if $\boldsymbol{N}'$ were white and Gaussian, then the random coding exponent of the MLAN channel would approach the Poltyrev random coding exponent as $G(\Lambda)\to\frac{1}{2\pi e}$. Theorem 4 specifies under what conditions this holds, and also extends this to the expurgated exponent.[11]

We first introduce the following definitions. Let R_u denote the covering radius of Λ, i.e., R_u is the radius of the smallest ball containing the Voronoi region $\mathcal{V}$. Also, let R_ℓ denote the effective radius of the Voronoi region, i.e., the radius of a sphere having the same volume as $\mathcal{V}$. Finally, substituting the effective VNR $\mu = e^{2(C-R)}$ from (58) in the Poltyrev exponent $E_P(\mu)$ given in (56), we get (65) at the bottom of the page. The following theorem is proved in Appendix A.

Theorem 4 (Random-Coding Error Exponent for a Fixed Shaping Lattice): For any n-dimensional lattice Λ, the error exponent of the Λ-MLAN channel satisfies

$$E_\Lambda(R) \ge E_P\left(e^{2(C-R-\epsilon_2(\Lambda))}\right) - \epsilon_1(\Lambda) \tag{66}$$

where $E_P(\cdot)$ is the Poltyrev exponent

$$\epsilon_1(\Lambda) \triangleq \log\left(\frac{R_u}{R_\ell}\right) + \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n} \tag{67}$$

and

$$\epsilon_2(\Lambda) \triangleq \log\left(\frac{R_u}{R_\ell}\right) + \frac{1}{2}\log 2\pi e G(\Lambda) \tag{68}$$

with R_u, R_ℓ, and $G(\Lambda)$ denoting the covering radius, effective radius, and normalized second moment of Λ, respectively, and G_n^* denoting the normalized second moment of an n-sphere.

To achieve the Poltyrev exponent $E_P(\mu)$ at $\mu = e^{2(C-R)}$, we would thus like $\epsilon_1(\Lambda)$ and $\epsilon_2(\Lambda)$ to be small. To that end, we confine ourselves to a more stringent class of shaping lattices Λ. Following a result of Rogers [29], [6], there exist lattices whose covering density, i.e., $\left(\frac{R_u}{R_\ell}\right)^n$, satisfies

$$1 \le \left(\frac{R_u}{R_\ell}\right)^n < c\cdot n\cdot(\log n)^a \triangleq \mathcal{R}(n) \tag{69}$$

for some positive constants c and a. We shall refer to such a sequence of lattices as "Rogers-good." By the proof of Lemma 1 of [32], this implies in particular that for such a sequence of

[11] Unfortunately, unlike for the case of capacity (corresponding to regular entropy, i.e., Rényi entropy of order $\bar\rho = 1$), there seems to be no direct way to bound the difference between $\frac{1}{n}h_{\bar\rho}(\boldsymbol{N}')$ and $h_{\bar\rho}(Z)$ in terms of $\log(2\pi e G(\Lambda))$.

$$E_P(e^{2(C-R)}) = \begin{cases} \frac{e^{2(C-R)}-1}{2} - (C-R), & \max(0, C-\frac{\log 2}{2}) \le R < C \\ C - R - \frac{1}{2}\log\frac{4}{e}, & \max(0, C-\log 2) \le R \le \max(0, C-\frac{\log 2}{2}) \\ \frac{1}{8}e^{2(C-R)}, & 0 < R \le \max(0, C-\log 2). \end{cases} \tag{65}$$

lattices $G(\Lambda^{(n)}) \to \frac{1}{2\pi e}$ as $n \to \infty$. Also, from (69) it follows that

$$\frac{1}{n}\log\left(\frac{R_u}{R_l}\right)^n \to 0$$

as $n \to \infty$.

Corollary 1: For a sequence of Rogers-good lattices $\Lambda^{(n)}$

$$\liminf_{n\to\infty} E_{\Lambda^{(n)}}(R) \geq E_P\left(e^{2(C-R)}\right) \tag{70}$$

where $E_P\left(e^{(2(C-R)}\right)$ is given in (65).

Remarks:

- Note that $E_P(\mu)$ vanishes at $\mu = 1$, i.e., at

  $$R = C = \frac{1}{2}\log(1 + P_X/P_N)$$

  which is the well-known capacity of the original AWGN channel.
- As noted, the exponent in (56) was derived by Poltyrev [28] in the context of coding for the unconstrained AWGN channel. In fact, the proof of Theorem 4 may be considered as an alternative simplified approach to proving Poltyrev's result. A similar simplification has been done previously in [20, Sec. VIII].
- In a preliminary version of this work, it was conjectured that (70) is in fact an equality, i.e., that the error exponent of the MLAN channel asymptotically equalts the Poltyrev exponent. A recent result [25], however, shows that a better error exponent can be achieved with an α which is different than α_{MMSE} and that, in fact, the random coding error exponent is (asymptotically) equal to that of the original power-constrained channel as given in (71). This surprising result implies that at least at high transmission rates, the MLAN transformation does not lose in error exponent.

B. Comparison With the Error Exponents of the Power-Constrained AWGN Channel

Denote the random coding error exponent of the original power-constrained channel (1), (3) by $E^r_{\mathrm{CON}}(\mu; \mathrm{SNR})$ (where R and μ are related via $\mu = e^{2(C-R)}$). This exponent is given by [22, p. 340] (71) and (72) (at the bottom of the page) for

$$\frac{1}{2}\log\left[\frac{1}{2}+\frac{\mathrm{SNR}}{4}+\frac{1}{2}\sqrt{1+\frac{\mathrm{SNR}^2}{4}}\right] \leq R \leq \frac{1}{2}\log(1+\mathrm{SNR}). \tag{73}$$

The expurgated exponent is given by [22, p. 342]

$$E^{\mathrm{ex}}_{\mathrm{CON}}(\mu, \mathrm{SNR}) = \frac{\mathrm{SNR}}{4}\left(1 - \sqrt{1 - \frac{\mu}{1+\mathrm{SNR}}}\right) \tag{74}$$

for

$$R \leq \frac{1}{2}\log\left[\frac{1}{2}+\frac{1}{2}\sqrt{1+\frac{\mathrm{SNR}^2}{4}}\right]. \tag{75}$$

Fig. 7 compares the exponents $E_{\mathrm{CON}}(R)$ and $E_P(R)$ for several SNR values. We note that at high SNR, the random coding and straight-line sections of $E_{\mathrm{CON}}(R)$ tend to the Poltyrev exponent. This can be seen more clearly in Fig. 8, where the random coding exponents are plotted as a function of the SNR and μ. Note that $E_P(\mu^{-1})$ does not depend on the SNR. We also note that at high SNR $E^{\mathrm{ex}}_{\mathrm{CON}}(R{=}0)$ is twice as large as $E^{\mathrm{ex}}_p(R{=}0)$.

VII. An Ensemble of Good Nested Lattice Codes

The scheme presented in Section IV-A assumes a nested pair of lattices such that the coarse lattice is good for quantization while the fine one is good for AWGN coding under ML decoding. In Section IV-C, we further assumed the existence of nested lattice pairs which allow Euclidean lattice decoding to be (asymptotically) optimal.

We now define, for any coding rate (35), a random ensemble of nested lattice pairs $\Lambda_1 \subset \Lambda$. We show that most members of the ensemble satisfy that the coarse lattice Λ is simultaneously Rogers-good (a good quantizer) and Poltyrev-good (a good channel code), while the fine lattice Λ_1 is Poltyrev-good. This will allow us to prove Theorem 5 which shows that the probability of error in the transmission scheme of Section IV-A satisfies $P_e = \Pr(\boldsymbol{N}' \notin \mathcal{V}_1) \leq e^{-n(E_\Lambda(R)-o_n(1))}$.

Clearly, by integer scaling a lattice we may obtain "self-similar" nested lattices for any *integer* nesting ratio. For example, Fig. 1 depicts a self-similar nested lattice pair of dimension two. The nine codewords are depicted as full circles. Note that the open circles are identical $\bmod \Lambda$ to full circles. Here, the nesting ratio is three. A lattice may also have a sublattice that is a scaled and rotated version of it; see [5]. In general, the pair of nested lattices discussed in this paper need not be similar and the nesting ratio does not have to be an integer. We note that in [27] a related construction of nested trellis codes is given that is better suited for applications.

We begin with a description of Loeliger's type A construction of a random $\bmod$-p lattice ensemble [26]. See [6] for a general definition of Construction A. The construction of a good n-dimensional lattice consists of the following steps [11]:

$$E^r_{\mathrm{CON}}(\mu; \mathrm{SNR}) \tag{71}$$

$$= \frac{1}{4}\left\{(\mathrm{SNR}+1+\mu) - (\mathrm{SNR}+1-\mu)\sqrt{1+\frac{\mathrm{SNR}+1}{\mathrm{SNR}}\frac{4}{\mathrm{SNR}+1+\mu}}\right\}$$

$$+ \frac{1}{2}\log\left\{\frac{\mathrm{SNR}+1}{\mu} - \frac{\mathrm{SNR}}{2\mu}(\mathrm{SNR}+1-\mu)\left[\sqrt{1+\frac{\mathrm{SNR}+1}{\mathrm{SNR}}\frac{4}{\mathrm{SNR}+1+\mu}}-1\right]\right\} \tag{72}$$

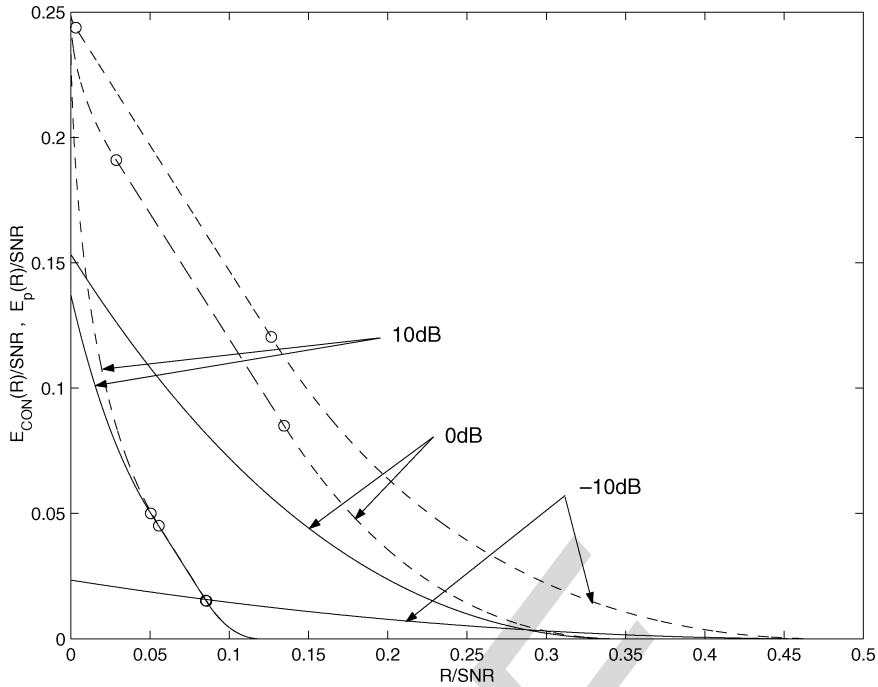

Fig. 7. Comparison of the random coding and expurgated error exponents of the power-constrained AWGN channel (dashed line) and the Poltyrev exponent (solid line). The circles in the figure separate the expurgated, straight-line, and random coding parts of the curves, respectively.

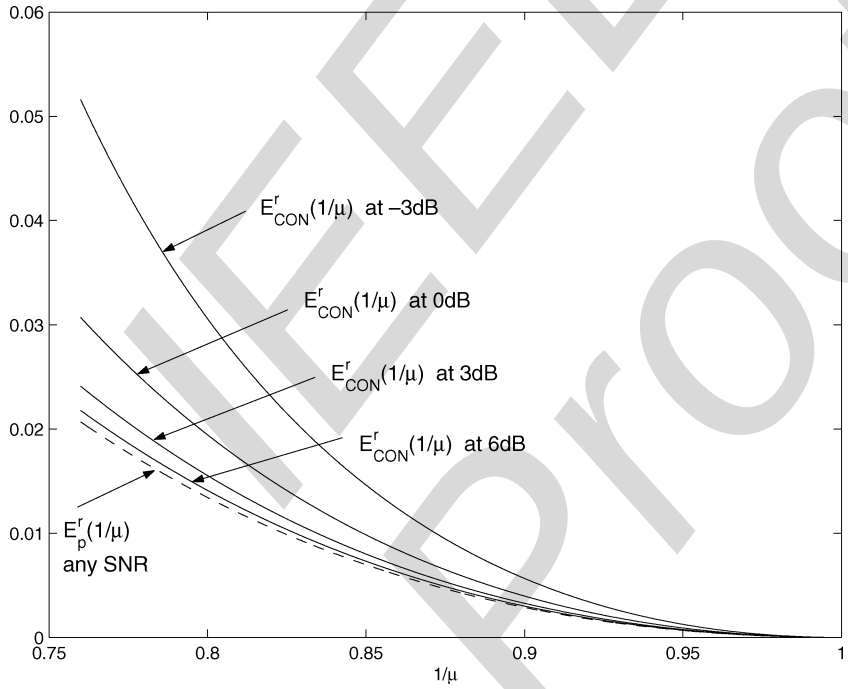

Fig. 8. Comparison of random coding exponent of the power-constrained AWGN channel (solid line) and the Poltyrev random coding exponent (dashed line).The curves depicted are all above the *critical* rates of the respective channels.

1) Draw a generating vector $\boldsymbol{g} = (g_1, \ldots, g_n)$ according to $g_i \sim \text{Unif}(0, \ldots, p-1)$ i.i.d., $i = 1, \ldots, n$.
2) Define the discrete codebook,

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{Z}_p^n : \boldsymbol{x} = (\boldsymbol{g} \cdot q) \bmod p \quad q = 0, \ldots, p-1.\}. \quad (76)$$

3) Apply Construction A to lift $\mathcal{C}$ to $\mathbb{R}^n$ and form the lattice:

$$\Lambda_1' = p^{-1}\mathcal{C} + \mathbb{Z}^n. \quad (77)$$

The goodness of this lattice ensemble for AWGN channel coding and for quantization is shown in [11]. We extend the discussion to the generation of a pair of *nested* lattices which is good for the MLAN channel. We use a transformed version of Λ_1' above as the fine lattice. As for a coarse (shaping) lattice, we use a lattice Λ that is simultaneously Rogers-good and Poltyrev-good. This is necessary for Euclidean decoding to be adequate, since a Euclidean decoder "ignores" the folding of the noise and hence we would like the probability of folding

to be (exponentially) small. In [11], it is shown that a lattice may indeed simultaneously be Rogers-good for covering and Poltyrev-good for channel coding. That is, we may take Λ (more precisely, the sequence of such lattices) such that the following two properties hold:

1) $\left(\frac{R_u}{R_\ell}\right)^n < c\cdot n\cdot(\log n)^a$ for some constants c and a, where R_u and R_ℓ are the covering radius and the effective radius, respectively, associated with Λ.
2) For any $\sigma^2 < \sigma^2(\mathcal{V})$

$$\Pr\{\mathcal{N}(0,\sigma^2) \notin \mathcal{V}\} < e^{-n[E_P(\mu)-o_n(1)]} \tag{78}$$

where $\mu = \frac{\sigma^2(\mathcal{V})}{\sigma^2}$ is the VNR of the coarse lattice (viewed as a channel code) relative to a noise $\mathcal{N}(0,\sigma^2)$.

Note that the first property implies that $\log(\frac{R_u}{R_\ell}) \to 0$ as $n \to \infty$. Let G denote the generating matrix of this lattice.

From the construction of the fine lattice Λ_1' above, we have that the n-dimensional cubic lattice $\mathbb{Z}^n$ may be viewed as nested in the resulting lattice, i.e., $\mathbb{Z}^n \subset \Lambda_1'$. The nesting ratio is given by

$$\gamma = \sqrt[n]{1/V_1'} = \sqrt[n]{p} \tag{79}$$

so the coding rate (35) is $R = \log(\gamma) = \frac{1}{n}\log(p)$. We now apply the linear transformation G to Λ_1' to obtain the modified lattice Λ_1 such that (Λ,Λ_1) is the desired nested lattice pair. Note that the transformation does not affect the nesting ratio. Since the unit cubic lattice $\mathbb{Z}^n$ is a sublattice of Λ_1', it follows that $\Lambda = G\mathbb{Z}^n$ is a sublattice of $\Lambda_1 = G\Lambda_1'$.

We may view the construction as starting with a self-similar pair of nested lattices $\Lambda \in p^{-1}\Lambda$ as depicted in Fig. 1. The nesting ratio at this point is $\sqrt[n]{p^n} = p$. We then dilute the lattice $p^{-1}\Lambda$ by picking one of its points, along with all its multiples modulo-Λ, and throwing away all the remaining points. This results in a new lattice Λ_1 and a nesting ratio of $\sqrt[n]{p}$. Since the total number of codewords is p, for a given rate R we must choose $p = [e^{nR}]$, where $[\cdot]$ denotes rounding to the nearest prime, and apply the above construction.

For large p, the resulting ensemble is "matched" to the Λ-MLAN channel, in the sense that the codewords of the fine lattice Λ_1 become uniform over the Voronoi region of Λ. Hence, a typical member of the ensemble approaches the optimum random-coding error exponent of this channel. These facts are proved in the next section.

VIII. Error Analysis in Euclidean Lattice Decoding

In this section, we prove that capacity as well as the Poltyrev exponent may be approached arbitrarily closely using nested lattices from the ensemble described in Section VII and a Euclidean decoding metric as defined in Section IV-C. Specifically, we prove the following theorem.

Theorem 5 (Error Exponent in Euclidean Lattice Decoding): For any rate $R < C = \frac{1}{2}\log(1+\text{SNR})$, there exists a sequence of n-dimensional nested lattice pairs $(\Lambda_1^{(n)},\Lambda^{(n)})$ whose coding rate as defined in (35) approaches R, and whose decoding error probability (48) under Euclidean lattice decoding satisfies

$$P_e = \Pr\{\boldsymbol{N}' \notin \mathcal{V}_1^{(n)}\} \le e^{-n\left(E_P\left(e^{2(C-R)}\right)-o_n(1)\right)} \tag{80}$$

where $o_n(1) \to 0$ as $n \to \infty$, and $E_p(\mu)$ is the Poltyrev exponent given in (56).

Since $E_P(\mu) > 0$ for all $\mu > 1$, Theorem 5 implies that for every rate R smaller than capacity, P_e goes to zero as $n \to \infty$, which is Theorem 3. Furthermore, since Euclidean decoding is suboptimal relative to ML decoding, Theorem 2 follows as well.

Proof: Assume the ensemble of nested lattices defined in the previous section with $\sigma^2(\mathcal{V}) = P_X$, covering radius R_u, effective radius R_ℓ, and coding rate R. Note that the coarse lattice Λ is fixed and not drawn at random. We wish to evaluate the error probability P_e in lattice decoding of a random member of this ensemble using the standard random-coding error exponent method [22] (as done, e.g., by Poltyrev [28] and Loeliger [26]). But this method assumes ML decoding, while as explained in Section IV-C, Euclidean decoding may not be ML. To overcome this difficulty, we first bound P_e by the probability of error in the presence of "truncated Gaussian" noise, $\boldsymbol{Z}_{\mathcal{V}}$, for which Euclidean decoding *is* optimal. To establish this bound, we need to define a few auxiliary random vectors.

Recall that $\boldsymbol{N}'' = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N}$ so that $\boldsymbol{N}' = \boldsymbol{N}'' \bmod \Lambda$. That is, $\boldsymbol{N}''$ is the effective noise prior to the modulo operation. In Lemmas 6 and 11 in Appendix A we show that there exists a Gaussian vector $\boldsymbol{Z}^* \sim \mathcal{N}(0, P_{Z^*}\cdot\boldsymbol{I})$ with

$$\begin{aligned}\frac{n}{n+2}\cdot\frac{P_XP_N}{P_X+P_N} \le P_{Z^*} &< \left(\frac{R_u}{R_\ell}\right)^2\frac{P_XP_N}{P_X+P_N}\\ &\le \sqrt[n/2]{\mathcal{R}(n)}\cdot\frac{P_XP_N}{P_X+P_N}\end{aligned} \tag{81}$$

such that

$$f_{\boldsymbol{N}''}(\boldsymbol{x}) < e^{\epsilon_1(\Lambda)\cdot n} f_{\boldsymbol{Z}^*}(\boldsymbol{x}) \tag{82}$$

where $\epsilon_1(\Lambda)$ is defined in (67), and $\mathcal{R}(n)$ is defined in (69). That is, the density of $\boldsymbol{N}''$ is not "much" greater than that of the density of a Gaussian distribution with a "slightly" greater variance. Note that the bound is uniform in $\boldsymbol{x}$, i.e., $\epsilon_1(\Lambda)$ does not depend on $\boldsymbol{x}$. Thus, we may bound the probability of error by

$$P_e = \Pr\{\boldsymbol{N}' \notin \mathcal{V}_1\} \le \Pr\{\boldsymbol{N}'' \notin \mathcal{V}_1\} \le e^{\epsilon_1(\Lambda)\cdot n}\Pr\{\boldsymbol{Z}^* \notin \mathcal{V}_1\}. \tag{83}$$

Unfortunately, we cannot apply the random-coding error exponent of Theorem 4 to bound $\Pr\{\boldsymbol{Z}^* \notin \mathcal{V}_1\}$, since $\boldsymbol{Z}^*$ is not a modulo-Λ noise. Also, we cannot apply it to bound $\Pr\{\boldsymbol{Z}^* \bmod \Lambda \notin \mathcal{V}_1\}$, because for this noise $\mathcal{V}_1$ is not an ML region. Instead, we shall bound this probability in terms of $\Pr\{\boldsymbol{Z}_{\mathcal{V}} \notin \mathcal{V}_1\}$, where $\boldsymbol{Z}_{\mathcal{V}}$ is a truncated version of $\boldsymbol{Z}^*$ limited to the Voronoi region of Λ. That is, $\boldsymbol{Z}_{\mathcal{V}}$ has the following distribution:

$$f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x}) = \begin{cases}\frac{1}{1-\epsilon_t(n)} f_{\boldsymbol{Z}^*}(\boldsymbol{x}), & \boldsymbol{x}\in\mathcal{V}\\ 0, & \text{otherwise}\end{cases} \tag{84}$$

where

$$\epsilon_t(n) \triangleq \Pr\{\boldsymbol{Z}^* \notin \mathcal{V}\} \tag{85}$$

is the probability of truncation. Since $\mathcal{V}_1 \subset \mathcal{V}$, we have

$$\Pr(\boldsymbol{Z}_{\mathcal{V}} \in \mathcal{V}_1) = \Pr(\boldsymbol{Z}^* \in \mathcal{V}_1)/\Pr(\boldsymbol{Z}^* \in \mathcal{V}).$$

Thus,

$$\begin{aligned}\Pr(\boldsymbol{Z}^* \notin \mathcal{V}_1) &= 1 - \Pr(\boldsymbol{Z}^* \in \mathcal{V}_1) && (86)\\ &= 1 - \Pr(\boldsymbol{Z}_{\mathcal{V}} \in \mathcal{V}_1)\Pr(\boldsymbol{Z}^* \in \mathcal{V}) && (87)\\ &\le \Pr(\boldsymbol{Z}_{\mathcal{V}} \notin \mathcal{V}_1) + \Pr(\boldsymbol{Z}^* \notin \mathcal{V}). && (88)\end{aligned}$$

Note that from (81), the equivalent VNR of the coarse lattice (viewed as a channel code) relative to Z^* is

$$\mu = \frac{P_X}{P_{Z^*}} \ge 1 + \frac{P_X}{P_N} - o_n(1) = e^{2C} - o_n(1)$$

so from (78) the second term in (88) is upper-bounded by

$$\Pr(\boldsymbol{Z}^* \notin \mathcal{V}) \le e^{-n\left(E_P\left(e^{2C}\right)-o_n(1)\right)}. \tag{89}$$

We now turn to evaluate the first term in (88), $\Pr(\boldsymbol{Z}_{\mathcal{V}} \notin \mathcal{V}_1)$. Consider a $(\Lambda, \boldsymbol{Z}_{\mathcal{V}})$-MLAN channel

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}_{\mathcal{V}} \bmod \Lambda. \tag{90}$$

The next lemma shows that when Λ is simultaneously good in the above meaning, the exponent of this channel is arbitrarily close to the Poltyrev exponent for large enough dimension n.

Lemma 3: If Λ is Rogers-good and Poltyrev-good, then the random coding exponents of the $(\Lambda, \boldsymbol{Z}_{\mathcal{V}})$-MLAN channel satisfy

$$E_\Lambda(R; \boldsymbol{Z}_{\mathcal{V}}) \ge E_P(e^{2(C-R)} - o_n(1)) - o_n(1). \tag{91}$$

The lemma is proved in Appendix B.

We next show that we may replace the random code with a lattice from the ensemble defined in the previous section without affecting the error exponent. Consider first the random code (nonlattice) ensemble obtained by applying a uniform distribution over the fine grid $\left(p^{-1}\cdot\Lambda\right)\cap\mathcal{V}$. Denote the union of the random coding and expurgated error exponent corresponding to this ensemble by

$$E_\Lambda(R; \boldsymbol{Z}_{\mathcal{V}}, p) = \max\left(E_\Lambda^r(R; \boldsymbol{Z}_{\mathcal{V}}, p), E_\Lambda^{\mathrm{ex}}(R; \boldsymbol{Z}_{\mathcal{V}}, p)\right) \tag{92}$$

where, as above, $p = [e^{nR}]$. The next lemma is proved in Appendix C.

Lemma 4: If Λ is Rogers-good, then

$$E_\Lambda^r(R; \boldsymbol{Z}_{\mathcal{V}}, p) > E_\Lambda^r(R; \boldsymbol{Z}_{\mathcal{V}}) - o_n(1). \tag{93}$$

The claim for the expurgated exponent may be proved similarly.

Consider now the ensemble of nested codes defined in Section VII. It can be seen that each codeword in the ensemble is uniformly distributed over the basic grid $\left(p^{-1}\cdot\Lambda\right)\cap\mathcal{V}$. Furthermore, the distribution of the difference between any two codewords is also uniform. The pairwise distribution is thus identical to that obtained by drawing each codeword independently and uniformly over the basic grid $\left(p^{-1}\cdot\Lambda\right)\cap\mathcal{V}$ as done in the random code ensemble. Therefore (see [22]), these two ensembles have the same random coding error exponent. It may also be shown that with probability going to one (as $n\to\infty$) a lattice drawn from the proposed nested lattice ensemble will satisfy the expurgated error exponent bound in (92). Thus, the probability of error for this ensemble of nested lattices when used over a $(\Lambda, \boldsymbol{Z}_{\mathcal{V}})$-MLAN channel with ML decoding is governed by the error exponent $E_\Lambda(R; \boldsymbol{Z}_{\mathcal{V}}, p)$. Furthermore, since $\mathcal{V}_1 \subset \mathcal{V}$ and the density of $\boldsymbol{Z}_{\mathcal{V}}$ is proportional to $e^{-\|\boldsymbol{x}\|^2/2P_{Z^*}}$ inside $\mathcal{V}$ and zero elsewhere, it follows that Euclidean decoding is ML for this channel. Thus,

$$\Pr(\boldsymbol{Z}_{\mathcal{V}} \notin \mathcal{V}_1) \le e^{-nE_\Lambda^r(R; \boldsymbol{Z}_{\mathcal{V}}, p)}. \tag{94}$$

We can now combine (83), (88), (89), Lemma 3, Lemma 4, and (94), to obtain

$$\begin{aligned}&\Pr(\boldsymbol{N}' \notin \mathcal{V}_1)\\ &\le e^{\epsilon_1(\Lambda)\cdot}\cdot\left[e^{-n\left(E_P\left(e^{2(C-R)}\right)-o_n(1)\right)} + e^{-n\left(E_P\left(e^{2C}\right)-o_n(1)\right)}\right] && (95)\\ &\le e^{-n\left(E_P\left(e^{2(C-R)}\right)-o_n(1)\right)} && (96)\end{aligned}$$

where the second inequality follows because the first exponent dominates. This establishes Theorem 5. □

IX. Conclusion

We have demonstrated that using nested lattice codes in conjunction with an MMSE-scaled transformation of the AWGN channel into a modulo additive noise channel, lattice codes can achieve capacity using *lattice decoding*. It should be noted, however, that the precise definition of lattice encoding and decoding used throughout this work differs somewhat from that in previous works.

This transformation of the original power-constrained channel into a modulo additive-noise channel, though sufficient for achieving capacity, is not strictly information lossless. The error exponent is lower-bounded by the Poltyrev exponent, which was derived in [28] in the context of coding for the *unconstrained* AWGN channel.

As illuminated by Forney [18], the combination of MMSE estimation with a dithered lattice code presented here offers a useful connection between Wiener and Shannon theories. Recent work indeed indicates that the underlying principle may find application in diverse areas of digital communications, see, e.g., [23]. The random dither can be replaced in practice by a suitable deterministic translation of the fine lattice [18]. As discussed in Section V, for finite-dimensional (e.g., uncoded) modulation, the best linear estimator slightly deviates from the MMSE solution. Section V also presents equivalent forms of the estimation–lattice–decoding scheme.

Similar observations were made in the source coding context, by incorporating filters with dithered lattice quantizers [34].

Here the role of shaping is accomplished by entropy coding. As in the scaled MLAN transformation, with "good" lattices and optimum filters, entropy coded dithered lattice quantization achieves Shannon's rate-distortion function for Gaussian sources.

The proposed encoding scheme may easily be generalized to nonwhite Gaussian noise/linear Gaussian intersymbol interference (ISI) channels. The scheme also is related to "dirty paper" coding techniques. In particular, the coarse lattice component of the nested code plays a role similar to that of the "lattice strategy" for canceling interference known to the transmitter [12], stemming from the work of Costa on the "dirty paper" channel [8]. Indeed, the present work was directly motivated by [12]. In this respect, it confirms that various lattice-theoretic schemes such as trellis shaping and precoding for ISI channels may be extended so as to achieve capacity at *any* SNR and there is no "inherent" precoding loss. See [35], [27], [13] for a detailed account.

The notion of "good" nested lattices is central to our approach. Such codes are useful for "structured binning" [3], [35]. As mentioned in Section VII, one approach to the construction of such codes is by using self-similar lattices [5]. However, this approach is limited and it is not clear that any nesting ratio may be approached with self-similar lattices. The construction given in Section VII is more general and allows for any nesting ratio. Furthermore, it may readily be interpreted in terms of conventional coding techniques in the spirit of trellis shaping [17].

Appendix A
Random Coding Error Exponents of MLAN Channel

In this appendix, we prove the following two propositions which together constitute Theorem 4.

Proposition 1:

$$E_\Lambda^r(R) \geq E_P^r\left(e^{2(C-R-\epsilon_2(\Lambda))}\right) - \epsilon_1(\Lambda) \tag{97}$$

for $\max(0, C - \frac{\log 2}{2}) \leq R < C$ where $\epsilon_1(\Lambda)$ and $\epsilon_2(\Lambda)$ are defined in (67) and (68), respectively.

Proposition 2:

$$E_\Lambda^x(R) \geq E_P^x\left(e^{2(C-R-\epsilon_2(\Lambda))}\right) - \epsilon_1(\Lambda) \tag{98}$$

for $0 < R \leq \max(0, C - \log 2)$.

Before proving the first proposition, we introduce two lemmas. We use the following identity that relates the Poltyrev random coding exponent to that of an "infinite-dimensional" MLAN channel.

Lemma 5 (Poltyrev Exponent as a "Spherical" MLAN Exponent I):

$$\max_{0<\rho\leq 1} \rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z) - R\right] = E_P^r(\mu) \tag{99}$$

with $\mu = e^{2(C-R)}$ and $Z \sim \mathcal{N}(0, \frac{P_X P_N}{P_X + P_N})$.

Proof: Let $\delta^* = \frac{e^R}{\sqrt{2\pi e P_X}}$. Using the the fact that for Rényi entropy (as for Shannon entropy), $h_{\bar{\rho}}(aX) = h_{\bar{\rho}}(X) + \log a$, we can rewrite (99) as

$$\max_{0<\rho\leq 1} -\rho h_{\bar{\rho}}(\delta^* Z) = E_P^r(\mu). \tag{100}$$

Define $Z_2 = \delta^* Z$. The Rényi entropy of order $\bar{\rho}$ of a (generic) Gaussian random variable X with variance P_X is

$$h_{\bar{\rho}}(X) = h(X) - \frac{1}{2}\log e\bar{\rho}^{\frac{1}{1-\bar{\rho}}} \tag{101}$$

$$= \frac{1}{2}\left[\log 2\pi P_X - \frac{1}{1-\bar{\rho}}\log\bar{\rho}\right] \tag{102}$$

$$= \frac{1}{2}\left[\log 2\pi P_X + \frac{1+\rho}{\rho}\log(1+\rho)\right] \tag{103}$$

substituting $\bar{\rho} = \frac{\rho}{1+\rho}$. We therefore have

$$\rho h_{\bar{\rho}}(Z_2) = \frac{1}{2}\left[\rho\log 2\pi P_{Z_2} + (1+\rho)\log(1+\rho)\right]. \tag{104}$$

Taking the derivative of $-\rho h_{\bar{\rho}}(Z_2)$ with respect to ρ, we get

$$\frac{d}{d\rho}\left[-\rho h_{\bar{\rho}}(Z_2)\right] = -\frac{1}{2}\left[\log 2\pi P_{Z_2} + \log(1+\rho) + 1\right]. \tag{105}$$

Thus, an extremum occurs when

$$\log(1+\rho) = -\log 2\pi e P_{Z_2} \tag{106}$$

or, equivalently, when

$$\rho = \frac{1}{2\pi e P_{Z_2}} - 1. \tag{107}$$

It is easy to verify that this extremum is indeed a maximum. Substituting (106) and (107) in (104), we get

$$\max_\rho -\rho h_{\bar{\rho}}(Z_2)$$

$$= -\frac{1}{2}\left[\frac{\log 2\pi P_{Z_2}}{2\pi e P_{Z_2}} - \log 2\pi P_{Z_2} - \frac{1}{2\pi e P_{Z_2}}\cdot\log 2\pi e P_{Z_2}\right] \tag{108}$$

$$= \frac{1}{2}\left[-\frac{1}{2\pi e P_{Z_2}} - \log 2\pi e P_{Z_2} + 1\right]. \tag{109}$$

From the definition of Z_2 (57), we get

$$P_{Z_2} = \delta^{*2} P_Z = \frac{e^{2R}}{2\pi e P_X}\cdot\frac{P_X P_N}{P_X + P_N}$$

$$= \frac{e^{2R}}{2\pi e(1+\text{SNR})} = \frac{1}{2\pi e\mu}. \tag{110}$$

Substituting (110) for P_{Z_2} in (109) we get

$$\max_{0<\rho<1} -\rho h_{\bar{\rho}}(Z_2) = \frac{1}{2}\left[(\mu - 1) - \log\mu\right]. \tag{111}$$

Finally, from (107) and (110) we note that $\rho = 1$ corresponds to a rate R satisfying

$$1 = \frac{1}{2\pi e/(2\pi e\frac{1+\text{SNR}}{e^{2R}})} - 1 \tag{112}$$

from which we obtain the critical rate

$$R = C - \frac{\log 2}{2}. \tag{113}$$

□

We next define a number of auxiliary random variables. Let R_u be the covering radius of Λ. Denote by $\mathcal{B}(R_u)$ a ball of radius

R_u and let σ^2 be the second moment per dimension of $\mathcal{B}(R_u)$. We have (see, e.g., [32], for details)

$$\sigma^2 = \frac{1}{n}\frac{1}{|\mathcal{B}(R_u)|}\int_{\mathcal{B}(R_u)} \|\boldsymbol{x}\|^2 d\boldsymbol{x} = G_n^* \cdot |\mathcal{B}(R_u)|^{2/n} = \frac{R_u^2}{n+2} \tag{114}$$

where G_n^* denotes the normalized second moment of an n-sphere. Note that σ^2 is the second moment of a ball containing $\mathcal{V}$, which has second moment P_X. Thus, $P_X < \sigma^2$.

Define the following:

- $\boldsymbol{Z}_1 \sim \mathcal{N}(0, \sigma^2 \cdot \boldsymbol{I}^n)$ where $\boldsymbol{I}^n$ is the identity matrix of dimension n;
- $\boldsymbol{Z}^* = (1-\alpha)\boldsymbol{Z}_1 + \alpha\boldsymbol{N}$;
- $\boldsymbol{Z} = \mathcal{N}\left(0, \frac{P_X P_N}{P_X+P_N}\right)$.

The variance of $\boldsymbol{Z}^*$ is related to that of $\boldsymbol{Z}$ by the following lemma.

Lemma 6:

$$\frac{n}{n+2}\cdot\frac{P_X P_N}{P_X+P_N} \le \mathrm{Var}(Z^*) = (1-\alpha)^2\sigma^2 + \alpha^2 P_N < \left(\frac{R_u}{R_\ell}\right)^2 \frac{P_X P_N}{P_X+P_N}. \tag{115}$$

Proof: Recall that $\boldsymbol{B} \sim \mathrm{Unif}(\mathcal{B}(R_u))$ and that R_ℓ satisfies $\mathrm{Vol}(\mathcal{B}(R_\ell)) = \mathrm{Vol}(\mathcal{V})$. Since a ball has the smallest normalized second moment, it follows that

$$\frac{1}{n}E\|\boldsymbol{U}\|^2 \ge \frac{1}{n}E\left\|\boldsymbol{B}\cdot\frac{R_\ell}{R_u}\right\|^2 \tag{116}$$

$$= \left(\frac{R_\ell}{R_u}\right)^2 \sigma^2 \tag{117}$$

$$= \left(\frac{R_\ell}{R_u}\right)^2 \mathrm{Var}(Z_1). \tag{118}$$

Now from (118) and (23), we see that

$$\mathrm{Var}(Z^*)$$

$$= (1-\alpha)^2\frac{1}{n}\mathrm{Var}(\boldsymbol{Z}_1) + \alpha^2\frac{1}{n}\mathrm{Var}(\boldsymbol{N}) \tag{119}$$

$$\le (1-\alpha)^2\frac{1}{n}\left[\left(\frac{R_u}{R_\ell}\right)^2\frac{1}{n}E\|\boldsymbol{U}\|^2\right] + \alpha^2\frac{1}{n}\mathrm{Var}(\boldsymbol{N}) \tag{120}$$

$$\le \left(\frac{R_u}{R_\ell}\right)^2\left[(1-\alpha)^2\frac{1}{n}E\|\boldsymbol{U}\|^2 + \alpha^2\frac{1}{n}\mathrm{Var}(\boldsymbol{N})\right] \tag{121}$$

$$= \left(\frac{R_u}{R_\ell}\right)^2 \cdot \frac{P_X P_N}{P_X+P_N}. \tag{122}$$

On the other hand, we have

$$\frac{1}{n}E\|\boldsymbol{U}\|^2 \le \frac{1}{n}R_u^2 = \frac{n+2}{n}\cdot\sigma^2 \tag{123}$$

from which follows the left inequality in (115). □

We are now ready to prove Proposition 1.

Proof of Proposition 1: We may bound the random coding error exponent of the Λ-MLAN channel as follows:

$$E_\Lambda^r(R)$$

$$= \max_{0<\rho\le 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}h_{\bar{\rho}}(\boldsymbol{N}') - R\right] \tag{124}$$

$$\ge \max_{0<\rho\le 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}h_{\bar{\rho}}(\boldsymbol{N}'') - R\right] \tag{125}$$

$$\ge \max_{0<\rho\le 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}h_{\bar{\rho}}(\boldsymbol{Z}^*) - R\right] - \epsilon_1(\Lambda) \tag{126}$$

$$= \max_{0<\rho\le 1}\rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z^*) - R - \frac{1}{2}\log 2\pi e G(\Lambda)\right] - \epsilon_1(\Lambda) \tag{127}$$

$$= \max_{0<\rho\le 1}\rho\Big[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z) - [h_{\bar{\rho}}(Z^*) - h_{\bar{\rho}}(Z)] - R - \frac{1}{2}\log 2\pi e G(\Lambda)\Big] - \epsilon_1(\Lambda) \tag{128}$$

$$\ge \max_{0<\rho\le 1}\rho\Big[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z) - R - \log\left(\frac{R_u}{R_\ell}\right) - \frac{1}{2}\log 2\pi e G(\Lambda)\Big] - \epsilon_1(\Lambda) \tag{129}$$

$$= \max_{0<\rho\le 1}\rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z) - (R+\epsilon_2(\Lambda))\right] - \epsilon_1(\Lambda) \tag{130}$$

$$= E_P^r\left(e^{2(C-R-\epsilon_2(\Lambda))}\right) - \epsilon_1(\Lambda) \tag{131}$$

where (125) follows since the function $p(x)^\alpha$ is convex-$\cap$ for $0<\alpha<1$; (126) follows by Lemma 9 proved below; (129) follows by Lemma 6 and since $h_\beta(aX) = h_\beta(X) + \log a$; and (131) follows by Lemma 5. □

Expurgated Exponent: We next bound the expurgated error exponent [22] of the Λ-MLAN channel. Since for a modulo additive noise channel the expurgated exponent is achieved by a uniform input, we have (132)–(134) at the bottom of the following page. The last expression my be rewritten as follows. For the Λ-MLAN channel, define the generalized Bhattacharyya distance of order $\rho \ge 0$ by

$$D_\rho^{\mathrm{Bhatt}}(\Lambda; \boldsymbol{N}') = \log\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f_{\boldsymbol{N}'}(\boldsymbol{y}) f_{\boldsymbol{N}'}([\boldsymbol{y}+\boldsymbol{x}] \bmod \Lambda)}\,d\boldsymbol{y}\right)^{\frac{1}{\rho}}. \tag{135}$$

Recall the definition of $\boldsymbol{N}''$, the effective noise prior to folding, i.e., $\boldsymbol{N}'' = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N}$. For the noise $\boldsymbol{N}''$, define

$$D_\rho^{\mathrm{Bhatt}}(\boldsymbol{N}'') = \log\int_{\boldsymbol{x}\in\mathbb{R}^n}\left(\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{f_{\boldsymbol{N}''}(\boldsymbol{y}) f_{\boldsymbol{N}''}(\boldsymbol{y}+\boldsymbol{x})}\,d\boldsymbol{y}\right)^{\frac{1}{\rho}}. \tag{136}$$

We similarly define $D_\rho^{\mathrm{Bhatt}}(\boldsymbol{Z}^*)$. Thus, we may write

$$E_\Lambda^{\mathrm{ex}}(R; \boldsymbol{N}') = \sup_{\rho\ge 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}D_\rho^{\mathrm{Bhatt}}(\Lambda; \boldsymbol{N}') - R\right]. \tag{137}$$

We have the following lemma.

Lemma 7:

$$D_\rho^{\mathrm{Bhatt}}(\Lambda;\boldsymbol{N}') \le D_\rho^{\mathrm{Bhatt}}(\boldsymbol{N}''). \tag{138}$$

Proof: See (139)–(145) at the bottom of the page, where (141) and (143) follow since $x^\alpha + y^\alpha \ge (x+y)^\alpha$ for positive x and y and $0 \le \alpha \le 1$. □

We also have the following identity (analogous to Lemma 5) that relates the Poltyrev expurgated exponent to that of an "infinite-dimensional" spherical MLAN channel.

Lemma 8 (Poltyrev Exponent as a "Spherical" MLAN Exponent II):

$$\sup_{\rho\ge 1}\rho\left[\frac{1}{2}\log 2\pi e P_X - D_\rho^{\mathrm{Bhatt}}(Z) - R\right] = E_P^r(\mu) \tag{146}$$

for $0 < R \le \max(0, C - \log 2)$ with $\mu = e^{2(C-R)}$.

Proof: Let us first compute $D_\rho^{\mathrm{Bhatt}}(Z)$. We use the following property of Gaussian distributions:

$$\begin{aligned} f_Z(y+x)f_Z(x) &= f_{Z/\sqrt{2}}(y+x/2)f_{\sqrt{2}Z}(x) && (147)\\ &= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})}e^{-\frac{(y+x/2)^2}{2(\sigma/\sqrt{2})^2}}\cdot\frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)}e^{-\frac{x^2}{2(\sqrt{2}\sigma)^2}} && (148)\end{aligned}$$

where $\sigma^2 = \mathrm{Var}(Z) = \frac{P_X P_N}{P_X+P_N}$. This follows since

$$\begin{aligned} f_Z(y)f_Z(y+x) &= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y+x)^2}{2\sigma^2}} && (149)\\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2 e^{-\frac{2y^2+2yx+x^2}{2\sigma^2}} && (150)\\ &= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})}\frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)}e^{-\frac{(\sqrt{2}y+x/\sqrt{2})^2}{2\sigma^2}}\cdot e^{-\frac{(x/\sqrt{2})^2}{2\sigma^2}} && (151)\\ &= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})}e^{-\frac{(y+x/2)^2}{2(\sigma/\sqrt{2})^2}}\cdot\frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)}e^{-\frac{x^2}{2(\sqrt{2}\sigma)^2}}. && (152)\end{aligned}$$

$$\begin{aligned} E_\Lambda^{\mathrm{ex}}(R) &= \sup_{\rho\ge1}\rho\left[-\frac{1}{n}\log\int_{\boldsymbol{t}_1\in\mathcal{V}}\int_{\boldsymbol{t}_2\in\mathcal{V}}\frac{1}{V^2}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f(\boldsymbol{y}|\boldsymbol{t}_1)f(\boldsymbol{y}|\boldsymbol{t}_2)}d\boldsymbol{y}\right)^{\frac{1}{\rho}}d\boldsymbol{t}_1 d\boldsymbol{t}_2 - R\right]\\ &= \sup_{\rho\ge1}\rho\Big[\frac{1}{n}\log V^2 - R\\ &\quad - \frac{1}{n}\log\int_{\boldsymbol{t}_1\in\mathcal{V}}\int_{\boldsymbol{t}_2\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f_{\boldsymbol{N}'}([\boldsymbol{y}-\boldsymbol{t}_1]\bmod\Lambda)f_{\boldsymbol{N}'}([\boldsymbol{y}-\boldsymbol{t}_2]\bmod\Lambda)}d\boldsymbol{y}\right)^{\frac{1}{\rho}}d\boldsymbol{t}_1 d\boldsymbol{t}_2\Big] && (132)\\ &= \sup_{\rho\ge1}\rho\Big[\frac{1}{n}\log V^2 - R\\ &\quad - \frac{1}{n}\log\int_{\boldsymbol{t}_1\in\mathcal{V}}\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f_{\boldsymbol{N}'}([\boldsymbol{y}-\boldsymbol{t}_1]\bmod\Lambda)f_{\boldsymbol{N}'}([\boldsymbol{y}-\boldsymbol{t}_1-\boldsymbol{x}]\bmod\Lambda)}d\boldsymbol{y}\right)^{\frac{1}{\rho}}d\boldsymbol{x}d\boldsymbol{t}_1\Big] && (133)\\ &= \sup_{\rho\ge1}\rho\left[\frac{1}{n}\log V^2 - \frac{1}{n}\log V\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f_{\boldsymbol{N}'}(\boldsymbol{y})f_{\boldsymbol{N}'}([\boldsymbol{y}+\boldsymbol{x}]\bmod\Lambda)}d\boldsymbol{y}\right)^{\frac{1}{\rho}}d\boldsymbol{x} - R\right]. && (134)\end{aligned}$$

$$\begin{aligned} D_\rho^{\mathrm{Bhatt}}(\Lambda;\boldsymbol{N}') &= \log\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{f_{\boldsymbol{N}'}(\boldsymbol{y})f_{\boldsymbol{N}'}([\boldsymbol{y}+\boldsymbol{x}]\bmod\Lambda)}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (139)\\ &= \log\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sqrt{\sum_{\lambda\in\Lambda}f_{\boldsymbol{N}''}(\boldsymbol{y}+\lambda)\sum_{\lambda'\in\Lambda}f_{\boldsymbol{N}'}((\boldsymbol{y}+\boldsymbol{x})\bmod\Lambda+\lambda')}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (140)\\ &\le \log\int_{\boldsymbol{x}\in\mathcal{V}}\left(\int_{\boldsymbol{y}\in\mathcal{V}}\sum_{\lambda\in\Lambda}\sum_{\lambda'\in\Lambda}\sqrt{f_{\boldsymbol{N}''}(\boldsymbol{y}+\lambda)f_{\boldsymbol{N}'}((\boldsymbol{y}+\boldsymbol{x})\bmod\Lambda+\lambda')}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (141)\\ &= \log\int_{\boldsymbol{x}\in\mathcal{V}}\left(\sum_{\lambda'\in\Lambda}\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{f_{\boldsymbol{N}''}(\boldsymbol{y})f_{\boldsymbol{N}'}((\boldsymbol{y}+\boldsymbol{x})\bmod\Lambda+\lambda')}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (142)\\ &\le \log\int_{\boldsymbol{x}\in\mathcal{V}}\sum_{\lambda'\in\Lambda}\left(\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{f_{\boldsymbol{N}''}(\boldsymbol{y})f_{\boldsymbol{N}'}((\boldsymbol{y}+\boldsymbol{x})\bmod\Lambda+\lambda')}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (143)\\ &= \log\int_{\boldsymbol{x}\in\mathbb{R}^n}\left(\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{f_{\boldsymbol{N}''}(\boldsymbol{y})f_{\boldsymbol{N}'}(\boldsymbol{y}+\boldsymbol{x})}d\boldsymbol{y}\right)^{\frac{1}{\rho}} && (144)\\ &= D_\rho^{\mathrm{Bhatt}}(\boldsymbol{N}'') && (145)\end{aligned}$$

We obtain

$$D_\rho^{\mathrm{Bhatt}}(Z)$$
$$= \log \int_{x\in\mathbb{R}} \left(\int_{y\in\mathbb{R}} \sqrt{f_Z(x) f_Z(x+y)} dy \right)^{\frac{1}{\rho}} dx \tag{153}$$
$$= \log \int_{x\in\mathbb{R}} \left(\int_{y\in\mathbb{R}} \sqrt{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x+y)^2}{2\sigma^2}}} dy \right)^{\frac{1}{\rho}} dx \tag{154}$$
$$= \log \int_{x} \left(\int_{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+x/2)^2}{2\sigma^2}} e^{-\frac{x^2}{2(2\sigma)^2}} dy \right)^{\frac{1}{\rho}} dx \tag{155}$$
$$= \log \int_{x} \left(e^{-\frac{x^2}{2(2\sigma)^2}} \int_{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+x/2)^2}{2\sigma^2}} dy \right)^{\frac{1}{\rho}} dx \tag{156}$$
$$= \log \int_{x} e^{-\frac{x^2}{2(2\sqrt{\rho}\sigma)^2}} dx \tag{157}$$
$$= \log \sqrt{2\pi} 2\sqrt{\rho}\sigma \tag{158}$$
$$= \frac{1}{2}\log 8\pi\sigma^2\rho. \tag{159}$$

Plugging (159) into the left-hand side of (146) and substituting $\delta^* = \frac{e^R}{\sqrt{2\pi e P_X}}$, it is left to show that

$$\sup_{\rho\geq 1} \rho[-\log\delta^* - \frac{1}{2}\log 8\pi\rho\sigma^2] = E_P^r(\mu) \tag{160}$$

for $0 < R \leq \max(0, C - \log 2)$ with $\mu = e^{2(C-R)}$. Differentiating the left side of (160), we get

$$\frac{d}{d\rho}\rho\left[-\log\delta^* - \frac{1}{2}\log 8\pi\rho\sigma^2\right]$$
$$= \left[-\log\delta^* - \frac{1}{2}\log 8\pi\rho\sigma^2\right] - \rho\frac{1}{2}\frac{8\pi\sigma^2}{8\pi\sigma^2\rho} \tag{161}$$
$$= -\log\delta^* - \frac{1}{2}\log 8\pi\sigma^2\rho - \frac{1}{2}. \tag{162}$$

Equating to zero, we obtain

$$\log\frac{1}{\delta^{*2}\cdot 8\pi\sigma^2\rho} = 1 \tag{163}$$

or equivalently

$$\rho = \frac{1}{\delta^{*2}\cdot 8\pi e\sigma^2}. \tag{164}$$

Substituting (164) into (160), we have

$$\sup_{\rho\geq 1} \rho[-\log\delta^* - \frac{1}{2}\log 8\pi\rho\sigma^2]$$
$$= \frac{1}{\delta^{*2}\cdot 8\pi e\sigma^2}\left[-\log\delta^* + \frac{1}{2}\log\delta^{*2}e\right]$$
$$= \frac{1}{\delta^{*2}\cdot 16\pi e\sigma^2}. \tag{165}$$

Taking into account that $\sigma^2 = \frac{P_X P_N}{P_X+P_N}$, we get

$$\sup_{\rho\geq 1} \rho[-\log\delta^* - \frac{1}{2}\log 8\pi\rho\sigma^2] = \frac{1}{\delta^{*2}\cdot 16\pi e\sigma^2} = \frac{\mu}{8}. \tag{166}$$

Finally, note that from (164), we have that $\rho = 1$ corresponds to a rate satisfying

$$1 = \frac{2\pi e P_X/e^{2R}}{8\pi\frac{P_X P_N}{P_X+P_N}} \tag{167}$$

or

$$R = \frac{1}{2}\log\left(\frac{2\pi e P_X}{8\pi e\frac{P_X P_N}{P_X+P_N}}\right) = \frac{1}{2}\log\frac{1+\mathrm{SNR}}{4} = C - \log 2. \tag{168}$$

□

We are now ready to prove Proposition 2.

Proof of Proposition 2: We bound the MLAN expurgated exponent by (169)–(180) at the top of the following page, where (171) follows by Lemma 7, (172) follows by Lemma 10, (178) follows by Lemma 6, and (180) follows from Lemma 8. □

The straight-line part of the bound on $E_\Lambda(R; \boldsymbol{N}')$ in Theorem 4 now follows by combining the results for the random coding exponent and the expurgated exponent.

Lemma 9: For any $\rho > 0$

$$\frac{1}{n}\rho h_{\bar{\rho}}(\boldsymbol{N}'') < \frac{1}{n}\rho h_{\bar{\rho}}(\boldsymbol{Z}^*) + \epsilon_1(\Lambda). \tag{181}$$

Proof: Using Lemma 11, which is proved below, for any $\rho > 0$

$$\frac{1}{n}\rho h_\alpha(\boldsymbol{N}'') \leq \frac{1}{n}\cdot\rho\cdot\frac{1}{\rho}\log\|e^{n\epsilon_1(\Lambda)} f_{\boldsymbol{Z}^*}(\boldsymbol{x})\|_{\bar{\rho}} \tag{182}$$
$$= \frac{1}{n}\log\|f_{\boldsymbol{Z}^*}(\boldsymbol{x})\|_{\bar{\rho}} + \epsilon_1(\Lambda) \tag{183}$$
$$= \frac{1}{n}\rho h_\alpha(\boldsymbol{Z}^*) + \epsilon_1(\Lambda). \tag{184}$$

□

Lemma 10: For any $\rho > 1$

$$\frac{1}{n}\rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{N}'') < \frac{1}{n}\rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{Z}^*) + \epsilon_1(\Lambda). \tag{185}$$

Proof: Using Lemma 11 which is proved below, for any $\rho > 0$

$$\frac{1}{n}\rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{N}'')$$
$$\leq \frac{1}{n}\rho\log\int_{\boldsymbol{x}\in\mathbb{R}^n}\left(\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{e^{n\epsilon_1(\Lambda)} f_{\boldsymbol{Z}^*}(\boldsymbol{y}) e^{n\epsilon_1(\Lambda)} f_{\boldsymbol{Z}^*}(\boldsymbol{y}+\boldsymbol{x})}d\boldsymbol{y}\right)^{\frac{1}{\rho}}$$
$$= \frac{1}{n}\rho\log\int_{\boldsymbol{x}\in\mathbb{R}^n}\left(\int_{\boldsymbol{y}\in\mathbb{R}^n}\sqrt{f_{\boldsymbol{Z}^*}(\boldsymbol{y}) f_{\boldsymbol{Z}^*}(\boldsymbol{y}+\boldsymbol{x})}d\boldsymbol{y}\right)^{\frac{1}{\rho}} + \epsilon_1(\Lambda)$$
$$= \frac{1}{n}\rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{Z}^*) + \epsilon_1(\Lambda). \tag{186}$$

□

Lemma 11:

$$\frac{1}{n}\log\frac{f_{\boldsymbol{N}''}(\boldsymbol{x})}{f_{\boldsymbol{Z}^*}(\boldsymbol{x})} = \epsilon_1(\Lambda). \tag{187}$$

Proof: Let $\boldsymbol{B}$ denote a random vector uniformly distributed over a ball of radius R_u and $f_{\boldsymbol{B}}(\cdot)$ denotes its density

$$f_{\boldsymbol{B}}(\boldsymbol{x}) = \begin{cases} (\gamma_n\cdot R_u^n)^{-1}, & \|\boldsymbol{x}\| \leq R_u \\ 0, & \text{elsewhere} \end{cases} \tag{188}$$

where γ_n is the volume of a unit sphere of dimension n. Since $\boldsymbol{U}$ is uniformly distributed over $\mathcal{V}$, for any $\boldsymbol{x}\in\mathcal{V}$ we have

$$\frac{f_{\boldsymbol{U}}(\boldsymbol{x})}{f_{\boldsymbol{B}}(\boldsymbol{x})} = \frac{\mathrm{Vol}\left(\mathcal{B}(R_u)\right)}{\mathrm{Vol}(\mathcal{V})} = \left(\frac{R_u}{R_\ell}\right)^n. \tag{189}$$

Thus,

$$f_{\boldsymbol{U}}(\boldsymbol{x}) \leq \left(\frac{R_u}{R_\ell}\right)^n f_{\boldsymbol{B}}(\boldsymbol{x}). \tag{190}$$

$$E_{\Lambda}^{\rm ex}(R;\boldsymbol{N}') = \tag{169}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}D_{\rho}^{\rm Bhatt}(\Lambda;\boldsymbol{N}') - R\right] \tag{170}$$

$$\geq \sup_{\rho\geq 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}D_{\rho}^{\rm Bhatt}(\boldsymbol{N}'') - R\right] \tag{171}$$

$$\geq \sup_{\rho\geq 1}\rho\left[\frac{1}{n}\log V - \frac{1}{n}D_{\rho}^{\rm Bhatt}(\boldsymbol{Z}^*) - R\right] - \epsilon_1(\Lambda) \tag{172}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{2}\log\frac{P_X}{G(\Lambda)} - \frac{1}{n}D_{\rho}^{\rm Bhatt}(\boldsymbol{Z}^*) - R\right] - \epsilon_1(\Lambda) \tag{173}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{n}\log 2\pi eP_X - \frac{1}{n}D_{\rho}^{\rm Bhatt}(\boldsymbol{Z}^*) - R - \log 2\pi eG(\Lambda)\right] - \epsilon_1(\Lambda) \tag{174}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{2}\log 2\pi eP_X - \left(D_{\rho}^{\rm Bhatt}(Z^*) + D_{\rho}^{\rm Bhatt}(Z)\right)\right. \tag{175}$$

$$\left. - D_{\rho}^{\rm Bhatt}(Z) - R - \log 2\pi eG(\Lambda)\right] - \epsilon_1(\Lambda) \tag{176}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{2}\log 2\pi eP_X - D_{\rho}^{\rm Bhatt}(Z) - R - \sqrt{\frac{{\rm Var}(Z^*)}{{\rm Var}(Z)}} - \log 2\pi eG(\Lambda)\right] - \epsilon_1(\Lambda) \tag{177}$$

$$\geq \sup_{\rho\geq 1}\rho\left[\frac{1}{2}\log 2\pi eP_X - D_{\rho}^{\rm Bhatt}(Z) - R - \log 2\pi eG(\Lambda) - \log\left(\frac{R_u}{R_\ell}\right)\right] - \epsilon_1(\Lambda) \tag{178}$$

$$= \sup_{\rho\geq 1}\rho\left[\frac{1}{2}\log 2\pi eP_X - D_{\rho}^{\rm Bhatt}(Z) - (R+\epsilon_2(\Lambda))\right] - \epsilon_1(\Lambda) \tag{179}$$

$$= E_P^r\left(e^{2(C-R-\epsilon_2(\Lambda))}\right) - \epsilon_1(\Lambda) \tag{180}$$

We next observe that for any $\boldsymbol{x}$

$$\frac{1}{n}\log\frac{f_{\boldsymbol{B}}(\boldsymbol{x})}{f_{\boldsymbol{Z}_1}(\boldsymbol{x})} \leq \frac{1}{2}\log 2\pi eG_n^* + \frac{1}{n} \tag{191}$$

where $\boldsymbol{Z}_1$ is defined after (114), i.e., is Gaussian having the same second moment as $\boldsymbol{B}$ defined in (188). To see this note that for $\|\boldsymbol{x}\| = \sqrt{n}\sigma$ we have

$$-\frac{1}{n}\log f_{\boldsymbol{Z}_1}(\boldsymbol{x}) = \frac{1}{2}\log 2\pi e\sigma^2 = h(Z_1). \tag{192}$$

Using (114) we have that

$$\frac{f_{\boldsymbol{Z}_1}(R_u)}{f_{\boldsymbol{Z}_1}(\sqrt{n}\sigma)} = \exp\left\{-\frac{(n+2)\sigma^2}{2\sigma^2} + \frac{n\sigma^2}{2\sigma^2}\right\} = \frac{1}{e}. \tag{193}$$

Combining (192) and (193), we get that for $\|\boldsymbol{x}\| = R_u$, we have

$$-\frac{1}{n}\log f_{\boldsymbol{Z}_1}(\boldsymbol{x}) = h(Z_1) - \frac{1}{n}\log\frac{1}{e} = h(Z_1) + \frac{1}{n}. \tag{194}$$

We also have for any $\boldsymbol{x}$ such that $\|x\| \leq R_u$

$$-\frac{1}{n}\log f_{\boldsymbol{B}}(\boldsymbol{x}) = \frac{1}{n}\log{\rm Vol}(\mathcal{B}(R)) \tag{195}$$

$$= \frac{1}{2}\log\frac{\sigma^2}{G_n^*} \tag{196}$$

$$= \frac{1}{2}\log 2\pi e\sigma^2 - \frac{1}{2}\log 2\pi eG_n^* \tag{197}$$

$$= h(Z_1) - \frac{1}{2}\log 2\pi eG_n^*. \tag{198}$$

Since $f_{\boldsymbol{Z}_1}(\boldsymbol{x})$ is monotonically decreasing with $\|\boldsymbol{x}\|$, we have that (194) together with (198) imply that for any $\boldsymbol{x}$

$$\frac{1}{n}\log\frac{f_{\boldsymbol{B}}(\boldsymbol{x})}{f_{\boldsymbol{Z}_1}(\boldsymbol{x})} < \frac{1}{2}\log 2\pi eG_n^* + \frac{1}{n}. \tag{199}$$

We thus get

$$\frac{1}{n}\log\frac{f_{\boldsymbol{U}}(\boldsymbol{x})}{f_{\boldsymbol{Z}_1}(\boldsymbol{x})} = \frac{1}{n}\log\frac{f_{\boldsymbol{U}}(\boldsymbol{x})}{f_{\boldsymbol{B}}(\boldsymbol{x})} + \frac{1}{n}\log\frac{f_{\boldsymbol{B}}(\boldsymbol{x})}{f_{\boldsymbol{Z}_1}(\boldsymbol{x})}$$

$$\leq \log\left(\frac{R_u}{R_\ell}\right) + \frac{1}{2}\log 2\pi eG_n^* + \frac{1}{n}$$

$$= \epsilon_1(\Lambda). \tag{200}$$

Recall that

$$\boldsymbol{N}'' = (1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N} \tag{201}$$

and

$$\boldsymbol{Z}^* = (1-\alpha)\boldsymbol{Z}_1 + \alpha\boldsymbol{N}. \tag{202}$$

It follows from (200)–(202) that

$$\frac{1}{n}\log\frac{f_{\boldsymbol{N}''}(\boldsymbol{x})}{f_{\boldsymbol{Z}^*}(\boldsymbol{x})} \leq \epsilon_1(\Lambda). \tag{203}$$

□

Appendix B
Proof of Lemma 3: Exponent of Truncated Gaussian MLAN Channel

The random coding error exponent of the $(\Lambda, \boldsymbol{Z}_{\mathcal{V}})$-MLAN channel is given by

$$E_{\Lambda}(R;\boldsymbol{Z}_{\mathcal{V}}) = \max_{0\leq\rho\leq 1}\rho\big[\frac{1}{n}\log V - \frac{1}{n}h_{\bar{\rho}}(\boldsymbol{Z}_{\mathcal{V}}) - R\big]. \tag{204}$$

We further have (recall that $\boldsymbol{Z}_{\mathcal{V}}$ is truncated version of $\boldsymbol{Z}^*$, see (84))

$$h_\alpha(\boldsymbol{Z}_{\mathcal{V}}) = \frac{1}{1-\alpha}\log\int_{\mathcal{V}}\left(\frac{f_{\boldsymbol{Z}^*}(\boldsymbol{x})}{1-\epsilon_t(n)}\right)^\alpha d\boldsymbol{x} \tag{205}$$

$$< \frac{1}{1-\alpha}\log\int_{\mathbb{R}^n}\left(\frac{f_{\boldsymbol{Z}^*}(\boldsymbol{x})}{1-\epsilon_t(n)}\right)^\alpha d\boldsymbol{x} \tag{206}$$

$$= \frac{\alpha}{1-\alpha}\log\frac{1}{1-\epsilon_t(n)} + \frac{1}{1-\alpha}\log\int_{\mathbb{R}^n} f_{\boldsymbol{Z}^*}(\boldsymbol{x})^\alpha d\boldsymbol{x}. \tag{207}$$

Taking $\alpha = \bar{\rho} = \frac{1}{1+\rho}$ we get

$$\rho h_{\bar{\rho}}(\boldsymbol{Z}_{\mathcal{V}}) < \rho h_{\bar{\rho}}(\boldsymbol{Z}^*) + \log\frac{1}{1-\epsilon_t(n)}. \tag{208}$$

Similarly, we have

$$\rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{Z}_{\mathcal{V}}) < \rho D_\rho^{\mathrm{Bhatt}}(\boldsymbol{Z}^*) + \log\frac{1}{1-\epsilon_t(n)}. \tag{209}$$

Therefore, following the steps in the proof of Theorem 4 we get

$$E_\Lambda(R;\boldsymbol{Z}_{\mathcal{V}}) = E_P(\mu) - o_n(1). \tag{210}$$

This completes the proof. □

Appendix C
Proof of Lemma 4: Exponent Robust to Fine Quantization of Input

Consider the random coding error exponent corresponding to a uniform distribution over the basic grid $(p^{-1}\cdot\Lambda)\cap\mathcal{V}$. It is given by

$$E_\Lambda^r(R;\boldsymbol{Z}_{\mathcal{V}},p) = \max_{0<\rho\le 1}\left[E_\Lambda^0(R;\boldsymbol{Z}_{\mathcal{V}},p) - \rho R\right] \tag{211}$$

where we have (212) at the bottom of the page. Compare this with the random coding exponent corresponding to a uniform input, which is given by

$$E_\Lambda^r(R;\boldsymbol{Z}_{\mathcal{V}}) = \max_{0<\rho\le 1}\rho\left[E_\Lambda^0(R;\boldsymbol{Z}_{\mathcal{V}}) - \rho R\right] \tag{213}$$

where we have (214) at the bottom of the page. We next show that for any $\boldsymbol{x}\in\mathcal{V}$ and $\boldsymbol{z}\in p^{-1}\cdot\mathcal{V}$

$$\frac{1}{|p^{-1}\cdot\mathcal{V}|}\int_{\boldsymbol{z}\in p^{-1}\cdot\mathcal{V}} f_{Z_{\mathcal{V}}}([\boldsymbol{x}-\boldsymbol{z}] \bmod \Lambda)d\boldsymbol{z} = (1+o_n(1))f_{Z_{\mathcal{V}}}(\boldsymbol{x} \bmod \Lambda) \tag{215}$$

where Λ is assumed to be any Rogers-good lattice. Consider a ball of radius R_ℓ and volume $|\mathcal{B}(R_\ell)| = V$. We have

$$G_n^*\cdot|\mathcal{B}(R_\ell)|^{2/n} = \frac{R_\ell^2}{(n+2)}. \tag{216}$$

Since $P_X = G(\Lambda)V^{2/n}$, this gives

$$R_\ell = \sqrt{(n+2)\frac{G_n^*}{G(\Lambda)}P_X} = \sqrt{\frac{n+2}{n}\frac{G_n^*}{G(\Lambda)}}\cdot\sqrt{nP_X}. \tag{217}$$

For Rogers-good lattices, we have

$$\frac{G(\Lambda)}{G_n^*}\to 1 \quad\text{and}\quad \frac{R_u}{R_\ell} = \mathcal{R}(n)^{1/n} = 1+o_n(1).$$

Combined with (216), this implies that for any $\boldsymbol{x}\in\mathcal{V}$

$$||\boldsymbol{x}|| \le R_u = \mathcal{R}(n)^{1/n}R_\ell = (1+o_n^*(1))\sqrt{nP_X} \tag{218}$$

where here $o_n^*(1) = \sqrt{\frac{n+2}{n}\frac{G_n^*}{G(\Lambda)}}\cdot\mathcal{R}(n)^{1/n} - 1$. Recalling that $p = [e^{nR}]$, we also have for any $\boldsymbol{z}\in p^{-1}\cdot\mathcal{V}$

$$||\boldsymbol{z}|| \le \frac{R_u}{p} = (1+o_n^*(1))\frac{\sqrt{nP_X}}{e^{nR}}. \tag{219}$$

For any $\boldsymbol{x}\in\mathcal{V}$

$$f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x}) = \frac{1}{1-\epsilon_t(n)}\cdot\frac{1}{\sqrt{(2\pi P_{Z^*})^n}}\cdot e^{\frac{-\sum_{i=1}^n x_i^2}{2P_{Z^*}}} \propto e^{\frac{-\sum_{i=1}^n x_i^2}{2P_{Z^*}}}. \tag{220}$$

By the Cauchy–Schwarz inequality

$$||\boldsymbol{x}||^2 - 2||\boldsymbol{x}||||\boldsymbol{z}|| + ||\boldsymbol{z}||^2 \le ||\boldsymbol{x}+\boldsymbol{z}||^2 \le ||\boldsymbol{x}||^2 + 2||\boldsymbol{x}||||\boldsymbol{z}|| + ||\boldsymbol{z}||^2. \tag{221}$$

Therefore,

$$f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x})e^{\frac{-||\boldsymbol{z}||^2-2||\boldsymbol{x}||||\boldsymbol{z}||}{2P_{Z^*}}} \le f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x}+\boldsymbol{z}) \le f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x})e^{\frac{+||\boldsymbol{z}||^2+2||\boldsymbol{x}||||\boldsymbol{z}||}{2P_{Z^*}}}. \tag{222}$$

Now

$$\frac{||\boldsymbol{z}||^2+2||\boldsymbol{x}||||\boldsymbol{z}||}{2P_{Z^*}} \le \frac{nP_X/p^2 + 2nP_X/p}{2P_{Z^*}}(1+o_n^*(1)) = \frac{nP_X}{e^{nR}P_{Z^*}}(1+o_n(1)) = o_n(1). \tag{223}$$

The last two inequalities imply that

$$f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x}+\boldsymbol{z} \bmod \Lambda) = (1+o_n(1))f_{\boldsymbol{Z}_{\mathcal{V}}}(\boldsymbol{x}). \tag{224}$$

$$E_\Lambda^0(R;\boldsymbol{Z}_{\mathcal{V}},p) = -\frac{1}{n}\log\int_{\boldsymbol{y}\in\mathcal{V}}\left(\frac{1}{|(p^{-1}\cdot\Lambda)\cap\mathcal{V}|}\sum_{\boldsymbol{x}\in(p^{-1}\cdot\Lambda)\cap\mathcal{V}} f_{Z_{\mathcal{V}}}([\boldsymbol{y}-\boldsymbol{x}] \bmod \Lambda)^{\frac{1}{1+\rho}}\right)^{1+\rho} d\boldsymbol{y}. \tag{212}$$

$$E_\Lambda^0(R;\boldsymbol{Z}_{\mathcal{V}}) = -\frac{1}{n}\log\int_{\boldsymbol{y}\in\mathcal{V}}\left(\frac{1}{V}\int_{\boldsymbol{x}\in\mathcal{V}} f_{Z_{\mathcal{V}}}([\boldsymbol{y}-\boldsymbol{x}] \bmod \Lambda)^{\frac{1}{1+\rho}}d\boldsymbol{x}\right)^{1+\rho} d\boldsymbol{y} = -\frac{1}{n}\log\int_{\boldsymbol{y}\in\mathcal{V}}\left(\frac{1}{|(p^{-1}\cdot\Lambda)\cap\mathcal{V}|}\sum_{\boldsymbol{x}\in(p^{-1}\cdot\Lambda)\cap\mathcal{V}}\int_{\boldsymbol{z}\in p^{-1}\cdot\mathcal{V}}\frac{1}{|p^{-1}\cdot\mathcal{V}|} f_{Z_{\mathcal{V}}}([\boldsymbol{y}-\boldsymbol{x}-\boldsymbol{z}] \bmod \Lambda)^{\frac{1}{1+\rho}}d\boldsymbol{z}\right)^{1+\rho} d\boldsymbol{y}. \tag{214}$$

Consequently, we obtain (215). Substituting (215) into (214), it follows that

$$E_\Lambda^0(R; \boldsymbol{Z}_\mathcal{V}, p) - E_\Lambda^0(R; \boldsymbol{Z}_\mathcal{V}) = o_n(1). \tag{225}$$

This proves the lemma. □

ACKNOWLEDGMENT

The authors wish to thank G. Poltyrev for early discussions which contributed to the development of some of the ideas in this work. The authors are also indebted to G. D. Forney, Jr. and to H. A. Loeliger for very helpful comments.

REFERENCES

[1] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2201–2214, Aug. 2002.

[2] A. Barg and G. D. Forney Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2568–2573, Sept. 2002.

[3] R. J. Barron, B. Chen, and G. W. Wornell, "On the duality between information embedding and source coding with side information and some applications," in *Proc. Int. Symp. Information Theory (ISIT)*, Washington, DC, June 2001, p. 300.

[4] A. R. Calderbank, "The art of signaling: Fifty years of coding theory," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2561–2595, Oct. 1998.

[5] J. H. Conway, E. M. Rains, and N. J. A. Sloane, "On the existence of similar sublattices," *Canad. J. Math.*, submitted for publication.

[6] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.

[7] ——, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, Mar. 1982.

[8] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[9] R. de Buda, "Some optimal codes have structure," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 893–899, Aug. 1989.

[10] ——, "The upper error bound of a new near-optimal code," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 441–445, July 1975.

[11] U. Erez, S. Litsyn, and R. Zamir, "Lattices that are good for (almost) everything," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Trans. Inform. Theory*, submitted for publication.

[13] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, submitted for publication.

[14] U Erez and S. ten Brink, "Approaching the dirty paper limit for canceling known interference," in *Proc. 41st Annual Allerton Conf. Communication, Control, and Computing*, Allerton House, Monticello, IL, Oct. 2003, pp. 799–808.

[15] U. Erez and R. Zamir, "Error exponents of modulo additive noise channels with side information at the transmitter," *IEEE Trans. Inform. Theory*, vol. 47, pp. 210–218, Jan. 2001.

[16] G. D. Forney Jr., "Multidimensional constellations – Part II: Voronoi constellations," *IEEE J. Select. Areas Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.

[17] ——, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, pp. 281–300, Mar. 1992.

[18] ——, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Allerton House, Monticello, IL, Oct. 2003, pp. 430–439.

[19] ——, "Coset codes-I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.

[20] G. D. Forney Jr., M. D. Trott, and S. Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 820–850, May 2000.

[21] G. D. Forney Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2384–2415, Oct. 1998.

[22] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[23] H. El Gamal, G. Caire, and M. O. Damen, "On the role of MMSE in lattice decoding: Achieving the optimal diversity-vs-multiplexing tradeoff," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Allerton House, Monticello, IL, Oct. 2003, pp. 231–241.

[24] T. Linder, C. Schlegel, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1735–1737, Sept. 1993.

[25] T. Liu, P. Moulin, and R. Koetter, "On error exponents of nested lattice codes for the AWGN channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[26] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1767–1773, Nov. 1997.

[27] T. Philosof, U. Erez, and R. Zamir, "Combined shaping and precoding for interference cancellation at low SNR," in *Proc. Int. Symp. Information Theory (ISIT2003)*, Yokohama, Japan, June 2003, p. 68.

[28] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 409–417, Mar. 1994.

[29] C. A. Rogers, *Packing and Covering*. Cambridge, U.K.: Cambridge Univ. Press, 1964.

[30] D. Slepian, "Group codes for the Gaussian channel," *Bell Syst. Tech. J.*, vol. 47, pp. 575–602, 1968.

[31] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, Jan. 1998.

[32] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1152–1159, July 1996.

[33] ——, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.

[34] ——, "Information rates of pre/post filtered dithered quantizers," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1340–1353, Sept. 1996.

[35] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250–1276, June 2002.