AWGN-Goodness is Enough: Capacity-Achieving Lattice Codes based on Dithered Probabilistic Shaping

Antonio Campello and Daniel Dadush

Abstract

In this paper we show that any sequence of infinite lattice constellations which is good for the unconstrained Gaussian channel can be shaped into a capacity-achieving sequence of codes for the power-constrained Gaussian channel under lattice decoding and non-uniform signalling. Unlike previous results in the literature, our scheme holds with no extra condition on the lattices (e.g. quantization-goodness or vanishing flatness factor), thus establishing a direct implication between AWGN-goodness, in the sense of Poltyrev [17], and capacity-achieving codes. Our analysis uses properties of the discrete Gaussian distribution in order to obtain precise bounds on the probability of error and achievable rates. In particular, we obtain a simple characterization of the finite-blocklength behavior of the scheme, showing that it approaches the optimal dispersion coefficient for *high* signal-to-noise ratio. We further show that for *low* signal-to-noise ratio the discrete Gaussian over centered lattice constellations can not achieve capacity, and thus a shift (or "dither") is essentially necessary.

I. INTRODUCTION

Coded modulation schemes for the Gaussian channel can be usually constructed from infinite constellations in $\mathbb{R}^n$, shaped according to the power constraint of the channel. In order to analyse such infinite constellations independently of the power, Poltyrev [17] defined the notion of codes which are good for the unconstrained Gaussian channel. In this setting, a code is an infinite discrete subset of $\mathbb{R}^n$, and any point can be transmitted. Since the usual code rate is infinite in this case, the optimal *normalized log density* (NLD) replaces the notion of capacity. The NLD measures the logarithm, per dimension, of the number of points of an IC per unit of volume. An optimal sequence of ICs with vanishing probability of error is called *AWGN-good*, and corresponds to the "most economic" constellations that can achieve reliable communication. The most popular ICs are *lattices*, since their symmetries allow for construction of efficient encoding/decoding schemes. Since the work of Poltyrev in 1994, the notion of AWGN-goodness has become an important widely used benchmark and building block for the construction of lattice codes.

The authors are listed in alphabetical order.

A. Campello is with the Department of Electrical and Electronic Engineering, Imperial College London, UK (e mail: a.campello@imperial.ac.uk).

D. Dadush is with the Centrum Wiskunde & Informatika, Netherlands (e mail: dadush@cwi.nl).

Intuitively, AWGN-good ICs should be able to produce capacity-achieving codes in the power-constrained setting, using nearest neighbor decoding and a carefully chosen shaping technique. Nevertheless, all known schemes in the literature that convert lattices into codes for the constrained Gaussian channel under *lattice decoding* entail some additional property. For instance, Erez and Zamir [6] proved that an AWGN-good sequence of lattices can be converted into capacity-achieving codes for the Gaussian channel, *provided that it can be nested to another lattice sequence* which is also AWGN-good and has optimal normalized second-moment. Recently, Ling and Belfiore [12] have shown that a sequence of lattices along with probabilistic shaping can achieve the capacity of the Gaussian channel above a certain signal-to-noise ratio (SNR) *provided that it has vanishing flatness factor*, which relates to how quickly a Gaussian random vector becomes equidistributed over cosets of the lattice as its standard deviation increases.

The objective of this paper is to show that indeed AWGN-goodness is a *sufficient* property for building capacity-achieving lattices in the Gaussian channel, with no extra condition. In order to do so, we employ a technique based on non-equiprobable signalling using discrete Gaussian distributions centered at a (non-zero) randomly generated point in $\mathbb{R}^n$. Following [16], we call such technique *dithered probabilistic shaping* (DPS). From a practical "separation" point of view, this means that the design problem of good lattices for the AWGN channel can be completely *decoupled*: one can focus entirely on the design problem for the unconstrained channel, which can be then coupled with *plug-and-play* DPS techniques into a good code for the constrained channel.

A. Main Result
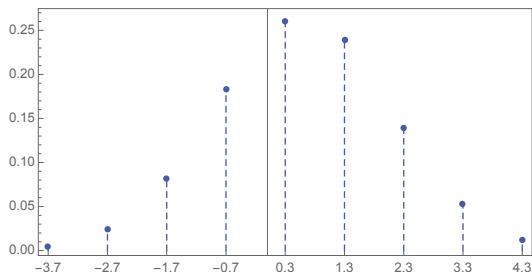
(a) A one-dimensional discrete Gaussian with support in $\mathbb{Z} + 0.3$
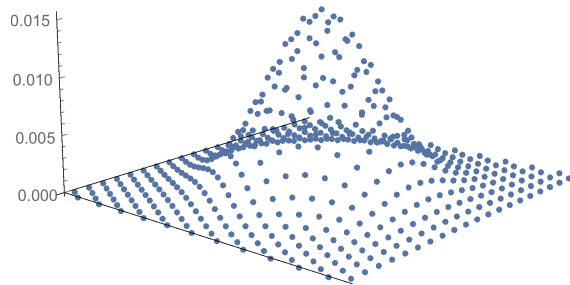
(b) A two-dimensional centered discrete Gaussian with support in a hexagonal lattice

Fig. 1: Illustration of the discrete Gaussian distribution

Let $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2)$ be an n-dimensional Gaussian random vector. A sequence of *AWGN-good* lattices $(\Lambda_n)_{n=1}^\infty$ with volume $V(\Lambda_n)$ is such that the probability of error (i.e. the probability that $\mathbf{W}$ leaves the Voronoi cell of Λ_n) vanishes for any NLD

$$\delta(\Lambda_n) \triangleq -\frac{1}{n} \log V(\Lambda_n) < -\frac{1}{2} \log(2\pi e \sigma_w^2) \triangleq \delta^*. \tag{1}$$

The discrete Gaussian distribution $\mathcal{D}_{\Lambda+\mathbf{t},\sigma_s^2}$ is the distribution taking values in $\Lambda + \mathbf{t}$ whose mass of $\mathbf{x} + \mathbf{t}$ is proportional to $e^{-\|\mathbf{x}+\mathbf{t}\|/2\sigma_s^2}$ (see Figure 1). This distribution has finite power and can thus be used in

the transmission over a Gaussian channel with average power constraint P. In a lattice Gaussian coding scheme, the sent signal is chosen according to $\mathcal{D}_{\Lambda+\mathbf{t},\sigma_s^2}$, while the received points are suitably scaled and then decoded using lattice decoding. Our main result is the following:

Theorem 1 (Informal). *Let $(\Lambda_n)_{n=1}^\infty$ be an AWGN-good sequence of lattices. There exist $(\mathbf{t}_n)_{n=1}^\infty$ such that the lattice coding scheme with distribution $\mathcal{D}_{\Lambda_n+\mathbf{t}_n,P}$ is capacity-achieving in the Gaussian channel.*

A more quantitative statement can be found in Theorem 2, along with an analysis of the rate of convergence to capacity. For instance, we show that the existence of $\mathbf{t}_n$ in Theorem 1 holds with constant probability and, using [10], that the gap to capacity has order $\approx (2n)^{-1/2}(Q^{-1}(\varepsilon) + \sqrt{8})$ where Q^{-1} is the inverse-error function. Apart for the $\sqrt{8}$ term (which we believe is an artifact of our analysis), this corresponds to the optimal convergence behavior for the Gaussian channel for high SNR [18]. In particular, as $\varepsilon \to 0$, the optimal dispersion is approached at high SNR.

A key point for the proof of the main results is the *sampling lemma* (Lemma 2), that says that the dithered lattice Gaussian is equivalent to a continuous Gaussian. As a further point of interest, we observe using [4] that the coding gain of a lattice is closely related to the so-called *smoothing parameter* of the dual lattice. At a high level, the discrete Gaussian over Λ^* sampled above the smoothing parameter "looks like" a continuous Gaussian, which justifies the name. The study of this parameter has played a fundamental role in our understanding of the discrete Gaussian, and has led to many important developments such as tight transference theorems in the geometry of numbers [1], new lattice based cryptographic schemes [9], [15], [19], and the recent development of the so-called reverse Minkowski inequality [20], [21]. This relation, discussed in Section IV, gives an alternative viewpoint for quantifying AWGN-goodness that we hope will find future use.

B. To dither or not to dither

Since the seminal work [6], nested lattice constellations are known to be capacity achieving in the Gaussian channel with lattice decoding, provided that the constellations are suitably shifted by a vector known both at the transmitter and at the receiver. The process of randomizing the choice of the shift vector, known as *dithering*, greatly facilitates the analysis but may introduce additional design complexity. An intriguing open problem is whether dithering is indeed necessary. For Voronoi constellations (under MMSE scaled lattice decoding), this problem is considered in [25], [26], where the necessity of a dither is argued for low SNR.

Recently, Ling and Belfiore [12] have shown that undithered lattice coding, with probabilistic shaping according to a discrete Gaussian distribution, is capacity achieving for a threshold[1] SNR greater than e. In Section V of the present paper, we show that (a non-zero) dither is indeed necessary in the low SNR regime. Specifically we show (Theorem 3) that if a sequence of lattices is shaped according to a *centered* Gaussian distribution with variance parameter equals to the power constraint of the channel, it fails to achieve any positive rates for SNR < 1. This motivates the use of dither in our scheme. Interestingly, in the low SNR regime our scheme uses properties of the discrete Gaussian when the flatness factor is large

---

[1] In fact, the arguments in [12] can be slightly improved in order to reduce the threshold to SNR $> e - 1$

(or "below smoothing", in the computer science jargon), a setting where previous analytical techniques break down. As it turns out, in this regime, dithering is not only a matter of simplification but essentially necessary. It is worth noting that although lattices based on polar codes [24] can achieve capacity for all SNRs, their construction employ randomly generated bits which, ultimately, play the role of a dither.

C. Related Works

Our main result strongly uses non-uniform signalling, which could be implemented with probabilistic shaping techniques. In a broader scope, besides the theoretically appealing properties of probabilistic shaping for Gaussian channels [11], [12], the techniques became popular in the last years due to their prospective applications to non-linear optical-fiber transmissions [7], [22]. The idea of *dithered probabilistic shaping* (DPS) was previously considered in [16], who proposed a method to achieve high shaping gains of a constellation with small dimension and low complexity.

Besides showing the sufficiency of the AWGN goodness figure of merit, our analysis closes the gap of uniform signalling [12] for low SNR. Although this is not the usual regime for wireless channels, it has attracted a recent interest due to secrecy applications in low profile covert communications [23], where signals are transmitted with very low power. Understanding how probabilistic shaping schemes behave for low SNR might find use in such scenarios. In addition to AWGN channel coding, there is an increasing literature in the use of the discrete Gaussian in other information-theoretic scenarios where our simplified approach could play a role. A few examples are Gaussian wiretap channels [13], fading wiretap channels [14], and compound channels [3].

II. NOTATION AND PRELIMINARY RESULTS

We consider a real-valued AWGN channel with average-power constraint P and noise variance σ_w^2. Denote the received signal of the channel, given input $\mathbf{x} \in \mathbb{R}^n$, by $\mathbf{y} = \mathbf{x} + \mathbf{w}$, where $\mathbf{w}$ is drawn from the distribution $\mathcal{N}(0, \sigma_w^2 \mathbf{I}_n)$. A (full-rank) lattice $\Lambda \subset \mathbb{R}^n$ is a discrete subgroup not contained in any proper subspace of $\mathbb{R}^n$.

Definition 1. *A lattice code for the Gaussian channel consists of*
1) *A lattice $\Lambda \subset \mathbb{R}^n$.*
2) *A finite-entropy probability distribution $\mathcal{D}$ taking values in Λ such that, for $\mathbf{X} \sim \mathcal{D}$, $E[\|\mathbf{X}\|^2] \leq nP$.*
3) *A decoding function $g : \mathbb{R}^n \to \Lambda$.*

The error probability of a lattice code is $P_e(\Lambda) \triangleq \Pr(g(\mathbf{X} + \mathbf{W}) \neq \mathbf{X})$, where $\mathbf{W} \sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_n)$. The *rate* of the code is given by $R = (1/n)\mathbb{H}(\mathbf{X})$, where

$$\mathbb{H}(\mathbf{X}) \triangleq - \sum_{\mathbf{x} \in \Lambda} \Pr(\mathbf{X} = \mathbf{x}) \log \Pr(\mathbf{X} = \mathbf{x})$$

is the entropy of distribution $\mathcal{D}$ (all logarithms are with respect to base e and rates are calculated in *nats*). We can similarly define lattice codes for a translation $\Lambda + \mathbf{t}$, where $\mathbf{t} \notin \Lambda$, with the obvious modifications.

If $\mathcal{D}$ is the uniform distribution over the points of Λ in a compact set $\mathcal{S}$ (e.g. a ball or the Voronoi region of a sub-lattice), and zero otherwise, this corresponds to the classic "deterministic" shaping. In this case, a set of possible uniformly distributed messages $\{1, 2, \ldots, e^{nR}\}$ can be mapped into the signal

space by simply labeling the points in $\Lambda \cap \mathcal{S}$. When $\mathcal{D}$ is non-uniform, messages can be mapped into the signal space by means of *probabilistic shaping* techniques (e.g. [16], [12, Sec IV] [2], [26, Sec 6.5]).

Let $\text{SNR} \triangleq P/\sigma_w^2$ and $C(\text{SNR}) \triangleq 1/2\log(1 + \text{SNR})$.

Definition 2. *A sequence of lattice codes of increasing dimension Λ_n with distributions $\mathcal{D}_n$ is said to be capacity-achieving in the Gaussian channel with $\text{SNR} = P/\sigma_w^2$ if $P_e(\Lambda_n) \to 0$ and*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{H}(\mathbf{X}_n) = C(\text{SNR}), \tag{2}$$

where $\mathbf{X}_n \sim \mathcal{D}_n$.

Definition 3. *Let*

$$f_\sigma(\mathbf{x}) = \frac{e^{-\|\mathbf{x}\|^2/2\sigma^2}}{(\sqrt{2\pi\sigma^2})^n}.$$

The discrete Gaussian distribution $\mathcal{D}_{\Lambda+\mathbf{t},\sigma^2}$ is the discrete distribution assuming values in $\Lambda + \mathbf{t}$ such that the mass of each point $\mathbf{x} + \mathbf{t} \in \Lambda + \mathbf{t}$ is proportional to $f_\sigma(\mathbf{x} + \mathbf{t})$.

If $\mathcal{D}$ in Definition 1 is a discrete Gaussian distribution assuming values in $\Lambda + \mathbf{t}$, the corresponding code is called a *lattice Gaussian code*. Some discrete Gaussians are illustrated in Figure 1. If a point in a lattice Gaussian code is transmitted over a Gaussian channel, we have the following fact:

Lemma 1 (Equivalence between MAP and lattice decoder [12]). *Let $\hat{\mathbf{x}}$ be the output of the maximum-a-posteriori decoder for a lattice Gaussian code with distribution $\mathcal{D}_{\Lambda+\mathbf{t},\sigma_s^2}$. We have*

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\Lambda+\mathbf{t}} \|\alpha\mathbf{y} - \mathbf{x}\|,$$

where $\alpha = \sigma_s^2/(\sigma_s^2 + \sigma_w^2)$ is called the MMSE *(or Wiener) coefficient.*

Therefore, the optimal MAP decoder is obtained by MMSE pre-processing followed by lattice decoding. Let $\mathbf{W}_{\text{eff}} = (\alpha - 1)\mathbf{X} + \alpha\mathbf{W}$ be the effective noise in this process. For an n-dimensional lattice $\Lambda \subset \mathbb{R}^n$, we define the Voronoi cell $\mathcal{V}(\Lambda)$ of Λ by

$$\mathcal{V}(\Lambda) = \left\{ \mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{y}\|^2/2, \text{ for any } \mathbf{y} \in \Lambda \right\}$$

namely, the set of all points closer to the origin than any other lattice point. The volume of Λ, denoted by $V(\Lambda)$, corresponds to the volume of the Voronoi cell $\mathcal{V}(\Lambda)$. From Lemma 1, the probability of error of the maximum-a-posteriori decoder is

$$P_e(\Lambda) = \Pr(\mathbf{W}_{\text{eff}} \notin \mathcal{V}(\Lambda)).$$

III. CAPACITY ACHIEVING CODES FROM AWGN-GOOD LATTICES

In this section, we will show that any AWGN-good lattice Λ can be used to construct a capacity achieving code for the AWGN-channel. The caveat will be the use of a random dither as well as the use of non-equiprobable encoding, namely discrete Gaussian encoding. We will later argue on Section V that the dither is essentially necessary for this construction to hold for all SNR.

A. Dithered Coding Scheme

Let us now examine an ensemble of coding schemes in $\mathbb{R}^n$ based on a n-dimensional lattice $\Lambda \subset \mathbb{R}^n$. Define the inverse error function of Λ by

$$\mathrm{err}_\varepsilon^{-1}(\Lambda) = \min \left\{ s : \Pr_{\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)} [x \notin s\mathcal{V}(\Lambda)] \leq \varepsilon \right\} . \tag{3}$$

Note that for a AWGN-Good family $(\Lambda_n)_{n=1}^\infty$, $V(\Lambda_n) = 1 \ \forall i \in [n]$, for any fixed $\varepsilon > 0$ we have that $\lim_{n \to \infty} \mathrm{err}_\varepsilon^{-1}(\Lambda) \to \sqrt{2\pi e}$. We also define the *normalized volume-to-noise ratio* (NVNR) of lattice at probability of error ε as:

$$\mu(\Lambda, \varepsilon) = \mathrm{err}_\varepsilon^{-1}(\Lambda)^2 V(\Lambda)^{2/n}. \tag{4}$$

This definition indeed corresponds to the usual NVNR (e.g. [26, Def. 3.3.3]), rephrased in light of the error function (3). For convenience, we further normalize $\mu(\Lambda, \varepsilon)$ by $2\pi e$ and define

$$\gamma(\Lambda, \varepsilon) = \frac{\mu(\Lambda, \varepsilon)}{2\pi e}. \tag{5}$$

This ratio can be interpreted as the "modulation loss" of Λ. For an AWGN-good sequence, $\log \gamma(\Lambda_n, \varepsilon)$ vanishes as $n \to \infty$.

Let σ_s^2, σ_w^2 denote the average power and error per symbol respectively and let $\sigma_{\mathrm{eff}}^2 := \frac{\sigma_s^2 \sigma_w^2}{\sigma_s^2 + \sigma_w^2}$.

Let $\Lambda \subset \mathbb{R}^n$ be a lattice. Given a fixed probability of error $\varepsilon > 0$, we will now examine a family of dithered coding schemes which on average have decoding error at most 6ε, power per symbol $\sigma_s^2(1 + 4/\sqrt{n})$, and distance to the capacity depending only on the relationship between the inverse error function and the lattice determinant. The existence of a good code from our family will then follow by the union bound appropriately applied.

For the purpose of finding a good code, we shall examine the family of encoding distributions $D_{\Lambda + \mathbf{t}, \sigma_s}$, indexed by $\mathbf{t} \in \mathbb{R}^n$, i.e. the discrete Gaussian on $\Lambda + \mathbf{t}$ of parameter σ_s. For a given $\mathbf{t}$, the decoding function $g_\mathbf{t}$ will be maximum likelihood decoding as Lemma 1, namely for a noisy signal $\mathbf{y} \in \mathbb{R}^n$, the decoded signal corresponds to

$$g_\mathbf{t}(\mathbf{y}) := \arg \min_{\mathbf{x} \in \Lambda + \mathbf{t}} \left\| \mathbf{x} - \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2} \mathbf{y} \right\| .$$

To prove our bounds on the average properties of such codes we need a distribution on shifts $\mathbf{t}$. For this distribution, we pick the natural choice $\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$, i.e. the shift is distributed according to the maximum entropy input distribution satisfying the average power constraint.

The main theorem we will prove in this section is as follows.

Theorem 2. *Let $\Lambda \subset \mathbb{R}^n$ be an n-dimensional lattice, $n \geq 16$. For power $\sigma_s^2 = P$, noise variance $\sigma_w^2 > 0$ and probability of error $\varepsilon > 0$, let $\sigma_{\mathrm{eff}}^2 = \frac{\sigma_s^2 \sigma_w^2}{\sigma_s^2 + \sigma_w^2}$, $s_\varepsilon = \mathrm{err}_\varepsilon^{-1}(\Lambda) \cdot \sigma_{\mathrm{eff}}$ and $\gamma(\Lambda, \varepsilon) = \mathrm{err}_\varepsilon^{-1}(\Lambda)^2 V(\Lambda)^{2/n}/(2\pi e)$. Then for $\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$ with probability at least $1/2$ the coding distribution $\mathbf{X} \sim D_{s_\varepsilon \Lambda + \mathbf{T}, \sigma_s}$ equipped with maximum likelihood decoding satisfies:*

 1) *The decoding error is bounded by 6ε.*
 2) *The squared power per symbol is at most $(1 + 4/\sqrt{n})\sigma_s^2$.*
 3) *The per symbol gap to capacity is bounded by $\frac{1}{2} \log \gamma(\Lambda, \varepsilon) + \frac{2}{\sqrt{n}} + \frac{4}{n}$.*

For notational convenience, we will assume in the remainder that $\sigma_{\text{eff}} \cdot \text{err}_\varepsilon^{-1}(\Lambda) = 1$ (note that this can be achieved by simply scaling the lattice). With this normalization, we will be able to achieve a low probability of error by choosing the codes directly from Λ (instead of a scaling of Λ).

To prove the theorem, we will rely on the subsequent lemmas which characterize the behavior of a randomly dithered channel. We first present these lemmas and prove Theorem 2 at the end of the section.

The following simple sampling lemma shows the random variables corresponding to either sampling $\mathbf{T}$ from $\mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$ and a) returning $\mathbf{T}$ or b) returning a discrete Gaussian sample from $D_{\Lambda+\mathbf{T}, \sigma_s}$ have the same distribution.

Lemma 2 (Sampling Lemma). *Let $\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$ and let $\mathbf{X} \sim D_{\Lambda+\mathbf{T}, \sigma_s}$. Then $\mathbf{T}$ and $\mathbf{X}$ are identically distributed.*

Proof. We need only show that $\mathbf{X}$ has the same probability density as $\mathbf{T}$. For $\mathbf{w} \in \mathbb{R}^n$, note that $\mathbf{X}$ can only hit $\mathbf{w}$ if $\mathbf{w} \in \Lambda + \mathbf{T} \Leftrightarrow \mathbf{T} \in \Lambda + \mathbf{w}$. Therefore, for any measurable set $A \subseteq \mathbb{R}^n$, we have that

$$
\begin{aligned}
\Pr[\mathbf{X} \in A] &= \int_{\mathbb{R}^n} f_{\sigma_s}(\mathbf{t}) \Pr[Y \in A | T = \mathbf{t}] \mathrm{dt} \\
&= \int_{\mathbb{R}^n} f_{\sigma_s}(\mathbf{t}) \frac{f_{\sigma_s}(A \cap (\Lambda + \mathbf{t}))}{f_{\sigma_s}(\Lambda + \mathbf{t})} \mathrm{dt} \\
&= \int_{\mathcal{V}(\Lambda)} \sum_{\mathbf{x} \in \Lambda + \mathbf{c}} f_{\sigma_s}(\mathbf{x}) \frac{f_{\sigma_s}(A \cap (\Lambda + \mathbf{c}))}{f_{\sigma_s}(\Lambda + \mathbf{c})} \mathrm{dc} \\
&= \int_{\mathcal{V}(\Lambda)} f_{\sigma_s}(A \cap (\Lambda + \mathbf{c})) \mathrm{dc} = \int_A f_{\sigma_s}(\mathbf{x}) \mathrm{dx},
\end{aligned}
$$

as needed. $\qquad\square$

Remark 1. *Note that the sampling lemma still holds if we take $\mathbf{T} \sim \mathcal{N}(0, \sigma_s^2 \mathbb{I}_n) \bmod \Lambda$, where the mod-$\Lambda$ operation maps a point in $\mathbb{R}^n$ to a coset representative in the Voronoi cell of Λ. The ℓ_∞ distance between $\mathcal{N}(0, \sigma_s^2 \mathbb{I}_n) \bmod \Lambda$ and a uniform distribution in $\mathcal{V}(\Lambda)$, normalized by $1/V(\Lambda)$, is the so-called flatness factor (e.g. [12]). If the flatness factor of Λ is small, roughly speaking, one could replace $\mathbf{T}$ in Lemma 2 by a uniform dither. As we will see later in Section V, this condition is too stringent for low SNR.*

We now show that expected probability of error of our lattice family is exactly ε. In what follows, $\mathbf{T} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$ will be our random shift of Λ, $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbb{I}_n)$ will be the channel noise, $\mathbf{X} \sim D_{\Lambda+\mathbf{T}, \sigma_s}$ will denote the coding distribution, and $g_\mathbf{T}$ is our decoding function. We recall our assumption that $\sigma_{\text{eff}} \cdot \text{err}_\varepsilon^{-1}(\Lambda) = 1$.

Lemma 3 (Error Probability Bound). *For $\gamma \geq 1$,*

$$
\Pr_{\mathbf{T}}[\Pr_{\mathbf{X}}[g_\mathbf{T}(\mathbf{X} + \mathbf{W}) \neq \mathbf{X}] \geq \gamma\varepsilon] \leq 1/\gamma .
$$

Proof. Let $\alpha = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2}$. Recall that $g_\mathbf{T}(\mathbf{X} + \mathbf{W}) \neq \mathbf{X}$ if and only if

$$
(1 - \alpha)\mathbf{X} + \alpha\mathbf{W} \notin \mathcal{V}(\Lambda) .
$$

By the sampling lemma $\mathbf{X}$ has distribution $\mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbb{I}_n)$. Given that $\mathbf{X}$ and $\mathbf{W}$ are independent Gaussians, we have that $(1-\alpha)\mathbf{X} + \alpha\mathbf{W}$ is distributed as $\mathcal{N}(\mathbf{0}, \sigma_{\text{eff}}^2 \mathbb{I}_n)$ since by construction

$$\sigma_{\text{eff}}^2 := \frac{\sigma_s^2 \sigma_w^2}{\sigma_s^2 + \sigma_w^2} = (1-\alpha)^2 \sigma_s^2 + \alpha^2 \sigma_w^2 \ .$$

By assumption on Λ we have that $\sigma_{\text{eff}} \cdot \text{err}_\varepsilon^{-1}(\Lambda) = 1$, and thus

$$\Pr[(1-\alpha)\mathbf{X} + \alpha\mathbf{W} \notin \mathcal{V}(\Lambda)] = \Pr_{\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_n)}[\sigma_{\text{eff}} \mathbf{Z} \notin \mathcal{V}(\Lambda)]$$

$$= \Pr_{\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_n)}[\mathbf{Z} \notin \text{err}_\varepsilon^{-1}(\Lambda)\mathcal{V}(\Lambda)] = \varepsilon \ .$$

Using the above, by Markov's inequality we have that

$$\Pr_{\mathbf{T}}[\Pr_{\mathbf{X}}[g_{\mathbf{T}}(\mathbf{X}+\mathbf{W}) \neq \mathbf{X}] \geq \gamma\varepsilon] \leq \frac{\Pr[g_{\mathbf{T}}(\mathbf{X}+\mathbf{W}) \neq \mathbf{X}]}{\gamma\varepsilon} = \frac{1}{\gamma} \ ,$$

as needed. $\qquad\square$

The next lemma shows that average power per symbol is very close to the desired limit. The proof proceeds by a comparison to the continuous Gaussian followed by a standard Chernoff bound.

Lemma 4. *For any $\varepsilon > 0$, we have that*

$$1. \Pr_{\mathbf{T}}[\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}]] \geq (1+\varepsilon)n] \leq e^{-(\varepsilon^2/4 - \varepsilon^3/6)n}$$

$$2. \Pr_{\mathbf{T}}[\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}]] \leq (1-\varepsilon)n] \leq e^{-(\varepsilon^2/4 + \varepsilon^3/6)n}$$

Proof. We first prove the upper bound. Since $\mathbf{X}/\sigma_s \sim \mathcal{N}(0, \mathbb{I}_n)$, a standard computation reveals that $\mathbb{E}[e^{\alpha\|\mathbf{X}/\sigma_s\|^2}] = (1-2\alpha)^{-n/2}$ for $\alpha < 1/2$. By the Chernoff bound

$$\Pr_{\mathbf{T}}[\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}] \geq (1+\varepsilon)n] \leq \min_{\alpha \in (0, 1/2)} \frac{\mathbb{E}_{\mathbf{T}}[e^{\alpha\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}]}]}{e^{\alpha(1+\varepsilon)n}}$$

$$\leq \min_{\alpha \in (0, 1/2)} \frac{\mathbb{E}[e^{\alpha\|(\mathbf{X}/\sigma_s)\|^2}]}{e^{\alpha(1+\varepsilon)n}} \quad (\text{ by Jensen's inequality })$$

$$= \min_{\alpha \in (0, 1/2)} \left(\frac{e^{-(1+\varepsilon)\alpha}}{\sqrt{1-2\alpha}} \right)^n$$

$$= (\sqrt{1+\varepsilon}e^{-\varepsilon/2})^n < e^{-(\varepsilon^2/4 - \varepsilon^3/6)n} \left(\text{ setting } \alpha = \frac{\varepsilon}{2(1+\varepsilon)} \right) \ .$$

Similarly, for the lower bound, we have

$$\Pr_{\mathbf{T}}[\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}] \leq (1-\varepsilon)n] \leq \min_{\alpha > 0} \frac{\mathbb{E}_{\mathbf{T}}[e^{-\alpha\mathbb{E}[\|(\mathbf{X}/\sigma_s)\|^2|\mathbf{T}]}]}{e^{-\alpha(1-\varepsilon)n}}$$

$$\leq \min_{\alpha > 0} \frac{\mathbb{E}[e^{-\alpha\|(\mathbf{X}/\sigma_s)\|^2}]}{e^{\alpha(1-\varepsilon)n}} \quad (\text{ by Jensen's inequality })$$

$$= \min_{\alpha > 0} \left(\frac{e^{(1-\varepsilon)\alpha}}{\sqrt{1+2\alpha}} \right)^n$$

$$= (\sqrt{1-\varepsilon}e^{\varepsilon/2})^n < e^{-(\varepsilon^2/4 + \varepsilon^3/6)n} \left(\text{ setting } \alpha = \frac{\varepsilon}{2(1-\varepsilon)} \right) \ .$$

□

We now argue that the entropy of the coding distribution is large with good probability. In particular, we would like to know that for most choices $\mathbf{t}$ for $\mathbf{T}$ that $\mathbb{H}(\mathbf{X}|\mathbf{T} = \mathbf{t})$ is large. Recall that $\mathbf{X}|\mathbf{T} = \mathbf{t}$ is distributed as $D_{\Lambda+\mathbf{t},\sigma_s}$, where a direct computation gives

$$
\begin{aligned}
\mathbb{H}(\mathbf{X}|\mathbf{T} = \mathbf{t}) &= \sum_{\mathbf{y}\in\Lambda+\mathbf{t}} -\log\left(\frac{f_{\sigma_s}(\mathbf{y})}{f_{\sigma_s}(\Lambda+\mathbf{t})}\right)\frac{f_{\sigma_s}(\mathbf{y})}{f_{\sigma_s}(\Lambda+\mathbf{t})} \\
&= \log\left((\sqrt{2\pi\sigma_s^2})^n f_{\sigma_s}(\Lambda+\mathbf{t})\right) + \frac{1}{2}\mathbb{E}_{\mathbf{Z}\sim D_{\Lambda+\mathbf{t},\sigma_s}}[\|\mathbf{Z}/\sigma_s\|^2] \ .
\end{aligned}
\tag{6}
$$

Thus, to prove that the entropy is large we must show that both $f_{\sigma_s}(\Lambda+\mathbf{t})$ and $\mathbb{E}[\|\mathbf{Z}/\sigma_s\|^2]$ are large with good probability over $\mathbf{t}$. Note that the latter condition is essentially given by Lemma 4, so we focus now on the former.

The following lemma uses a bound on the negative moment of $f_{\sigma_s}(\Lambda+\mathbf{t})$ to show that it is unlikely to be too small.

Lemma 5.

$$
\Pr_{\mathbf{T}}[f_{\sigma_s}(\Lambda+\mathbf{T}) \leq \frac{\varepsilon}{V(\Lambda)}] \leq \varepsilon \ .
$$

Proof. To begin, we have that

$$
\begin{aligned}
\mathbb{E}_{\mathbf{T}}[f_{\sigma_s}(\Lambda+\mathbf{T})^{-1}] &= \int_{\mathbb{R}^n} f_{\sigma_s}(\mathbf{t})f_{\sigma_s}(\Lambda+\mathbf{t})^{-1}\mathrm{dt} = \int_{\mathbb{R}^n/\Lambda}\sum_{\mathbf{t}\in\Lambda+\mathbf{c}} f_{\sigma_s}(\mathbf{t})f_{\sigma_s}(\Lambda+\mathbf{c})^{-1}\mathrm{dc} \\
&= \int_{\mathbb{R}^n/\Lambda} f_{\sigma_s}(\Lambda+\mathbf{c})f_{\sigma_s}(\Lambda+\mathbf{c})^{-1}\mathrm{dc} = \int_{\mathbb{R}^n/\Lambda}\mathrm{dc} = V(\Lambda) \ .
\end{aligned}
$$

By Markov's inequality, we have

$$
\Pr_{\mathbf{T}}[f_{\sigma_s}(\Lambda+\mathbf{T}) \leq \frac{\varepsilon}{V(\Lambda)}] \leq \varepsilon\,\frac{\mathbb{E}_{\mathbf{T}}[f_{\sigma_s}(\Lambda+\mathbf{T})^{-1}]}{V(\Lambda)} = \varepsilon \ ,
$$

as needed. □

Proof of Theorem 2. By scaling the lattice, we may assume as above that $\sigma_{\mathrm{eff}}\cdot\mathrm{err}_\varepsilon^{-1}(\Lambda) = 1$ and thus that $s_\varepsilon = 1$. By Lemmas 3, 4, 5 we have that

a. $\Pr_{\mathbf{T}}[\Pr_{\mathbf{X}}[g_{\mathbf{T}}(\mathbf{X}+\mathbf{W}) \neq \mathbf{X}] \geq 6\varepsilon] \leq 1/6$.
b. $\Pr_{\mathbf{T}}[\mathbb{E}[\|\mathbf{X}/\sigma_s\|^2|\mathbf{T}] \geq n + 4\sqrt{n}] \leq e^{-4+64/(6\sqrt{n})} \leq e^{-4/3}$ for $n \geq 16$ (setting $\varepsilon = 4/\sqrt{n}$).
c. $\Pr_{\mathbf{T}}[\mathbb{E}[\|\mathbf{X}/\sigma_s\|^2|\mathbf{T}] \leq n - 4\sqrt{n}] \leq e^{-4}$ (setting $\varepsilon = 4/\sqrt{n}$).
d. $\Pr_{\mathbf{T}}[f_{\sigma_s}(\Lambda+\mathbf{T}) \leq \frac{1}{e^4 V(\Lambda)}] \leq e^{-4}$.

Since $1/6 + e^{-4/3} + 2e^{-4} \leq 1/2$, by the union bound we have $\mathbf{T}$ satisfies the complement of all the above events with probability at least $1/2$. Let $\mathbf{t} \in \mathbb{R}^n$ be such a setting of $\mathbf{T}$. Clearly, the complement of (a) implies that the decoding error is at most 6ε, and the complement of (b) implies that averaged squared power per symbol is at most $\frac{1}{n}(n + 4\sqrt{n})\sigma_s^2 = (1 + 4/\sqrt{n})\sigma_s^2$.

It now remains to show the gap to capacity, i.e. a lower bound on the entropy of $D_{\Lambda+\mathbf{t},\sigma_s}$. By equation 6 and the complement of (c) and (d), we have that

$$
\begin{aligned}
\frac{1}{n}\mathbb{H}(\mathbf{X}|\mathbf{T}=\mathbf{t}) &= \frac{1}{n}\log\left((\sqrt{2\pi\sigma_s^2})^n f_{\sigma_s}(\Lambda+\mathbf{t})\right) + \frac{1}{2n}\mathbb{E}_{\mathbf{Z}\sim D_{\Lambda+\mathbf{t},\sigma_s}}[\|\mathbf{Z}/\sigma_s\|^2] \\
&\geq \log\left(\sqrt{2\pi\sigma_s^2}/(e^{4/n}V(\Lambda)^{1/n})\right) + \frac{1}{2}(1-4/\sqrt{n}) \\
&= \frac{1}{2}\log\left(2\pi e\sigma_s^2/V(\Lambda)^{2/n}\right) - (2/\sqrt{n}+4/n) \\
&= \frac{1}{2}\log\left(\sigma_s^2\mathrm{err}^{-1}(\Lambda)^2/\gamma(\Lambda,\varepsilon))\right) - (2/\sqrt{n}+4/n) \\
&= \frac{1}{2}\log\left(1+\sigma_s^2/\sigma_w^2\right) - (1/2\log\gamma(\Lambda,\varepsilon)+2/\sqrt{n}+4/n) \quad \left(\sigma_{\mathrm{eff}}\cdot\mathrm{err}_\varepsilon^{-1}(\Lambda)=1\right).
\end{aligned}
$$

The theorem now follows recalling that the capacity of the Gaussian channel is $\frac{1}{2}\log(1+\sigma_s^2/\sigma_w^2)$. $\qquad\square$

IV. DISCUSSION

A. Finite Blocklength

Next we discuss how the behavior of the proposed scheme compares to the best finite-blocklength codes for the Gaussian channel. The optimal rate to which a code can converge to capacity is given, in nats per dimension, by [18]:

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

where

$$
V = \lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{-n(C-R)^2}{2\log\varepsilon} = \frac{1}{2}\left(1-\frac{1}{(\mathrm{SNR}+1)^2}\right)
$$

is called the *dispersion* of the channel and ε is the target probability of error. Notice that $V\approx 1/2$ for high signal-to-noise ratio. There is an analogous result for unconstrained constellations, in which case the dispersion is exactly $1/2$. If $\delta_\varepsilon(n)$ denotes the maximum NLD of a constellation for which the probability of error is at most ε, then [10]:

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

where $\delta^* = -\frac{1}{2}\log\left(2\pi e\sigma_w^2\right)$ is the Poltyrev limit. The value $\delta_\varepsilon(n)$ also dictates the optimal behavior of an AWGN-good sequence of lattices. Indeed,

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

Therefore for high SNR the lattice Gaussian scheme approaches optimal dispersion, provided that it is coupled with an infinite lattice with optimal behavior. We further observe that [26, p 131] conjectured that the gap to capacity of non-equiprobable signalling is upper bounded by

$$
\frac{1}{2}\log\left(\mu(\Lambda,\varepsilon)G(\Lambda)\right),
$$

where $G(\Lambda)$ is the *normalized-second moment* of Λ. Noting that $G(\Lambda) > 1/2\pi e$, up to lower order terms which vanish as $n\to\infty$, the expression for the gap to capacity in Theorem 2 supports the conjecture.

B. Relation to the Smoothing Parameter

We now record a new connection between AWGN-Goodness of a lattice and the smoothing parameter of its *dual lattice*. As explained in the last two sections, a useful quantitative measure of the AWGN-Goodness of a lattice Λ with respect to a desired target error ε is its *normalized volume to noise ratio* $\gamma(\Lambda, \varepsilon)$, where we recall that up to lower order terms, $\ln(\gamma(\Lambda, \varepsilon))/2$ upper bounds the gap to capacity of the discrete Gaussian coding scheme on Λ (appropriately scaled and shifted). We also recall the definition of *dual* lattice

$$\Lambda^* = \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{Z} \text{ for all } \mathbf{y} \in \Lambda\} .$$

The smoothing parameter $\eta_\varepsilon(\Lambda^*)$ of Λ^* [2] is defined to be the unique scaling $s > 0$ such that

$$\sum_{x \in \Lambda \backslash \{0\}} e^{-\|sx\|^2/2} = \varepsilon .$$

At a high level, the discrete Gaussian over Λ^* sampled *above the smoothing parameter* "looks like" a continuous Gaussian, which justifies the name.

The connection to the inverse error function was given in [4], in the context of understanding the complexity of approximating the smoothing parameter. It can be expressed as follows:

Lemma 6. *[4] For $\varepsilon \in [0, 1]$ and $\Lambda \subset \mathbb{R}^n$ an n-dimensional lattice, we have that:*

$$\eta_{\varepsilon/(1-\varepsilon)}(\Lambda^*) \leq \mathrm{err}_\varepsilon^{-1}(\Lambda) \leq 2 \cdot \eta_\varepsilon(\Lambda^*) .$$

The right hand side corresponds essentially to the union bound, though over the entire lattice instead of just the facets of the Voronoi cell, which should be tight when ε is very small (say exponentially small in the lattice dimension). The left hand side inequality is derived by estimating the Gaussian mass of lattice shifts of the Voronoi cell, and for this side it is unclear to us when it can be tight. Unfortunately, even for random lattices both sides can fail to be tight. In particular, for an n-dimensional random lattice Λ of determinant 1 and fixed constant error probability ε, $\eta_\varepsilon(\Lambda^*) \approx \sqrt{2\pi}$, whereas $\mathrm{err}_\varepsilon^{-1}(\Lambda) \approx \sqrt{2\pi e}$, thus the inverse error function is a $\sqrt{e}$ factor bigger than the dual smoothing parameter. Nevertheless, the above approximate characterization gives a new and possibly easier to check design criterion for good coding lattices.

We leave it as an open problem to understand whether there is a tighter connection between the smoothing parameter and the inverse error function. In particular, an interesting question is whether $\eta_\varepsilon(\Lambda^*) \approx \sqrt{2\pi} V(\Lambda)^{-1/n}$ [3] implies that $\mathrm{err}_\varepsilon^{-1}(\Lambda) \approx \sqrt{2\pi e} V(\Lambda)^{-1/n}$.

V. CONVERSE

In Section III, we have shown that the shifted (or "dithered") lattice Gaussian is capacity-achieving. To close this paper, we will argue that for very low signal-to-noise ratio the shift is essentially necessary. We will prove the following "converse":

---

[2] We normalize the smoothing parameter with $1/2$ in the exponent here instead of the usual π for simplicity.

[3] It is not hard to check via the Poisson summation formula that this is in fact a lower bound on the smoothing parameter.

Theorem 3. *Let $(\Lambda_n)_{n=0}^{\infty}$ be a sequence of Gaussian lattice codes, with corresponding distributions $D_{\Lambda_n, \sigma_s^2 = P}$. If $\sigma_s^2 / \sigma_w^2 < 1$ and*

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{H}(\mathbf{X}_n) = C(\text{SNR}), \tag{7}$$

then the probability of error $P_e(\Lambda_n)$ of the maximum-a-posteriori decoder is bounded away from 0.

Before exhibiting the proof, we provide a heuristic argument that justifies why Theorem 3 should be true for random lattices. The average of the Gaussian sum $f_{\sigma_s}(\Lambda)$ over a typical "random" lattice of volume V satisfies (e.g. [13, Lemma 3]):

$$\mathbb{E}\left[f_{\sigma_s}(\Lambda)\right] = (2\pi\sigma_s)^{-n/2} + V^{-1} \int f_{\sigma}(\mathbf{x}) \mathrm{d}x = (2\pi\sigma_s)^{-n/2} + V^{-1}.$$

For a sequence of lattices to have vanishing probability of error, the Poltyrev limit for the best NLD (1) implies that $V(\Lambda)^{2/n} > (2\pi e)\sigma_{\text{eff}}^2$ where $\sigma_{\text{eff}}^2 = \sigma_s^2 \sigma_w^2 / (\sigma_s^2 + \sigma_w^2)$ is the effective noise. Under this condition, the typical distribution D_{Λ, σ_s^2} of a random lattice has essentially all mass concentrated in the zero vector. Indeed, we have:

$$\mathbb{E}[P(\mathbf{X} = 0)] = \mathbb{E}[(\sqrt{2\pi\sigma_s^2})^{-n} / f_{\sigma_s}(\Lambda)] \geq (\sqrt{2\pi\sigma_s^2})^{-n} / \mathbb{E}[f_{\sigma_s}(\Lambda)] =$$

$$\overset{(a)}{=} (\sqrt{2\pi\sigma_s^2})^{-n} \left(\frac{(\sqrt{2\pi\sigma_s^2})^n}{V(\Lambda)^{-1}(\sqrt{2\pi\sigma_s^2})^n + 1} \right) \geq \frac{1}{\left(\frac{1}{e}\left(1 + \frac{\sigma_s^2}{\sigma_w^2}\right) \right)^{\frac{n}{2}} + 1} \to 1,$$

where the last limit holds for $\text{SNR} < e - 1$. In this case, it is not hard to see that the entropies of $\mathcal{D}_{\Lambda, \sigma_s^2}$ necessarily tends to zero, therefore no positive rate is achievable. Following this intuition, we prove Theorem 1 by showing that a centered lattice Gaussian with vanishing error probability has essentially all mass concentrated in the origin.

A. Proof of the Converse

We recall the definition of the effective noise $\mathbf{W}_{\text{eff}} = (\alpha - 1)\mathbf{X} + \alpha\mathbf{W}$, where $\alpha = \sigma_s^2 / (\sigma_s^2 + \sigma_w^2)$ is the MMSE coefficient. A tedious yet straightforward calculation gives the precise distribution of $\mathbf{w}_{\text{eff}}$:

Lemma 7. *The probability density function g of the effective noise $\mathbf{W}_{\text{eff}}$ is given by:*

$$g(\mathbf{w}_{\text{eff}}) = \frac{e^{\frac{-\|\mathbf{w}_{\text{eff}}\|^2}{2\sigma_{\text{eff}}^2}}}{\left(\sqrt{2\pi}\sigma_{\text{eff}}\right)^n} \frac{f_{\sqrt{\alpha}\sigma_s}(\Lambda + \mathbf{w}_{\text{eff}})}{f_{\sigma_s}(\Lambda)}, \tag{8}$$

where $\sigma_{\text{eff}}^2 = \sigma_w^2 \sigma_s^2 / (\sigma_w^2 + \sigma_s^2)$ is the variance of $\mathbf{w}_{\text{eff}}$.

Proof. See the appendix. $\square$

Lemma 8 (Relation between entropy and the mass of zero)**.** *Let $P_0(\Lambda)$ denote the probability that $\mathbf{X} = 0$, where $\mathbf{X} \sim \mathcal{D}_{\Lambda, \sigma_s^2}$. We have*

$$\frac{1}{n}\mathbb{H}(\mathbf{X}) \leq -\frac{1}{n}\log(P_0(\Lambda)) + \pi(1 - P_0(\Lambda)) + \frac{1.8e^{-1.7n}}{n}. \tag{9}$$

Proof. We will show the more general bound

$$\frac{1}{n}\mathbb{H}(\mathbf{X}) \leq -\frac{1}{n}\log P_0(\Lambda) + \alpha(1 - P_0(\Lambda)) + \frac{1}{n\phi(\alpha)}e^{-\alpha\phi(\alpha)n}, \tag{10}$$

for $\phi(\alpha) = 1 - (1/2\alpha)\log(2\alpha e)$ and any $\alpha > 1$. The lemma will follow by taking $\alpha = \pi$.

From the definition of entropy,

$$\frac{1}{n}\mathbb{H}(\mathbf{X}) = \frac{1}{n}\log\left((\sqrt{2\pi}\sigma_s)^n f_{\sigma_s}(\Lambda)\right) + \frac{1}{2n\sigma_s^2}E\left[\|\mathbf{X}\|^2\right]. \tag{11}$$

First notice that $P_0(\Lambda) = ((\sqrt{2\pi}\sigma_s)^n f_{\sigma_s}(\Lambda))^{-1}$, therefore the terms inside the logarithms in Eqs. (10) and (11) coincide. Now using Lemma 2.13 in [5], for $t \geq 1$, we have

$$P\left(\frac{1}{2n\sigma_s^2}\|\mathbf{X}\|^2 \geq t\right) \leq e^{-nt + \frac{n}{2}\log(2te)}.$$

Moreover $P\left(\|\mathbf{X}\|^2 \geq 2nt\sigma_s^2\right) \leq 1 - P_0(\Lambda)$. Therefore we obtain bound

$$E\left[\left\|\mathbf{X}/\sqrt{2n\sigma_s^2}\right\|^2\right] \leq (1 - P_0(\Lambda))\alpha + \int_\alpha^\infty P\left(\frac{1}{2n\sigma_s^2}\|\mathbf{X}\|^2 \geq t\right)dt$$

$$\leq (1 - P_0(\Lambda))\alpha + \int_\alpha^\infty e^{-nt\phi(\alpha)}dt$$

Evaluating the integral gives us (10).

$\square$

Probability of Error Analysis. Lemma 7 allows us to relate the probability that the effective noise lies outside the Voronoi cell of a lattice and the Gaussian mass of the point $\mathbf{0} \in \Lambda$. To this purpose we first note that Lemma 7 implies the following relations:

$$f_{\sigma_s}(\Lambda) = \int_{\mathbb{R}^n} \frac{e^{-\|\mathbf{w}\|^2/2\sigma_{\text{eff}}^2}}{(\sqrt{2\pi}\sigma_{\text{eff}})^n} f_{\sqrt{\alpha}\sigma_s}(\Lambda + \mathbf{w})d\mathbf{w} = \int_{\mathcal{V}(\Lambda)} f_{\sigma_{\text{eff}}}(\Lambda + \mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\Lambda + \mathbf{w})d\mathbf{w}. \tag{12}$$

and

$$P_e(\Lambda)f_{\sigma_s}(\Lambda) = \int_{\mathcal{V}(\Lambda)} f_{\sigma_{\text{eff}}}(\Lambda \backslash \{\mathbf{0}\} + \mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\Lambda + \mathbf{w})d\mathbf{w}. \tag{13}$$

Recalling that $P_0(\Lambda) = ((\sqrt{2\pi}\sigma_s)^n f_{\sigma_s}(\Lambda))^{-1}$, the above equality indeed relates the probability of the discrete Gaussian hitting zero and the probability of error. We will first bound $f_{\sigma_s}(\Lambda)$ which will imply a bound on $P_0(\Lambda)$ and on the entropy, by Lemma 8, showing the assertion in Theorem 3.

Proof of Theorem 3: We have

$$f_{\sigma_s}(\Lambda_n) = \int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\Lambda_n \backslash \{\mathbf{0}\} + \mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\Lambda_n + \mathbf{w})d\mathbf{w} + \int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\mathbf{w})d\mathbf{w}$$

$$+ \int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\Lambda_n \backslash \{\mathbf{0}\} + \mathbf{w})d\mathbf{w} \tag{14}$$

We proceed to bound the three terms in the right-hand side of (14). The first term is equal to $P_e(\Lambda_n)f_{\sigma_s}(\Lambda_n)$ while the second term satisfies

$$\int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\mathbf{w})d\mathbf{w} \leq \int_{\mathbb{R}^n} f_{\sigma_{\text{eff}}}(\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\mathbf{w})d\mathbf{w} = \frac{1}{(\sqrt{2\pi}\sigma_s)^n}.$$

Noting that $\sqrt{\alpha}\sigma_s = \sigma_{\text{eff}}(\sigma_s/\sigma_w)$ and using the assumption $\sigma_s/\sigma_w < 1$, the last term can be upper bounded as

$$\int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\Lambda_n\setminus\{\mathbf{0}\}+\mathbf{w})d\mathbf{w} \leq \int_{\mathcal{V}(\Lambda_n)} f_{\sigma_{\text{eff}}}(\Lambda_n\setminus\{\mathbf{0}\}+\mathbf{w})f_{\sqrt{\alpha}\sigma_s}(\mathbf{w})d\mathbf{w} \leq P_e(\Lambda_n)f_{\sigma_s}(\Lambda_n).$$

(15)

Combining altogether, we obtain the bound

$$f_{\sigma_s}(\Lambda_n) \leq 2P_e(\Lambda_n)f_{\sigma_s}(\Lambda_n) + \frac{1}{(\sqrt{2\pi}\sigma_s)^n} \Rightarrow P_e(\Lambda_n) \geq \frac{1}{2}(1 - P_0(\Lambda_n)).$$

This implies in turn that, if the probability of error $P_e(\Lambda_n) \to 0$, then $P_0(\Lambda_n) \to 1$ and, from Lemma 8, $(1/n)\mathbb{H}(\mathbf{X}_n) \to 0$. Conversely, if we force $(1/n)\mathbb{H}(\mathbf{X}_n)$ to tend to a positive value, $P_0(\Lambda_n)$ is bounded away from one, and therefore $P_e(\Lambda_n)$ is bounded away from zero. $\qquad\square$

VI. CONCLUSION

In this paper we have shown that dithered probabilistic shaping (DPS) can convert any lattice which is good for the unconstrained AWGN channel into a good code in the power-constrained setting. For instance any sphere-bound achieving lattice in the sense of [8] can be coupled with our results in order to produce capacity-achieving lattices. We stress the fact that previous schemes in the literature strictly need extra conditions other than AWGN-goodness, such as flatness-goodness or quantization-goodness.

Efficient explicit implementations of DPS for a broad range of dimensions and rates are an interesting further research direction. Improving the second-order analysis in order to achieve the right dispersion for all SNRs is a further point of interest and left as an open problem.

Finally, although the heuristic argument exhibited in Section V reveals that the centered distribution should fail to achieve capacity for SNR $< e - 1$, the actual proof only holds for SNR < 1, leaving inconclusive the values SNR $\in [1, e - 1]$. Furthermore, the scheme [12] fixes the variance parameter $\sigma_s^2 = P$ a priori, whereas one could potentially achieve better rates by choosing σ_s adaptively depending on the dimension. A stronger converse that can handle varying σ_s would strengthen our results. We believe that this would require completely new arguments.

ACKNOWLEDGEMENTS

APPENDIX

In this appendix we prove Lemma 7. Recall the MMSE coefficient α and the effective noise parameter σ_{eff}^2:

$$\alpha = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2}, \quad \sigma_{\text{eff}}^2 = \sigma_w^2\frac{\sigma_s^2}{\sigma_w^2 + \sigma_s^2} = \alpha\sigma_w^2.$$

Let $\tilde{\mathbf{X}} = (\alpha - 1)\mathbf{X}$, where $\mathbf{X} \sim D_{\Lambda, \sigma_s^2}$, We have that $\tilde{\mathbf{X}} \sim D_{(\alpha-1)\Lambda, (\alpha-1)^2 \sigma_s^2}$. Indeed, the probability of picking $\tilde{\mathbf{x}} \in (\alpha - 1)\Lambda$ is given by

$$\frac{f_{\sigma_s \sqrt{\alpha-1}}(\tilde{\mathbf{x}})}{f_{\sigma_s \sqrt{\alpha-1}}((\alpha-1)\Lambda)} = \frac{f_{\sigma_s \sqrt{\alpha-1}}(\tilde{\mathbf{x}})}{f_{\sigma_s}(\Lambda)}$$

Let $\tilde{\mathbf{W}} = \alpha \mathbf{W} \sim \mathcal{N}(0, \alpha^2 \sigma_s^2)$, and $\mathbf{W}_{\text{eff}} = \tilde{\mathbf{X}} + \tilde{\mathbf{W}}$. The distribution of the continuous variable $\mathbf{W}_{\text{eff}}$ is given by the convolution of the distributions of $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{W}}$, namely

$$
\begin{aligned}
g(\mathbf{w}_{\text{eff}}) &= \sum_{\tilde{\mathbf{x}} \in (\alpha-1)\Lambda} \frac{f_{\sigma_s \sqrt{\alpha-1}}(\tilde{\mathbf{x}})}{f_{\sigma_s}(\Lambda)} f_{\sigma_w \sqrt{\alpha}}(\mathbf{w}_{\text{eff}} - \tilde{\mathbf{x}}) \\
&= \frac{1}{(\sqrt{2\pi\sigma_w^2 \alpha^2})^n} \times \frac{1}{(\sqrt{2\pi\sigma_s^2})^n f_{\sigma_s}(\Lambda)} \times \sum_{\tilde{\mathbf{x}} \in \Lambda} e^{-\|\tilde{\mathbf{x}}\|^2/2\sigma_s^2 - \|\mathbf{w}_{\text{eff}} - (\alpha-1)\tilde{\mathbf{x}}\|^2/2\alpha^2 \sigma_w^2}. \\
&= \frac{1}{(\sqrt{2\pi\sigma_{\text{eff}}^2})^n} \times \frac{1}{(\sqrt{2\pi\alpha\sigma_s^2})^n f_{\sigma_s}(\Lambda)} \times \sum_{\tilde{\mathbf{x}} \in \Lambda} e^{-\|\tilde{\mathbf{x}}\|^2/2\sigma_s^2 - \|\mathbf{w}_{\text{eff}} - (\alpha-1)\tilde{\mathbf{x}}\|^2/2\alpha^2 \sigma_w^2}
\end{aligned}
\tag{16}
$$

To evaluate the exponents in the last expression we use the identity

$$\frac{\|\tilde{\mathbf{x}}\|^2}{2\sigma_s^2} + \frac{\|\mathbf{w}_{\text{eff}} - (\alpha-1)\tilde{\mathbf{x}}\|^2}{2\alpha^2 \sigma_w^2} = \frac{\|\mathbf{w}_{\text{eff}}\|^2}{2\sigma_{\text{eff}}^2} + \frac{\|\tilde{\mathbf{x}} + \mathbf{w}_{\text{eff}}\|^2}{2\alpha\sigma_s^2}$$

which gives

$$g(\mathbf{w}_{\text{eff}}) = \frac{e^{-\frac{\|\mathbf{w}_{\text{eff}}\|^2}{2\sigma_{\text{eff}}^2}}}{(\sqrt{2\pi\sigma_{\text{eff}}^2})^n} \times \frac{1}{(\sqrt{2\pi\alpha\sigma_s^2})^n f_{\sigma_s}(\Lambda)} \times \sum_{\tilde{\mathbf{x}} \in \Lambda + \mathbf{w}_{\text{eff}}} e^{-\frac{\|\tilde{\mathbf{x}}\|^2}{2\alpha\sigma_s^2}}. \tag{17}$$

This last expression for $g(\mathbf{w}_{\text{eff}})$ coincides with the one in Lemma 7.

REFERENCES

[1] W. Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(1):625–635, 1993.

[2] G. Böcherer. Probabilistic signal shaping for bit-metric decoding. In *IEEE International Symposium on Information Theory*, pages 431–435, June 2014.

[3] A. Campello, C. Ling, and J. C. Belfiore. Algebraic lattice codes achieve the capacity of the compound block-fading channel. In *IEEE International Symposium on Information Theory*, pages 910–914, July 2016.

[4] K.-M. Chung, D. Dadush, F.-H. Liu, and C. Peikert. On the lattice smoothing parameter problem. In *IEEE Conference on Computational Complexity (CCC)*, pages 230–241. IEEE Computer Soc., Los Alamitos, CA, 2013.

[5] D. Dadush, O. Regev, and N. Stephens-Davidowitz. On the closest vector problem with a distance guarantee. In *IEEE 29th Conference on Computational Complexity (CCC)*, pages 98–109, June 2014.

[6] U. Erez and R. Zamir. Achieving $1/2 \log(1 + \text{SNR})$ on the AWGN channel with lattice encoding and decoding. *IEEE Transactions on Information Theory*, 50(10):2293–2314, Oct 2004.

[7] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik. On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel. *Journal of Lightwave Technology*, 34(21):5063–5073, Nov 2016.

[8] G. D. Forney, M. D. Trott, and S.-Y. Chung. Sphere-bound-achieving coset codes and multilevel coset codes. *IEEE Transactions on Information Theory*, 46(3):820–850, May 2000.

[9] C. Gentry, C. Peikert, and V. Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *STOC*, pages 197–206, 2008.

[10] A. Ingber, R. Zamir, and M. Feder. Finite-dimensional infinite constellations. *IEEE Transactions on Information Theory*, 59(3):1630–1656, March 2013.

[11] F. R. Kschischang and S. Pasupathy. Optimal nonuniform signaling for gaussian channels. *IEEE Transactions on Information Theory*, 39(3):913–929, May 1993.

[12] C. Ling and J.-C. Belfiore. Achieiving AWGN channel capacity with lattice Gaussian coding. *IEEE Transactions on Information Theory*, 60(10):5918–5929, Oct. 2014.

[13] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé. Semantically secure lattice codes for the Gaussian wiretap channel. *IEEE Trans. Inform. Theory*, 60(10):6399–6416, Oct. 2014.

[14] L. Luzzi, C. Ling, and R. Vehkalahti. Almost universal codes for fading wiretap channels. In *IEEE International Symposium on Information Theory*, pages 3082–3086, July 2016.

[15] D. Micciancio and O. Regev. Worst-case to average-case reductions based on Gaussian measures. *SIAM J. Comput.*, 37(1):267–302 (electronic), 2007.

[16] N. Palgy and R. Zamir. Dithered probabilistic shaping. In *IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, pages 1–5, Nov 2012.

[17] G. Poltyrev. On coding without restrictions for the AWGN channel. *IEEE Transactions on Information Theory*, 40(2):409–417, Mar 1994.

[18] Y. Polyanskiy, H. V. Poor, and S. Verdu. Dispersion of Gaussian channels. In *IEEE International Symposium on Information Theory*, pages 2204–2208, June 2009.

[19] O. Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):Art. 34, 40, 2009.

[20] O. Regev and D. Dadush. Towards strong reverse minkowski-type inequalities for lattices. In *IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, 2016.

[21] O. Regev and N. Stephens-Davidowitz. A Reverse Minkowski Theorem. In *STOC*, 2017.

[22] B. P. Smith and F. R. Kschischang. A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications. *Journal of Lightwave Technology*, 30(13):2047–2053, July 2012.

[23] L. Wang, G. W. Wornell, and L. Zheng. Fundamental limits of communication with low probability of detection. *IEEE Transactions on Information Theory*, 62(6):3493–3503, June 2016.

[24] Y. Yan, L. Liu, C. Ling, and X. Wu. Construction of capacity-achieving lattice codes: Polar lattices. *CoRR*, abs/1411.0187, 2014.

[25] Y. Yona and E. Haim. The Effect of Dither on Modulo-Lattice Schemes in the Low SNR Regime. *unpublished manuscript*, 2017.

[26] R. Zamir. *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.