Lattices which are Good for (Almost) Everything

Uri Erez, Simon Litsyn* and Ram Zamir†
Dept. of Elect. Eng. - Systems, Tel Aviv University, ISRAEL

June 2, 2003

Abstract

We define an ensemble of lattices, and show that for asymptotically high dimension most of its members are simultaneously good as sphere packings, sphere coverings, AWGN channel codes and MSE quantization codes. These lattices are generated by applying Construction A to a random linear code over a prime field of growing size, i.e., by "lifting" the code to $\mathbb{R}^n$.

Keywords: sphere packing, sphere covering, MSE quantization, coding for unconstrained AWGN channel, lattice codes, Minkowski bound, Poltyrev exponent.

1 Introduction

In this work we consider the problem of existence of lattices in Euclidean space that are simultaneously asymptotically good in several coding related contexts. We begin with a description of the binary counterpart which motivated our study.

In n-dimensional *binary* Hamming space the problems of sphere packing, sphere covering, channel coding, and quantization are well known. The first two are of a combinatorial nature. In the packing problem we are interested in packing the greatest possible number of non-intersecting Hamming spheres of a given radius. This is equivalent to maximizing the number of codewords for a given minimum distance of the code. For a comprehensive survey see e.g.,

*The work of this author was supported in part by the ISF and the Binational Science Foundation, grant BSF1999-099.

†The work of this author was supported in part by the Israel Academy of Science research fund, grant ISF 65-01.

[25, 28]. The covering problem asks for a minimum size collection of spheres of a given radius, such that every point of space belongs to at least one sphere. This corresponds to covering codes, see e.g., [6]. The other two problems are of a probabilistic or information theoretical nature. The channel coding problem asks for the best arrangement of points in Hamming space such that for a given number of codewords and noise statistics, the probability of error of the maximum likelihood decoder is minimized. Of particular importance is the question of the greatest possible rate of the code that still enables this error probability to vanish asymptotically as the dimension n (length of the code) goes to infinity. This leads to the concept of channel capacity and error exponent. For a survey see e.g., [16]. The quantization problem seeks to minimize the required number of codewords such that the average distance of the points in Hamming space from their nearest codeword is not greater than a specified target distortion, see e.g., [2]. In what follows we deal with the asymptotic case of high dimension n. The bounds on the parameters are usually exponential in the dimension. For our purposes all bounds with the same exponent are considered equivalent.

All these problems are typically treated in information theory by *random coding* arguments. Basically all that is needed is to draw at random n-tuples according to a uniform distribution. The best known bounds, asymptotic in n, are obtained in this manner. For the packing problem it is the Gilbert bound [19], for the covering problem it is the Goblick-Cohen bound [20, 7]. For the channel coding problem the bound is given by the Shannon capacity [33] and the Elias-Gallager exponent [12]. For the quantization problem the bound is the Shannon rate-distortion bound [2].

The obtained codes however lack structure. It turns out that even if we impose a requirement that the code be *linear*, it is still possible to achieve the same asymptotic bounds. Indeed, for the packing problem this was shown by Varshamov [37], for the covering problem (from which the claim for quantization also follows) by Cohen [7], and for the channel coding problem by Gallager and Dobrushin [15, 11].

Moreover, the mentioned bounds hold true in a stronger sense, namely, for almost all linear codes. This was shown for the packing problem by Pierce [27], for the covering problem (from which the claim for quantization also follows) by Blinovskiĭ [3] (for the nonlinear case see [10]), and for the channel coding problem by Gallager and Dobrushin [15, 11]. In fact, as a code for the BSC channel the random linear ensemble is in a sense better than the totally random ensemble for channel coding. A typical code of the linear ensemble achieves the expurgated bound. This is not the case for a typical code of the totally random ensemble where expurgation of a relatively small number of codewords is necessary, see e.g., [1, 4].

A random linear ensemble is produced by a randomly chosen generator matrix with independent equiprobable binary entries. Since the same ensemble of linear codes generates a solution to all four problems, it follows (by a union bound argument) that *there exist binary linear codes that are good in all four senses simultaneously.*

In this paper we prove a similar result for analogous problems in Euclidean space, namely,

that *there exist lattices that are simultaneously good for sphere packing and covering, channel coding and quantization.* While the sphere packing/covering problems carry over in a straightforward manner from the Hamming to the Euclidean space, the corresponding extensions of the transmission/quantization problems require a word of caution. The standard scenarios of channel/source coding in Euclidean space are based on *bounded* codebooks [33]. In channel coding this is due to a transmitter power constraint, while in source coding this is the result of a fixed (or a limited) codeword length assumption. Nevertheless, for our discussion of lattice codes it is more convenient to assume *unbounded* codebooks. Poltyrev's notion of "unconstrained channels" provides a meaningful definition for the capacity of such codebooks in the channel coding context [29] (see also [14]), while "entropy constrained quantization" provides a corresponding notion of coding rate in the quantization framework [17, 42, 40]. Similarly to the binary Hamming case, we address here only asymptotic, in the space dimension, bounds. The bounds are again exponential in the dimension and are the best known bounds also for non lattice constellations (though sub-exponential terms may vary). We show that asymptotically there are lattices that achieve these bounds simultaneously.

For the Euclidean packing problem the best known bound is the Minkowski bound [26], for the covering problem it is the Rogers bound [30, 31], for the channel coding and Mean Squared Error (MSE) quantization problems the bounds are due to Poltyrev [29, 39] (which are related to Shannon's bounds for the power constrained channel and rate distortion of Gaussian sources, see [34, 33]). We say that a sequence of lattices is good for packing if it asymptotically achieves the Minkowski bound. A similar terminology is used when the Rogers and Poltyrev bounds are used.

It is well known that a lattice that is good according to *any* of the four criteria has a Voronoi region (see definition below) that in some sense is "close" to spherical. However, for a fixed dimension, an optimal lattice in one sense need not necessarily be optimal in another. For example, in three dimensional Euclidean space $\mathbb{R}^3$, the optimal lattice for packing is the FCC lattice while the optimal lattice for covering is the BCC lattice, see, e.g., [8]. Nonetheless, in this work we prove that asymptotically (in dimension) a lattice may be optimal in all four senses.

In order to prove that there exists a sequence of lattices simultaneously satisfying the four bounds, we adopt the approach used in the binary case. Namely, we attempt to define a *random ensemble* of lattices such that for almost all its members, the bounds are achieved. Moreover, we try to find a small and simply described ensemble.

Loeliger has previously taken this approach of defining an ensemble of lattices in [24]. Using this technique he proved that lattices achieve the Minkowski bound as well as achieve the Poltyrev *capacity* of the AWGN channel. We extend here the analysis to error exponents as well as treat also the covering and quantization problems.

In [5] Butler proved that there indeed exist lattices that are simultaneously good for

packing and covering. His proof utilizes the techniques of Rogers [32]. In the current work, we re-derive this claim as well as extend it to the channel coding and quantization settings. Moreover, our proof is significantly simpler.

The paper is organized as follows. Section 2 introduces the necessary notation and defines the four problems of interest. In Section 3 the random ensemble of lattices is introduced. Sections 4 to 7 each contains the proof that the defined ensemble is good for one of the four problems; simultaneous goodness is proved in Section 8. The work is summarized in Section 9. Technical proofs are relegated to the Appendix.

2 Lattice Properties

In order to give precise definitions of the problems we use the following notation:

- GRID : cubic grid (lattice) of step size $1/p$.
- Λ: n dimensional lattice nested in GRID, i.e., $\Lambda \subset$ GRID.
- $\|\cdot\|$: Euclidean norm, i.e., $\|\mathbf{x}\|^2 = \sum_{i=1}^n x_i^2$ [1].
- $A + B$: set sum. If A and B are sets in $\mathbb{R}^n$ then
$$A + B = \{x + y : x \in A, y \in B\}$$
- $\mathcal{V}$: Voronoi region of Λ, i.e,
$$\mathcal{V} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{c}\| \ \forall \ \mathbf{c} \in \Lambda\} \tag{1}$$
and
$$\mathbb{R}^n = \Lambda + \mathcal{V}. \tag{2}$$
- Vol$(\cdot)$: volume of a closed set in $\mathbb{R}^n$.
- $(x+y)^* = x + y \mod \mathbb{Z} = (x+y) - \lfloor x+y \rfloor$ where $x, y \in \mathbb{R}$ and $\lfloor \cdot \rfloor$ means rounding to the nearest non greater integer.
- CUBE$=[0,1)^n$
- $\mathcal{A}^* = \mathcal{A} \mod$ CUBE, where $\mathcal{A}$ is any set in $\mathbb{R}^n$.
- GRID* = GRID $\cap$ CUBE.
- $r\mathcal{B}$: n dimensional closed ball of radius r centered at the origin, $1 \cdot \mathcal{B} = \mathcal{B} =$ unit ball.
- $V_{\mathcal{B}}(r)$: volume of ball of radius r, i.e., $V_{\mathcal{B}}(r) = \text{Vol}(r\mathcal{B}) = r^n \text{Vol}(\mathcal{B})$.
- $\lceil \cdot \rceil_p$: rounding to the nearest greater prime.

[1] Extension to other norms is considered in the discussion.

2.1 Packing problem

Consider a lattice Λ with Voronoi region $\mathcal{V}$. For a given radius r the set $\Lambda + r\mathcal{B}$ is a packing in Euclidean space if for all lattice points $\mathbf{x}, \mathbf{y} \in \Lambda$ $(\mathbf{x} \neq \mathbf{y})$ we have

$$(\mathbf{x} + r\mathcal{B}) \cap (\mathbf{y} + r\mathcal{B}) = \emptyset.$$

That is, the balls do not intersect. Define the packing radius r_Λ^{pack} of the lattice by

$$r_\Lambda^{\text{pack}} = \sup\{r : \ \Lambda + r\mathcal{B} \text{ is a packing}\}. \tag{3}$$

Denote by r_Λ^{effec} the "effective radius" of the Voronoi region, meaning the radius of a ball

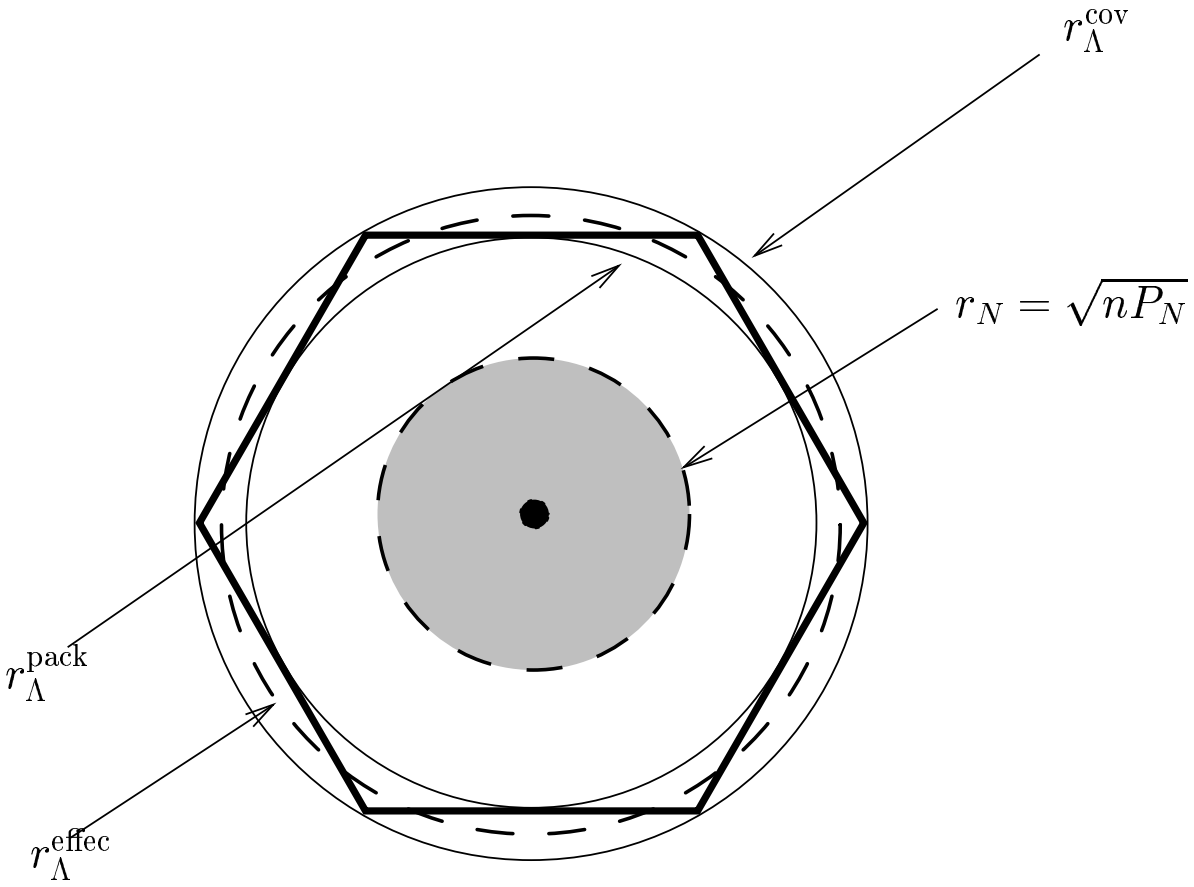

Figure 1: Geometric picture

having the same volume, so that r_Λ^{effec} is defined by

$$V_\mathcal{B}(r_\Lambda^{\text{effec}}) = \text{Vol}(\mathcal{V}) \tag{4}$$

or equivalently

$$r_{\Lambda}^{\text{effec}} = \left[\frac{\text{Vol}(\mathcal{V})}{V_{\mathcal{B}}(1)}\right]^{\frac{1}{n}} \tag{5}$$

where $V_{\mathcal{B}}(1)$ (the volume of a unit sphere) is given by (see e.g., [8])

$$V_{\mathcal{B}}(1) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} \approx \frac{1}{\sqrt{n\pi}}\left(\frac{n}{2e}\right)^{-n/2}. \tag{6}$$

Figure 1 gives the geometric picture of r_{pack} and $r_{\Lambda}^{\text{effec}}$ with respect to the Voronoi region, as well as the other radii to be defined below. Define the packing efficiency ρ_{pack} of a lattice Λ by

$$\rho_{\text{pack}}(\Lambda) = \frac{r_{\Lambda}^{\text{pack}}}{r_{\Lambda}^{\text{effec}}}. \tag{7}$$

We note that the packing efficiency $\rho_{\text{pack}}(\Lambda)$ is by definition not greater than one. We wish $\rho_{\text{pack}}(\Lambda)$ to be as large as possible. The density of the packing, i.e., the proportion of space taken up by the spheres is $(\rho_{\text{pack}}(\Lambda))^n$.

Let Λ_n denote an n-dimensional lattice. Define the optimal asymptotic packing efficiency by

$$\rho_{\text{pack}}^* = \limsup_n \sup_{\Lambda_n} \rho_{\text{pack}}(\Lambda_n).$$

The best known lower bound for ρ_{pack}^* is given by the Minkowski-Hlawka theorem [32]

$$\rho_{\text{pack}}^* \geq \frac{1}{2}. \tag{8}$$

Since $\rho_{\text{pack}}^* \leq 1$ from volume arguments, the bound implies a radius half as large as we could hope for. In fact, it is known that no efficient packing exists, in the sense that ρ_{pack}^* is strictly less than one. The best known upper bound was found by Kabatiansky and Levenshtein [21] and is given by (see [8])

$$\rho_{\text{pack}}^* \leq 0.6603. \tag{9}$$

Therefore, it is not known whether the Minkowski bound is tight. We say that a sequence of lattices is asymptotically good for packing if it achieves the Minkowski bound.

We note that for small dimension the optimal packing efficiency is a rather irregular (non monotonic) function of n. Indeed for $n = 1$ we have $\rho_{\text{pack}}(\Lambda_1) = 1$. Therefore packing is efficient in the one dimensional case but degrades as the dimension grows. A similar irregular behavior at small dimensions is also true in the covering problem but as we shall see, as $n \to \infty$ the optimal covering is efficient. On the other hand the performance of lattices as a function of the dimension n for the quantization problem and AWGN channel coding problem is more well behaved. It seems that the performance in these problems improves monotonically with dimension.

2.2 Covering problem

The associated notions for the covering problem are defined similarly to their packing counterparts. The set $\Lambda + r\mathcal{B}$ is a covering of Euclidean space if

$$\mathbb{R}^n \subseteq \Lambda + r\mathcal{B}.$$

That is, each point in space is covered by at least one ball. Define the covering radius of the lattice r_Λ^{cov} by

$$r_\Lambda^{\text{cov}} = \inf\{r : \ \Lambda + r\mathcal{B} \text{ is a covering}\}.$$

Define the covering efficiency $\rho_{\text{cov}}(\Lambda)$ of a lattice by

$$\rho_{\text{cov}}(\Lambda) = \frac{r_\Lambda^{\text{cov}}}{r_\Lambda^{\text{effec}}}.$$

We note that the covering efficiency $\rho_{\text{cov}}(\Lambda)$ is by definition not less than one. We wish $\rho_{\text{cov}}(\Lambda)$ to be as small as possible. The density of the covering, i.e., the average number of balls covering a point is $(\rho_{\text{cov}}(\Lambda))^n$.

Define the optimal asymptotic covering efficiency by

$$\rho_{\text{cov}}^* = \liminf_n \inf_\Lambda \rho_{\text{cov}}(\Lambda_n).$$

It is a result of Rogers [31] that ρ_{cov}^* satisfies

$$\rho_{\text{cov}}^* = 1.$$

This means that covering (in contrast to packing) may be asymptotically efficient, i.e., every point in space can be covered by at most a sub-exponential number of balls. See standard textbooks on packing and covering such as Rogers [32] and Conway and Sloane [8].

It will be shown later (see also [5]) that the following is also true:

$$\limsup_{n\to\infty} \sup_{\Lambda_n} \frac{\rho_{\text{pack}}(\Lambda_n)}{\rho_{\text{cov}}(\Lambda_n)} \geq \frac{1}{2},$$

i.e., that simultaneously good packings and coverings exist, as previously shown by Butler [5]. We say that a sequence of lattices is good for covering if it satisfies Rogers' bound, i.e., if it allows for efficient covering.

2.3 MSE quantization

In MSE quantization we associate to Λ a nearest neighbor quantizer $Q_\Lambda(\cdot)$ such that

$$Q(\mathbf{x}) = \mathbf{y}, \ \mathbf{y} \in \Lambda, \quad \text{if } \|\mathbf{x}-\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}'\| \ \forall\, \mathbf{y}' \in \Lambda \tag{10}$$

where ties are broken in a systematic manner. Equivalently

$$Q(\mathbf{x}) = \mathbf{y},\ \mathbf{y} \in \Lambda, \quad \text{if } \mathbf{x} \in \mathbf{y} + \mathcal{V}. \tag{11}$$

The second moment, $\sigma^2 = \sigma^2(\Lambda)$, of Λ is defined as the second moment per dimension of a uniform distribution over $\mathcal{V}$,

$$\sigma^2 = \frac{1}{\mathrm{Vol}(\mathcal{V})} \cdot \frac{1}{n} \int_{\mathcal{V}} \|\mathbf{x}\|^2 d\mathbf{x}. \tag{12}$$

A figure of merit of a lattice quantizer with respect to MSE distortion measure is the normalized second moment

$$G(\Lambda) = \frac{1}{\mathrm{Vol}(\mathcal{V})^{1+2/n}} \cdot \frac{1}{n} \int_{\mathcal{V}} \|\mathbf{x}\|^2 d\mathbf{x} = \frac{\sigma^2}{\mathrm{Vol}(\mathcal{V})^{2/n}}. \tag{13}$$

The minimum possible value of $G(\Lambda_n)$ over all lattices in $\mathbb{R}^n$ is denoted G_n. The normalized second moment of a sphere, denoted by G_n^*, approaches $\frac{1}{2\pi e}$ as the dimension n goes to infinity. The isoperimetric inequality implies that $G_n > G_n^* > \frac{1}{2\pi e}$ for all n. We also have $G_n \leq G_1 = \frac{1}{12}$.

The operational significance of this figure of merit comes from a result due to Gersho [17] (see also [38]) that states that under quite general conditions for high variable rate lattice quantization of a source $\mathbf{X}$ the distortion satisfies

$$D \approx G(\Lambda) e^{2(h_n(\mathbf{X}) - R)} \tag{14}$$

where D denotes the distortion, R is the quantization rate in nats and $h_n(\mathbf{X}) = \frac{1}{n} h(\mathbf{X})$ is the dimension-normalized differential entropy of the source (also in nats).

From rate distortion theory we have that the ultimate performance of a random code, defined by the Shannon lower bound, would yield the same formula for the distortion with $G(\Lambda)$ replaced by $\frac{1}{2\pi e}$, see [23]. Thus we refer to a sequence of lattices as good for MSE quantization if $G(\Lambda_n)$ tends to $\frac{1}{2\pi e}$. A result of Zamir, Feder and Poltyrev [39] states that

$$\lim_n G_n = \frac{1}{2\pi e} \tag{15}$$

i.e., that there exist good lattice quantizers.

We note that unlike in the AWGN channel coding problem, here it is not interesting to investigate the rate distortion error exponent for rates $R > R(D)$. As we shall see below, in the AWGN problem, when working below capacity $R < C$, the error probability decays exponentially with n. In rate distortion theory a similar analysis is done. For rates above the minimum possible to achieve a specified distortion D (see (14)), one may analyze probability that a point exceeds distortion D. However, as a consequence of the fact that we have a good covering lattice, this probability may be made zero for sufficiently large dimension, corresponding to an *infinite* exponent. Thus, we observe that the quantization problem differs greatly from the "dual" AWGN channel coding problem.

2.4 Coding for the unrestricted AWGN

Of the four applications of lattices considered in this work the precise notion of good lattices for coding over the AWGN channel was the most recently introduced. It appears in a 1994 paper by Poltyrev [29].

The AWGN channel model is given by the input/output relation

$$Y = X + N \tag{16}$$

where N is i.i.d. Gaussian noise of variance P_N . We define the "effective radius" of the noise vector by

$$r_N = \sqrt{nP_N}. \tag{17}$$

Note that by the Law of Large Numbers $\frac{1}{n}\|\mathbf{N}\|^2 \to P_N$ as $n \to \infty$, so $\|N\| \approx r_N$. Traditionally the scenario is such that the transmitter is subject to a power constraint, i.e., the input must satisfy $\frac{1}{n}\sum_i x^2 \leq P_X$. This means that only a finite subset of the lattice points is used as a codebook, typically the intersection of the lattice with a ball (or a thin spherical shell) of radius $\sqrt{nP_X}$. This approach originated in the work of De Buda [9] and was subsequently refined in [22, 24, 36]. When such an approach is taken, the Maximum Likelihood (ML) decoding regions are not the Voronoi regions of the lattice due to the boundedness of the codebook. Moreover, the decoding regions are not identical (up to translation), and in fact some are not bounded. This breaks the symmetry of the lattice structure in the decoding process and makes the definition of goodness of a lattice highly dependent on the precise mode of transmission.

An alternative approach was taken by Poltyrev [29]. He considered the problem of coding for the *unconstrained* AWGN channel. In this scenario any point of a lattice may be transmitted, corresponding to infinite power and transmission rate. For a given lattice the ML decoder will search for the lattice point that is nearest to the received vector. Therefore the probability of decoding error is the probability that the noise leave the Voronoi region of the transmitted lattice point

$$P_e(\Lambda, r_N) = \Pr\{\mathbf{N} \notin \mathcal{V}\}. \tag{18}$$

Since the rate of transmission is infinite, performance is measured with respect to the ratio of the radius of the Voronoi region and the "radius" of the noise. More precisely, define

$$\rho_{\text{AWGN}}(\Lambda, r_N) = \frac{r_\Lambda^{\text{effec}}}{r_N}. \tag{19}$$

Using the relation $G_n^* = \frac{1}{(n+2)[V_{\mathcal{B}}(1)]^{2/n}}$ (see e.g., [39]) we have $r_\Lambda^{\text{effec}} = \sqrt{(n+2)G_n^*}\,[\text{Vol}(\mathcal{V})]^{1/n}$. Recalling that $\lim_{n\to\infty} G_n^* = \frac{1}{2\pi e}$ [8] we therefore have

$$\rho_{\text{AWGN}}(\Lambda, r_N) = \frac{r_\Lambda^{\text{effec}}}{r_N} = \frac{[\text{Vol}(\mathcal{V})]^{1/n}}{\sqrt{2\pi e P_N}} + o(1) \tag{20}$$

where $o(1) \to 0$ as $n \to \infty$. We ask for a lattice with minimal probability of error for a given ratio $\rho_{\mathrm{AWGN}}(\Lambda, r_N)$. Poltyrev showed that reliable transmission is possible for $\rho_{\mathrm{AWGN}}(\Lambda, r_N) > 1$ and that the probability of error may be exponentially bounded, just as is done for the constrained AWGN channel [16], by the random coding bound and expurgated bound. He showed that the random coding error exponent for the unconstrained AWGN channel is given by

$$E_P^r(\rho_{\mathrm{AWGN}}) = \frac{1}{2}\left[(\rho_{\mathrm{AWGN}}^2 - 1) - \ln \rho_{\mathrm{AWGN}}^2\right] \tag{21}$$

for $1 \le \rho_{\mathrm{AWGN}} \le 2$, and the expurgated exponent is given by

$$E_P^{ex}(\rho_{\mathrm{AWGN}}) = \rho_{\mathrm{AWGN}}^2/8 \tag{22}$$

for $\rho_{\mathrm{AWGN}}^2 \ge 4$. Connecting $E_p^r(\rho_{\mathrm{AWGN}})$ and $E_P^{ex}(\rho_{\mathrm{AWGN}})$ by a straight line the Poltyrev exponent is given by

$$E_P(\rho_{\mathrm{AWGN}}) = \begin{cases} \frac{1}{2}\left[(\rho_{\mathrm{AWGN}}^2 - 1) - \ln \rho_{\mathrm{AWGN}}^2\right], & 1 \le \rho_{\mathrm{AWGN}}^2 \le 2 \\ \frac{1}{2}\ln\frac{e\rho_{\mathrm{AWGN}}^2}{4}, & 2 \le \rho_{\mathrm{AWGN}} \le 4 \\ \frac{\rho_{\mathrm{AWGN}}^2}{8}, & \rho_{\mathrm{AWGN}}^2 \ge 4 \end{cases} \tag{23}$$

More specifically, there exist lattices for which the probability of error satisfies the following exponential bound

$$P_e(\Lambda, r_N) < e^{-n(E_P(\rho_{\mathrm{AWGN}}) - o(1))}.$$

We note that this bound is exponentially tight in the random coding part. Note also that $E_P(\rho_{\mathrm{AWGN}})$ vanishes at $\rho_{\mathrm{AWGN}} = 1$ meaning that asymptotic reliable communication is possible as long as $\rho_{\mathrm{AWGN}} > 1$ (and not below it), so that $\rho_{\mathrm{AWGN}} = 1$ or $r_\Lambda^{\mathrm{effec}} = r_N$ has the significance of capacity in this scenario. As in the other problems, the performance of general non lattice codes is not superior to lattice codes (w.r.t. to the bounds) [29].

In [14] Poltyrev's unconstrained coding notion was linked to that of coding for the power constrained AWGN channel (via the notions of modulo-lattice transformation and Voronoi constellations), thereby giving it an operational significance.

Here, we re-derive Poltyrev's error exponent in a simplified way and show that the lattice ensemble is good for AWGN coding in the sense that, for any given ratio ρ_{AWGN}, it achieves the Poltyrev error exponent. In the next section we introduce the ensemble of lattices used. A nice property we obtain is that a lattice drawn from the ensemble attains the expurgated bound with high probability. This is in contrast to a totally random code for which expurgation is necessary. In a sense the expurgation is performed here on codebooks instead of codewords. This is similar to the behavior of random binary linear codes, as opposed to totally random binary codes [1].

3 A Random Ensemble of Lattices

Let k, n and p be integers such that $k \leq n$ and let G be a $k \times n$ (generating) matrix with elements in $\mathbb{Z}_p = \{0, \ldots, p-1\}$. We do not assume that G is necessarily full rank. The generation of an n-dimensional lattice by Construction A consists of the following steps:

1) Define the discrete codebook, $\mathcal{C} = \{\mathbf{x} = \mathbf{y} \cdot G : \mathbf{y} \in \mathbb{Z}_p^k\}$, where all the operations are over $\mathbb{Z}_p$ (that is, modulo-p). Thus $\mathbf{x} \in \mathbb{Z}_p^n$.

2) Map the code $\mathcal{C}$ into the unit cube by dividing all the components by p. Thus we define the finite set $\Lambda^* = p^{-1} \cdot \mathcal{C} \subseteq \text{GRID}$.

3) Replicate Λ^* over the entire Euclidean space $\mathbb{R}^n$ by integer tessellations, to form the lattice $\Lambda = \Lambda^* + \mathbb{Z}^n$. It is well known [8] (and elementary to prove) that Λ is indeed a lattice.

Example:
Set $n = 2$, $k = 1$ and $p = 11$. The underlying code is given by the generating matrix $G = [2, 3]$ so that

$$\mathcal{C} = \{x \cdot [2, 3] \mod 11 \ : \ x \in \mathbb{Z}_{11}\}$$

We embed the code "as is" in Euclidean space as depicted in Figure 2. Using this code we tessellate the whole of $\mathbb{R}^2$ resulting in the lattice

$$p\Lambda = \mathcal{C} + 11\mathbb{Z}^2.$$

This is depicted in Fig 3. The lower left quadrant corresponds to the $p \cdot \text{CUBE}$ region.

We note that $\mathbf{x}$ in Step 2 runs through $p^k = M$ vectors (not necessarily distinct, as G might not be full rank), M being the number of codewords. Let us index them as $\mathbf{y}_i$, $i = 0, \ldots, p^k - 1$. We assume that $\mathbf{y}_0 = 0^n = \mathbf{0}$, other than that the ordering is arbitrary. We correspondingly index the $M \stackrel{\Delta}{=} p^k$ points of Λ^* so that

$$\Lambda_i^* = p^{-1} \cdot \mathbf{y}_i \cdot G \quad , \quad i = 0, 1, \ldots, M-1.$$

The random ensemble of lattices we consider is generated as follows.

- Take p to be *prime*.
- Draw a $k \times n$ generator matrix G according to a uniform i.i.d. distribution over $\mathbb{Z}_p$, $G_{i,j} \sim \text{Unif}(0, \ldots, p-1)$, $\quad i = 1, \ldots, k \ ; \ j = 1, \ldots, n$.
- Apply Construction A, i.e., steps 1-3 as described above, to obtain the lattice Λ.

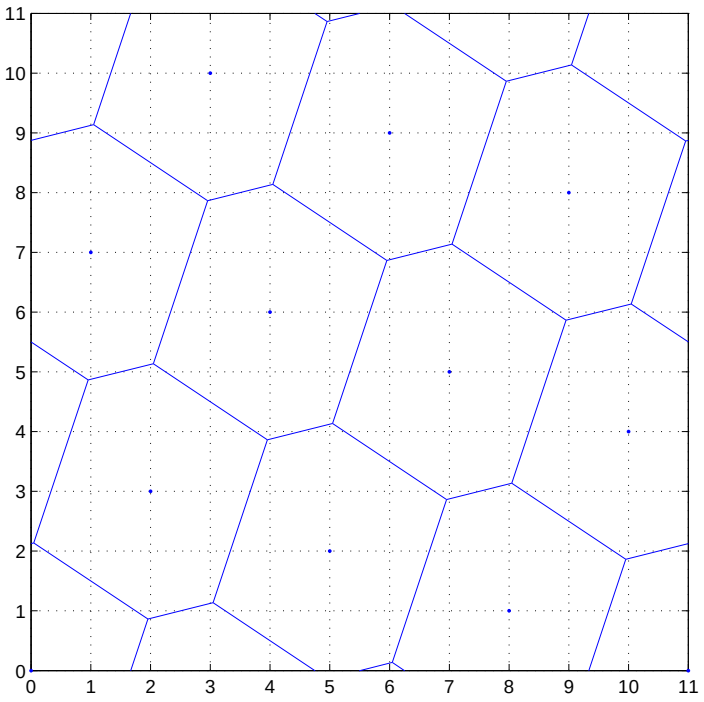

Figure 2: “Finite” lattice

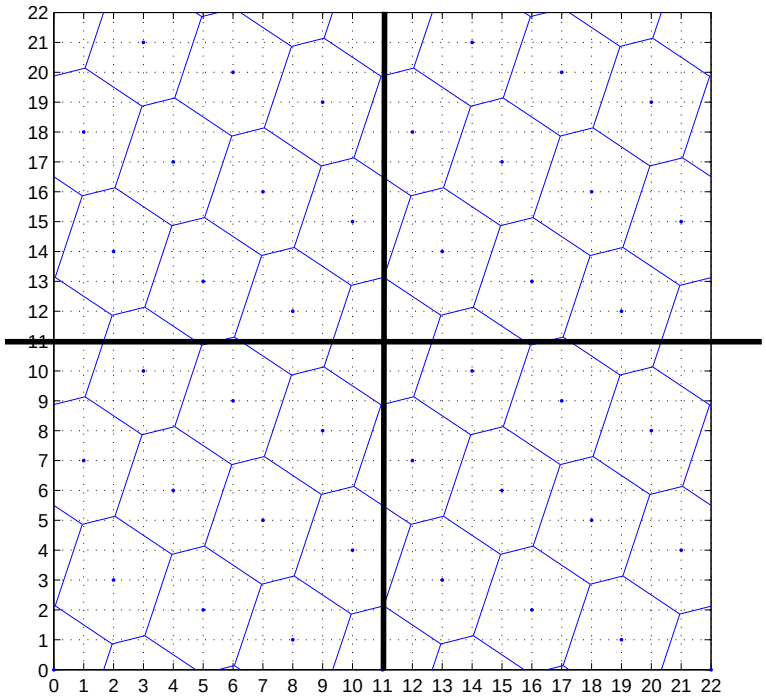

Figure 3: Forming a lattice by Construction A - tessellating space

The random ensemble is determined by n, k and p and in the sequel we refer to it as an (n, k, p) lattice ensemble. This ensemble was considered by Loeliger in [24] and has the following important properties.

1) $\Lambda_0^* = \mathbf{0}$ deterministically.

2) Λ_i^* is distributed uniformly over GRID* for $i = 1, \ldots, M-1$.

3) The difference $\Lambda_i^* - \Lambda_j^*$ is uniformly distributed over GRID* for all $i \neq j$.

The last two properties hold since p is prime (otherwise Λ_i^* might be restricted to some *sub*-grid).

By construction, the lattices we consider are periodic modulo the unit cube. All the problems we consider may thus be restated in equivalent terms in the realm of Λ^* in CUBE. Let r, $r < \frac{1}{2}$, designate the radius of a ball. We call $(r\mathcal{B})^*$ a (radius r) mod-sphere. We say that $(\Lambda^* + r\mathcal{B})^*$ is a mod-packing if

$$(\mathbf{x} + r\mathcal{B})^* \cap (\mathbf{y} + r\mathcal{B})^* = \emptyset. \quad \forall x, y \in \Lambda^* \quad x \neq y$$

That is, the mod-spheres do not intersect. Similarly we say that $(\Lambda^* + r\mathcal{B})^*$ is a mod-covering if CUBE $\subseteq (\Lambda^* + r\mathcal{B})^*$. It is easy to see that the condition $r < \frac{1}{2}$ ensures that $\Lambda + r\mathcal{B}$ is a packing (covering) iff $(\Lambda^* + r\mathcal{B})^*$ is a mod-packing (mod-covering). Therefore we use the two viewpoints interchangeably in the proofs below. Figure 4 demonstrates the relation between the mod-packing and the packing in $\mathbb{R}^n$ for $p = 3$ and $n = 2$. The lower left quadrant is CUBE and the three black points are Λ^*. The four disjoint full quarter circles form the set $(r\mathcal{B})^*$.

Note that if G is non-singular then there are $M = p^k$ codewords in GRID and the volume of a Voronoi region is p^{-k}. The probability that G is indeed non singular goes to one as $n \to \infty$. To see this, denote the rows of G by G_i, $i = 1, \ldots, k$. Consider any specific non trivial linear combination $\mathbf{x} = \sum_{i=1}^{k} c_i \cdot G_i$ determined by the coefficient vector $\mathbf{c} \neq \mathbf{0}$. For each component $j = 1, \ldots, n$ we have $\Pr\{\mathbf{x}_j = 0\} = p^{-1}$. Since the columns are statistically independent we get $\Pr\{\mathbf{x} = \mathbf{0}\} = p^{-n}$. Thus, applying the union bound over all possible linear combinations, i.e., all coefficient vectors $\mathbf{c}$, we have

$$\Pr\Big\{\text{rank}(G) < k\Big\} \leq \sum_{\mathbf{c} \neq \mathbf{0}} \Pr\left\{\sum_{i=1}^{k} c_i G_i = \mathbf{0}\right\} = p^{-(n-k)-1}. \tag{24}$$

We restrict our attention to ensembles such that $k < \beta n$ for some $0 < \beta < 1$ so that the latter probability goes to zero at least (p may also grow with n) exponentially. In the sequel we assume that G is full rank. If it is not, we treat that as a failure, adding a vanishing term to the probability that a lattice from the ensemble is not good.

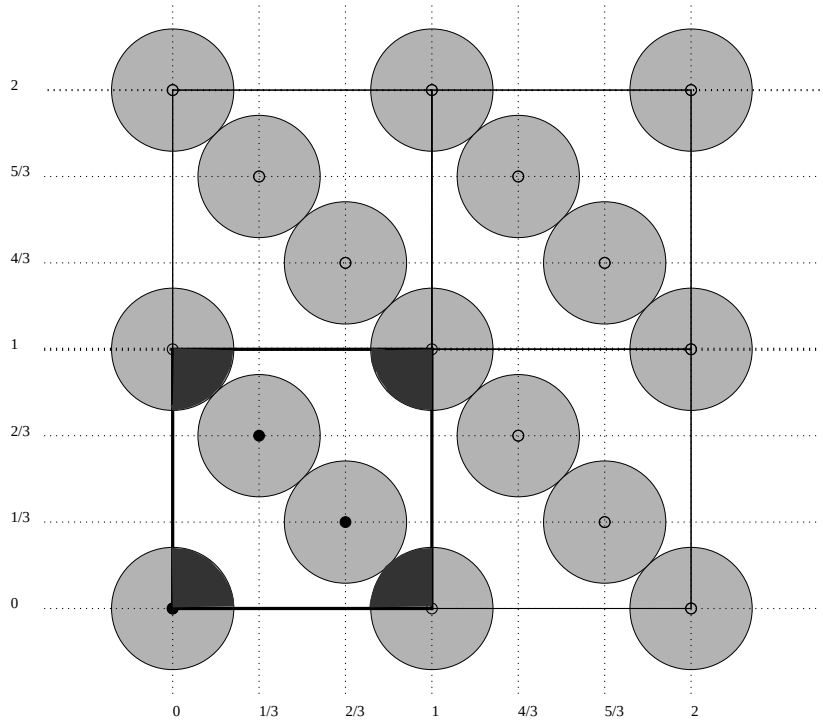

Figure 4: Modulo packing

In our analysis, we will hold r_Λ^{effec} (the effective radius of the Voronoi region) (approximately) constant while taking $n \to \infty$. Using (5) and (6), this in turn dictates that p and k grow with n such that

$$p^k = \frac{1}{V_{\mathcal{B}}(r_\Lambda^{\text{effec}})} = \frac{\Gamma(n/2+1)}{\pi^{n/2}(r_\Lambda^{\text{effec}})^n} \approx \sqrt{n\pi}\left(\frac{n}{2\pi\left(r_\Lambda^{\text{effec}}\right)^2}\right)^{n/2}. \tag{25}$$

To be more precise, since p has to be prime and k is an integer we cannot keep r_Λ^{effec} strictly constant. Rather, it suffices to choose p and k (as a function of n) such that r_Λ^{effec} as defined by (25) satisfies for all n

$$r_{\min} < r_\Lambda^{\text{effec}}(n) < 2 \cdot r_{\min} \tag{26}$$

where $0 < r_{\min} < \frac{1}{4}$ (in Section 7 we further restrict $r_{\min}$).

Note that (25) and (26) imply that if we hold k constant, then p must grow super-exponentially; if k grows linearly with n, then p grows polynomially with n. Furthermore, for k linear in n, i.e., $k = \beta n$ with $\beta < 1$, it follows from (25) that

$$\lim_{n\to\infty} n/p = 0. \tag{27}$$

That is, p goes to infinity "faster" than n. We will eventually require (needed for the covering problem) that k grow faster than $\log^2 n$. Note that this implies that

$$\log p = O(n/\log n). \tag{28}$$

We thus restrict the growth of p as a function of n. This is actually not a necessary condition for the theorems to hold but is imposed for convenience of proof (through equation (26)). Also note that even though p has to be prime equation (26) may be satisfied for every large enough n. To see that, define p^* to be the real number satisfying (25) for a radius of $2r_{\min}$, i.e., $p^{*k} = \frac{1}{V_{\mathcal{B}}(2r_{\min})}$. From (25) and (26) we get that p must satisfy

$$p^* < p < 2^{n/k} p^*. \tag{29}$$

Since we assume that $k < \beta n$ with $\beta < 1$ it follows that (29) may be satisfied as it is known (Bertrand's postulate, see e.g., [18]) that there is a prime number between n and $2n$ for any integer n.

The number of lattice points per unit volume grows super-exponentially with the dimension n. For more physical applications (power constrained transmission, quantization subject to a fixed distortion per dimension) it is more natural to have the lattice density grow exponentially with n. However, this is undesirable as we have seen that in order to be able to cast the two first problems in terms of mod-packing and mod-covering, we need to impose $r_\Lambda^{\text{effec}} < \frac{1}{2}$. Furthermore, to obtain the results for the AWGN channel coding problem we need the further restriction that r_Λ^{effec} be "small" as captured by $r_{\min}$ in (26); see (74). In the sequel we assume that $r_{\min}$ is an arbitrary number such that $r_{\min} \in (0, \frac{1}{4})$ unless otherwise stated (in Section 7). Thus, in order to use a lattice as a code we would in practice need to appropriately scale it.

4 Goodness for Packing

Actually, Loeliger [24] used the lattice ensemble defined above to prove that there exist lattices that are good for packing. We provide here a proof for completeness and uniformity of presentation.

Theorem 1 *A lattice Λ drawn from an (n, k, p) ensemble, where k is sub-linear in n and p,k satisfy (25),(26), is good for packing, i.e.,*

$$\rho_{\text{pack}}(\Lambda) \geq \frac{1}{2}$$

in probability as $n \to \infty$.

Note that Theorem 1 holds even for $k = 1$.
Proof: Denote the diagonal of an elementary cube of GRID* by $2d$,

$$d = \frac{1}{2}\frac{\sqrt{n}}{p}. \tag{30}$$

Let $r < \frac{r_\Lambda^{\text{effec}}}{2} - \frac{d}{2}$. Consider the mod sphere $(r\mathcal{B})^*$. We show that for large enough dimension, with high probability no other sphere intersects it, i.e.,

$$\lim_{n\to\infty} \Pr\left\{(r\mathcal{B})^* \cap ((\Lambda^* \setminus \{\mathbf{0}\}) + r\mathcal{B})^* = \emptyset\right\} = 1. \tag{31}$$

Consider $(\mathbf{x} + r\mathcal{B})^*$, a mod-sphere centered at $\mathbf{x} \in \text{GRID}^*$, It intersects the mod-sphere centered at the origin, $(r\mathcal{B})^*$, iff $\mathbf{x} \in (2r\mathcal{B})^*$ (see Fig 4). Let $\mathcal{S}_1$ denote the set of grid points that are within a (modulo) radius of $2r$ from the origin

$$\mathcal{S}_1 = (\text{GRID}^* \setminus \{\mathbf{0}\}) \cap (2r\mathcal{B})^*.$$

Therefore, the probability that a mod-sphere centered at Λ_i^* $(i \neq 0)$ intersects the mod-sphere at the origin satisfies

$$\Pr\left\{(r\mathcal{B})^* \cap (\Lambda_i^* + r\mathcal{B})^*\right\} \neq \emptyset\} = \Pr\left\{\Lambda_i^* \in \mathcal{S}_1\right\}$$

Let $\mathcal{S}_2$ stand for the set of grid points such their fine cubic cells of side $1/p$ are fully contained in a sphere of radius $2r + d$ from the origin

$$\mathcal{S}_2 = \left\{\mathbf{x} \in \text{GRID}^* : (\mathbf{x} + p^{-1} \cdot \text{CUBE})^* \subset ((2r + d)\mathcal{B})^*\right\}.$$

Notice that $\mathcal{S}_1 \subset \mathcal{S}_2$. We may therefore bound the cardinality of $\mathcal{S}_1$ as follows

$$|\mathcal{S}_1| \leq |\mathcal{S}_2| \leq \lfloor V_\mathcal{B}(2r + d)/\text{Vol}(p^{-1} \cdot \text{CUBE})\rfloor = \lfloor p^n V_\mathcal{B}(2r + d)\rfloor. \tag{32}$$

Consider some arbitrary index i $(i \neq 0)$. Since Λ_i^* is uniformly distributed over the p^n points of GRID*, using (32) we get

$$\begin{aligned}
\Pr\{(\Lambda_i^* + r\mathcal{B})^* \cap (r\mathcal{B})^* \neq \emptyset\} &= \Pr\{[\Lambda_i^* \in \mathcal{S}_1]\} && (33)\\
&= \frac{|\mathcal{S}_1|}{|\text{GRID}|} && (34)\\
&\leq \frac{\lfloor p^n V_\mathcal{B}(2r + d)\rfloor}{p^n} && (35)\\
&\leq V_\mathcal{B}(2r + d). && (36)
\end{aligned}$$

Thus by the union bound we have

$$\begin{aligned}
\Pr\left\{(\Lambda + r \cdot \mathcal{B}) \text{ is a packing}\right\} &\geq 1 - (M - 1) \cdot \Pr\{\Lambda_i^* \in \mathcal{S}_1\} && (37)\\
&> 1 - M \cdot V_\mathcal{B}(2r + d). && (38)
\end{aligned}$$

Now the number of codewords M and the effective radius are related by

$$M = \frac{\text{Vol(CUBE)}}{\text{Vol}(\mathcal{V})} = \frac{1}{\text{Vol}(\mathcal{V})} = \frac{1}{V_\mathcal{B}(r_\Lambda^{\text{effec}})}$$

so that

$$\Pr\{(\Lambda + r\cdot\mathcal{B}) \text{ is a packing}\} \geq 1 - \frac{V_{\mathcal{B}}(2r+d)}{V_{\mathcal{B}}(r_\Lambda^{\text{effec}})} \tag{39}$$

$$= 1 - \left(\frac{2r+d}{r_\Lambda^{\text{effec}}}\right)^n. \tag{40}$$

Recall that $d = \frac{1}{2}\frac{\sqrt{n}}{p}$ so by (27), $d \to 0$ as $n \to \infty$. Therefore as long as we have $\frac{r}{r_\Lambda^{\text{effec}}} < \frac{1}{2}$ we have

$$\lim_{n\to\infty} \Pr\{(\Lambda + r\cdot\mathcal{B}) \text{ is a packing}\} = 1. \tag{41}$$

Recalling the definition of ρ^*_{pack} (7), we conclude that the theorem is proved. □

We note again that p and k are required to satisfy (25). This leaves us with a wide range of possible values for k and p. We could even take $k = 1$. As we shall see next, however, such a choice would not allow us to prove that the ensemble is good for covering. We will ultimately find a sequence of pairs (k_n, p_n) such that the ensemble is good in all four senses.

5 Goodness for Covering

Theorem 2 *A lattice Λ drawn from an (n,k,p) ensemble, where k is sub-linear in n but grows faster than $\log^2 n$ and k,p satisfy (25),(26), is good for covering, i.e.,*

$$\rho_{\text{cov}}(\Lambda) \to 1$$

in probability as $n \to \infty$.

Proof: The proof is a variation on its binary counterpart as given in [3]; see also Chapter 12 of [6]. Consider the set Λ^* of the ensemble. Let k_1, k_2 be such that $k_1 + k_2 = k$. Denote the lattice obtained from the first k_1 rows of G by $\Lambda^*[k_1]$ and let $\Lambda^*[k_1+i]$ denote the lattice obtained from the first $k_1 + i$ rows of G. Consider an arbitrary point $\mathbf{x} \in$ CUBE. For an $r < \frac{1}{2}$, to be chosen later, let $\mathcal{S}_1(\mathbf{x})$ denote the set of grid points within a (modulo) distance $r - d$ from $\mathbf{x}$,

$$\mathcal{S}_1(\mathbf{x}) = \text{GRID} \cap (\mathbf{x} + (r-d)\mathcal{B})^* \tag{42}$$

where d is half the diagonal of an elementary cube of GRID* as defined in (30). Let $\mathcal{S}_2$ denote the set of grid points such that their fine cubic cells intersect a sphere of radius $r - 2d$ centered at $\mathbf{x}$

$$\mathcal{S}_2(\mathbf{x}) = \left\{\mathbf{y} \in \text{GRID} : (\mathbf{y} + p^{-1}\text{CUBE})^* \cap (\mathbf{x} + (r-2d)\mathcal{B})^* \neq \emptyset\right\}.$$

Notice that $\mathcal{S}_2 \subset \mathcal{S}_1$. We may therefore bound the cardinality of $\mathcal{S}_1$ as follows

$$|\mathcal{S}_1| \geq |\mathcal{S}_2| \geq \lceil V_{\mathcal{B}}(r-2d)/\mathrm{Vol}(p^{-1}\mathrm{CUBE})\rceil = \lceil p^n V_{\mathcal{B}}(r-2d)\rceil. \tag{43}$$

Therefore the probability that $\mathbf{x}$ is covered by a sphere of radius $(r-d)$ centered at any specific point of $\Lambda^*[k_1]$ is at least $V_{\mathcal{B}}(r-2d)/\mathrm{CUBE} = V_{\mathcal{B}}(r-2d)$, i.e.,

$$\Pr\{\mathbf{x} \in (\Lambda_i^*[k_1] + (r-d)\mathcal{B})^*\} \geq V_{\mathcal{B}}(r-2d) \quad i = 1, \ldots, M_1 - 1. \tag{44}$$

For any $\mathbf{x} \in \mathrm{CUBE}^*$, $\mathbf{x}$ is covered by $(\Lambda^*[k_1] + (r-d)\mathcal{B})^*$ if and only if at least one codeword is contained in $(\mathbf{x} + (r-d)\mathcal{B})^*$. For every $\mathbf{x} \in \mathrm{GRID}^*$ define the indicator random variable η_i for $i = 1, \ldots, M_1 - 1$, where $M_1 = p^{k_1}$,

$$\eta_i = \begin{cases} 1, & \text{if } \mathbf{x} \in (\Lambda_i^*[k_1] + (r-d)\mathcal{B})^* \\ 0, & \text{otherwise.} \end{cases} \tag{45}$$

Note, that we do not consider $i = 0$ since $\Lambda_0^*[k_1] = \mathbf{0}$ deterministically. Excluding $i = 0$ from our consideration, ensures that η_i is independent of both i and $\mathbf{x}$. Let

$$\chi = \sum_{i=1}^{M_1-1} \eta_i, \tag{46}$$

so that χ is equal to the number of nonzero codewords $(r-d)$-covering $\mathbf{x}$. Taking the expectation of χ and using (44) we get

$$E\{\chi\} = \sum_{i=1}^{M_1-1} E\{\eta_i\} \geq (M_1 - 1)V_{\mathcal{B}}(r-2d). \tag{47}$$

Using the pairwise independence of the η_i's (which implies that they are uncorrelated), we have

$$\begin{aligned} \mathrm{Var}(\chi) &= \sum_{i=1}^{M_1-1} \mathrm{Var}(\eta_i) && (48)\\ &= \sum_{i=1}^{M_1-1} \{E(\eta_i^2) - E(\eta_i)^2\} && (49)\\ &\leq \sum_{i=1}^{M_1-1} E(\eta_i^2) && (50)\\ &= \sum_{i=1}^{M_1-1} E(\eta_i) && (51)\\ &= E(\chi) && (52) \end{aligned}$$

Using (52), by Chebyshev's inequality, for any $\beta > 0$,

$$\Pr\Big\{|\chi - E(\chi)| > 2^{\beta}\sqrt{E(\chi)}\Big\} < \frac{\mathrm{Var}\,(\chi)}{2^{2\beta}E(\chi)} \leq 2^{-2\beta} \tag{53}$$

Let $\mu_\chi = E(\chi)$ and denote

$$\mu(\beta) = \mu_\chi - 2^{\beta}\sqrt{\mu_\chi}. \tag{54}$$

Then from (53) we have

$$\Pr\Big\{\chi < \mu(\beta)\Big\} < 2^{-2\beta} \tag{55}$$

We restrict are attention now to points of GRID*. We call a point $\mathbf{x} \in$ GRID* *remote* with respect to a set $\mathcal{A}$ if it is $(r-d)$-covered less than $\mu(\beta)$ "times" by $(\mathcal{A} + (r-d)\mathcal{B})^*$. Let $Q(\mathcal{A})$ stand for the set of remote points with respect to $\mathcal{A}$. Denote $Q_i = Q(\Lambda_1^*[k_1 + i])$ and $q_i = |Q_i|/|\text{GRID}^*| = |Q_i|/p^n$. Equation (55) then reads

$$E(q_0) < 2^{-2\beta}. \tag{56}$$

Using Markov's inequality, we estimate the deviation of this fraction of points from the mean,

$$\Pr\Big\{q_0 > 2^{\beta}E(q_0)\Big\} < 2^{-\beta}. \tag{57}$$

Therefore, using (56),

$$\Pr\{q_0 > 2^{-\beta}\} \;\; < \;\; 2^{-\beta} \tag{58}$$

Thus, in words, the probability that the fraction of remote points of GRID* be greater than $2^{-\beta}$ is smaller than $2^{-\beta}$. By taking $\beta \to \infty$ (but still keeping $\mu(\beta) > 0$) this probability can be seen to be arbitrarily small as $n \to \infty$. We note that this may be achieved with $k = k_1 = 1$. Therefore, we can obtain an "almost complete" $(r-d)$ covering of GRID*. From this we may conclude that we obtain with high probability an almost complete covering of CUBE with spheres of radius r.

We next show that taking $k \sim log^2 n$ we may obtain a perfect covering. Choose k_1 and p such that (25) and (26) are satisfied (k_1 playing the role of k) with k_1 growing faster than $\log^2 n$. Set

$$\beta = 2\log(\log n + \log\log p). \tag{59}$$

Note that $2^{-\beta}$ goes to zero as $n \to \infty$. We would like to have $\mu(\beta) > 0$. It is sufficient that $\mu > n^\lambda$ with $\lambda > 0$. By (47), this in turn will be satisfied if we choose a radius r that satisfies

$$p^{k_1} - 1 = \frac{n^\lambda}{V_{\mathcal{B}}(r-2d)} \tag{60}$$

Thus we may conclude that with this choice of parameters, for most lattices *almost all* points are covered by spheres of radius r.

However, we wish to have *all* points of CUBE covered. We next show that $\Lambda^*[k_1+k_2]$ allows a complete covering with $k_2 \approx \log n$. Consider the set $\mathcal{S} = \Lambda^*[k_1] \cup (\Lambda^*[k_1] + \boldsymbol{g}_{k_1+1}/p)^*$. Notice that

$$\Lambda^*[k_1+1] = \bigcup_{x=0}^{p-1} \left(\Lambda^*[k_1] + x \cdot \boldsymbol{g}_{k_1+1}/p\right)^* \tag{61}$$

Since the rows of G are statistically independent, we have

$$E(q_1|q_0) \leq E\left\{ \frac{Q(\mathcal{S})}{|\mathrm{GRID}^*|} \Big| q_0 \right\} = q_0^2. \tag{62}$$

Applying Markov's inequality, we have that given q_0,

$$\Pr\{q_1 > 2^{\gamma} E(q_1)\} < 2^{-\gamma}, \tag{63}$$

so that given $q_0 < 2^{-\beta}$

$$\Pr\{q_1 < 2^{\gamma-2\beta}\} \geq 1 - 2^{-\gamma}. \tag{64}$$

Therefore the probability over the whole ensemble that (64) is satisfied for $\Lambda[k_1+1]$ is at least $(1-2^{-\beta})(1-2^{-\gamma})$. Continuing this procedure we get

$$q_{k_2} < 2^{2^{k_2}(\gamma-\beta)-\gamma} \tag{65}$$

with probability $(1-2^{-\beta})(1-2^{-\gamma})^{k_2}$. We would like to choose k_2 such that

$$q_{k_2} < p^{-n} = 2^{-n \log p} \tag{66}$$

as this would imply that $q_{k_2} = 0$ since there are p^n points in GRID. It suffices that $\gamma = \beta - 1$ and

$$k_2 = \lceil \log n + \log\log p \rceil. \tag{67}$$

Recalling that $k = k_1 + k_2$ we conclude that Λ_k^* with probability at least equal to

$$(1-2^{-\beta})(1-2^{-\beta+1})^{(\log n + \log\log p)} \tag{68}$$

satisfies $q_{k_2} < p^{-n}$, i.e., that every $\mathbf{x} \in \mathrm{GRID}^*$ is covered by at least $\mu(\beta)$ balls of radius $(r-d)$. This implies that the probability that there be any $\mu(\beta)$-remote point $\mathbf{x} \in \mathrm{GRID}^*$ is arbitrarily small as $n \to \infty$. Now, if every point of GRID^* is covered by a sphere of radius $(r-d)$ this implies that CUBE is completely covered by M spheres of radius r. Therefore the probability of a complete covering with spheres of radius r goes to one with

$$M = p^{k_1+k_2} = \frac{n^{\lambda}}{V_{\mathcal{B}}(r-2d)} p^{k_2} \tag{69}$$

$$\leq \frac{n^{\lambda}}{V_{\mathcal{B}}(r-2d)} p^{(\log n + \log\log p)+1} \tag{70}$$

$$= \frac{n^{\lambda}}{V_{\mathcal{B}}(r-2d)} 2^{\log p[(\log n + \log\log p)+1]}. \tag{71}$$

We have

$$\frac{r}{r_\Lambda^{\text{effec}}} = \sqrt[n]{\frac{V_{\mathcal{B}}(r)}{V_{\mathcal{B}}(r-2d)} n^\lambda p^{k_2}} \tag{72}$$

$$= \left(\frac{r}{r-2d}\right) \cdot n^{\lambda/n} \cdot 2^{(\log p \log n + \log p \log\log p + \log p)/n}. \tag{73}$$

By (27) and (30) we have $d \to 0$ as $n \to \infty$ and thus also

$$\lim_{n\to\infty} \left(\frac{r}{r-2d}\right) = 1$$

Also since by (28) $p = O(n/\log n)$ as k grows faster than $\log^2 n$, we have

$$\lim_{n\to\infty} 2^{\log p(\log n + \log\log p)/n} = 1.$$

Thus we have that $\frac{r}{r_\Lambda^{\text{effec}}} \to 1$ as $n \to \infty$. This completes the proof. $\square$

6 Goodness for MSE Quantization

As was shown in [39], specifically in equation (27) therein, a lattice that is good for covering is necessarily good for quantization. More specifically,

Proposition 1 *For any lattice* Λ

$$G(\Lambda) \le G_n^* \cdot \frac{n+2}{n} \cdot \left(\rho_{\text{cov}}(\Lambda)\right)^2$$

where $G_n^* \to \frac{1}{2\pi e}$ *is the normalized second moment of an* n*-dimensional sphere.*

Combining this theorem and Theorem 2 we get:

Theorem 3 *A lattice* Λ *drawn from an* (n,k,p) *ensemble, where* k *is sub-linear in* n *but grows faster than* $\log^2 n$*, and p,k satisfy (25),(26), is good for quantization, i.e.,*

$$G(\Lambda) \to \frac{1}{2\pi e}$$

in probability as $n \to \infty$*.*

7 Goodness for AWGN channel coding

In this section we show that this ensemble achieves the Poltyrev random coding error exponent for any noise. In particular, we present a simple derivation for the Poltyrev exponent. We note that Loeliger [24] showed that the ensemble defined in Theorem 1 indeed achieves capacity, but did not point out that the ensemble also achieves the Poltyrev exponent. In this section we need to restrict $r_{\min}$, requiring it to satisfy

$$r_{\min}^2 = \frac{\rho_{\text{AWGN}}^2}{32 E_P(\rho_{\text{AWGN}})}. \tag{74}$$

We have the following theorem.

Theorem 4 *A lattice Λ drawn from an (n,k,p) ensemble, where k is sub-linear in n and p,k satisfy (25),(26) with $r_{\min}$ given by (74) is good for coding for the AWGN channel, i.e.,*

$$P_e(\Lambda, r_N) < e^{-n(E_P(\rho_{AWGN})-o(1))}$$

with probability tending to one as $n \to \infty$. Here $o(1) \to 0$ as $n \to \infty$, but may depend on ρ_{AWGN}.

Note that using the definition of ρ_{AWGN} (19) and of r_N (17) we have

$$P_N = \frac{r_\Lambda^{\text{effec}\,2}}{n\rho_{\text{AWGN}}^2}. \tag{75}$$

Thus, holding r_N fixed means that the variance of the noise decreases inversely with n. We also note that in this theorem we may take $k = 1$.

Proof: For a particular value of ρ_{AWGN} we wish to show that with high probability (w.r.t. the ensemble of lattices)

$$\Pr\{\mathbf{N} \notin \mathcal{V}\} < e^{-n(E_P(\rho_{AWGN})-o(1))}. \tag{76}$$

We prove the theorem through a number of lemmas. Define an auxiliary "truncated" Gaussian noise N_T so that its density is

$$f_{N_T}(x) = \begin{cases} \frac{1}{1-\epsilon_T} f_N(x), & \text{if } x \in [-1/2, 1/2) \\ 0, & \text{otherwise} \end{cases} \tag{77}$$

where

$$\epsilon_T = \Pr\{N \notin [-1/2, 1/2)\} \tag{78}$$

is the probability of truncation. Consider the modulo additive noise channel

$$Y = X + N_T. \tag{79}$$

Since $f_{\mathbf{N}_T}(\mathbf{x})$ is proportional to $f_{\mathbf{N}}(\mathbf{x})$ for any $\mathbf{x} \in$ CUBE it follows that for the noise N_T the decoding regions of Λ^* are the Voronoi regions, just as for the original noise $\mathbf{N}$. Thus the probability of error when transmitting a point of Λ^* over this channel is indeed

$$P_e(\Lambda, \mathbf{N}_T) \stackrel{\Delta}{=} \Pr\{\mathbf{N}_T \notin \mathcal{V}\}. \tag{80}$$

Denote also the random coding and expurgated exponents of this channel by $E^r(\Lambda, N_T)$ and $E^x(\Lambda, N_T)$ respectively.

Lemma 1 *For almost all lattices in the considered ensemble* $P_e(\Lambda, N_T) < e^{-n(E_P(\rho_{AWGN})-o(1))}$.

The proof is given in Appendix A.

We now bound the error probability of the original AWGN channel for Euclidean decoding with this noise by the error probability corresponding to the mod-N_T channel. For those lattices satisfying Lemma 1 we have

$$\begin{aligned}
\Pr(\mathbf{N} \in \mathcal{V}) &= \Pr(\mathbf{N} \in \text{CUBE}) \cdot \Pr(\mathbf{N} \in \mathcal{V} | \mathbf{N} \in \text{CUBE}) && (81)\\
&= [1-\epsilon_T]^n \Pr(\mathbf{N}_T \in \mathcal{V}) && (82)\\
&\geq [1 - 2 \cdot \Pr\{N > 1/2\}]^n \cdot \left[1 - e^{-n(E_P(\rho_{\text{AWGN}})-o(1))}\right] && (83)\\
&\geq [1 - 2n \cdot \Pr\{N > 1/2\}] \left[1 - e^{-n(E_P(\rho_{\text{AWGN}})-o(1))}\right]. && (84)
\end{aligned}$$

So it is sufficient to show that $\Pr\{N > 1/2\}$ decays exponentially in n with an exponent greater that $E_p(\rho_{\text{AWGN}})$. But for small enough P_N we have

$$\Pr\{N > 1/2\} \leq e^{-\frac{(1/2)^2}{2P_N}} = \exp\{-\frac{\rho_{\text{AWGN}}^2}{8{r_\Lambda^{\text{effec}}}^2} \cdot n\}.$$

Therefore, we would like to have

$$\frac{{\rho_{\text{AWGN}}}^2}{8{r_\Lambda^{\text{effec}}}^2} > E_P(\rho_{\text{AWGN}}).$$

This in turn is satisfied by (74). This completes the proof of the theorem. □

8 Simultaneous Goodness

Verifying that there exists a non-empty intersection of the admissible (n, k, p) sets of Theorems 1-4, the result now follows by the union bound.

Theorem 5 *A lattice Λ_n drawn from an (n,k,p) ensemble, where k is sub-linear in n but grows faster than $\log^2 n$ and k,p satisfy (25),(26) with $r_{\min}$ given by (74), is good in all four senses with probability tending to one as $n \to \infty$, i.e.,*

$$\rho_{\text{pack}}(\Lambda_n) \geq \frac{1}{2},\ \rho_{\text{cov}}(\Lambda_n) \to 1,\ G(\Lambda_n) \to \frac{1}{2\pi e}, \text{and } \frac{1}{n}\log P_e(\Lambda_n, r_N) \geq E_p(\rho_{\text{AWGN}})$$

in probability as $n \to \infty$.

To carry the analogies farther we may note the following. Let $\mathcal{S}$ denote the surface of an n-dimensional sphere of radius one and let Θ be a random direction uniformly distributed over $\mathcal{S}$. For a given lattice, define the radius of the basic Voronoi region in direction θ is defined as the intersection of a ray emanating from the origin in direction θ,

$$r_\Lambda(\theta) = \sup_{t:\ t\cdot\theta \in \mathcal{V}} t. \tag{85}$$

Thus $r_\Lambda(\Theta)$ is a random variable taking values in the interval $[r_\Lambda^{\text{pack}}, r_\Lambda^{\text{cov}}]$. The above results yield that for the defined lattice ensemble as $n \to \infty$,

$$\frac{r_\Lambda(\Theta)}{\max_\theta r_\Lambda(\theta)} = \frac{r_\Lambda(\Theta)}{r_\Lambda^{\text{cov}}} \to 1 \ \text{ in probability}$$

and

$$\frac{\min_\theta r_\Lambda(\theta)}{\max_\theta r_\Lambda(\theta)} = \frac{r_\Lambda^{pack}}{r_\Lambda^{cov}} \geq \frac{1}{2} \ \text{ in probability.}$$

Figure 1 depicts the various radii relevant to the different problems.

9 Summary and Extensions

We considered the problems of packing and covering with Euclidean spheres, MSE quantization and AWGN channel coding. Using random coding techniques we demonstrated that there exist lattices that are good for these four problems simultaneously. We saw that for the channel coding problem as well as for the packing problem we had a large freedom in the choice of the behavior of p, the cardinality of the prime field used, and k, the dimension of the underlying linear code, as a function of n. In fact for these problems it is sufficient to take $k = 1$, meaning that a good lattice may be generated by choosing a *single* point. It seems that the same may be true for the quantization problem. The covering problem on the other hand appears to be the most demanding. In order to obtain a good lattice for this problem we were forced to take k growing faster than $\log^2 n$. It remains to be studied whether this condition is indeed necessary or may be relaxed.

As discussed in Section 2, one way to use a lattice for transmission over a power constrained AWGN channel is to use a codebook which is the intersection of the lattice with a

sphere of radius $\sqrt{nPx}$. For the considered ensemble of lattices, we saw that for a constant ρ_{AWGN}, the corresponding noise variance decreases inversely with n. Thus, in such a scheme one would appropriately scale the lattice obtained from the ensemble according to the variance of the channel noise. The results are also directly applicable to the lattice coding scheme of [14] that allows to achieve the AWGN capacity using lattice encoding and decoding. Such a scheme incorporates a pair of nested lattices, the Voronoi region of a the coarse lattice serving as a shaping region, the fine lattice serving as a channel code. It turns out that it is important that the coarse lattice be also good for channel coding. Thus, a lattice obtained from the ensemble proposed in this work may serve as the coarse (shaping) region. The generation of the fine lattice code is described in [14]. Applications where the necessity for lattices which are simultaneously good for channel coding and MSE quantization also arises are in lattice based schemes in various multi-terminal problems. Among them are in the nested code approach for the Costa problem and the Wyner-Ziv problem, see [41].

These quadratic/Euclidean-norm/Gaussian problems are natural in $\mathbb{R}^n$. However, it may be readily verified that Theorem 1 and 2 equally hold for spheres corresponding to any norm. Specifically, consider the case of an r-th power norm:

$$\|\mathbf{x}\|_r = \left(\sum_i x_i^r \right)^{1/r} \qquad \text{r} \geq 1 \tag{86}$$

From the result for covering we get as a corollary, as done in the Euclidean case in Chapter 5, the existence of good quantizers for the single letter distortion measure $d(x, y) = (x - y)^r$. Likewise, the results for channel coding can be extended to more general noise distributions including the exponential family. By a union bound argument, this multi-norm/multi-metric optimality implies an even stronger notion of simultaneous goodness, namely, that there exist lattices which are not only good for the four criteria of sphere arrangements, but also for any finite collection of distortion measures and/or additive noise channels from the family above.

10 Appendix

A Proof of Lemma 1

The difference between any two codewords in the considered ensemble, i.e., $(\Lambda_i - \Lambda_j)^*$, is uniformly distributed over GRID*. Therefore it has the same distribution as that of an ensemble obtained by drawing each codeword according to a uniform distribution over GRID*. Thus, the ensemble has the same random coding and expurgated exponent as the totally random one (see [16, 35]). Denote the corresponding random coding exponent of (79) by $E_{\text{GRID}}^r(R; N_T)$ where the bar denotes averaging over the ensemble. Similarly, denote the

expurgated error exponent of (79) corresponding to a uniform distribution over GRID* by $E^x_{\text{GRID}}(R; N_T)$ and let

$$E_{\text{GRID}}(R; N_T) = \max\left[E^r_{\text{GRID}}(R; N_T), E^{ex}_{\text{GRID}}(R; N_T)\right] \tag{87}$$

Earlier we concluded in Theorem 1 that for almost all lattices from the considered ensemble, the minimum distance between lattice points satisfies the Minkowski bound. This means that the expurgated exponent is achieved for almost all lattices in the ensemble. Thus, the following applies to this subset (the averaging is only over this subset).

$$\bar{P}_e(\Lambda, \mathbf{N}_T) < e^{-n(E_{\text{GRID}}(R;N_T)-o(1))} \tag{88}$$

In Appendix B we show that $E^r_{\text{GRID}}(R; N_T)$ approaches, as $n \to \infty$, the exponent of a uniform distribution over CUBE, which we denote by $E^r(R; N_T)$. That is

$$E^r_{\text{GRID}}(R; N_T) > E^r(R; N_T) - o(1). \tag{89}$$

The claim for the expurgated exponent may be similarly proved. In Appendix C we show that

$$E^r(R; N_T) = E_P(\rho_{\text{AWGN}} - o(1)) \tag{90}$$

and

$$E^x(R; N_T) = E_P(\rho_{\text{AWGN}} - o(1)) \tag{91}$$

for the corresponding values of ρ_{AWGN}. Thus the lemma is proved.

B Proof of (89): Sensitivity of Error Exponent w.r.t. an input on GRID

Consider the random coding error exponent corresponding to the random code restricted to GRID*. It is given by

$$E^r_{\text{GRID}}(R; N_T) = \max_{0<s\leq 1} s\left[E^0_{\text{GRID}}(s; N_T) - sR\right] \tag{92}$$

where

$$E^0_{\text{GRID}}(s; N_T) = -\log \int_{y\in[0,1)} \left(\frac{1}{p}\sum_{i=0}^{p-1} f_{N_T}([y - i/p] \mod \mathbb{Z})^{\frac{1}{1+s}}\right)^{1+s} dy. \tag{93}$$

Compare this to the random coding exponent corresponding to a uniform input, which is given by

$$E^r(R; N_T) = \max_{0<s\leq 1} s\left[E^0(s; N_T) - sR\right] \tag{94}$$

where

$$E^0(s;N_T) = -\log \int_{y\in[0,1)} \left(\int_{\mathbf{x}\in[0,1)} f_{N_T}([y-x] \mod \mathbb{Z})^{\frac{1}{1+s}} d\mathbf{x} \right)^{1+s} dy \quad (95)$$

$$= -\log \int_{y\in[0,1)} \left(\frac{1}{p} \sum_{i=0}^{p-1} \int_{z=-\frac{1}{2p}}^{\frac{1}{2p}} p \cdot f_{N_T}([y-i/p-z] \mod \mathbb{Z})^{\frac{1}{1+s}} dx \right)^{1+s} dy \quad (96)$$

We next show that for any $x \in [0,1)$ and $z \in [-\frac{1}{2p}, \frac{1}{2p})$,

$$p \int_{z=-\frac{1}{2p}}^{\frac{1}{2p}} f_{N_T}([x-z] \mod \mathbb{Z}) = (1+o_n(1)) f_{N_T}(x) \quad (97)$$

For any $\mathbf{x} \in [0,1)$,

$$f_{N_T}(x) = \frac{1}{1-\epsilon_T} \cdot \frac{1}{\sqrt{2\pi P_N}} \cdot e^{\frac{-x^2}{2P_N}} \propto e^{\frac{-x^2}{2P_N}}. \quad (98)$$

Since

$$x^2 - 2|x||z| + z^2 \le (x+z)^2 \le x^2 + 2|x||z| + z^2 \quad (99)$$

we have

$$f_{N_T}(x) e^{\frac{-z^2-2|x||z|}{2P_N}} \le f_{N_T}(x+z \mod \mathbb{Z}) \le f_{N_T}(x) e^{\frac{+z^2+2|x||z|}{2P_N}} \quad (100)$$

Taking into account (75) and that $|x| \le 1$, $|z| \le 1/p$ and $r_\Lambda^{\text{effec}} \ge r_{\min}$ we have

$$\frac{z^2+2|x||z|}{2P_N} \le \frac{z^2+2z}{2P_N} \le \frac{3/p}{2\frac{r_\Lambda^{\text{effec}\,2}}{n\rho_{AWGN}^2}} \le \frac{3}{2} \cdot \frac{\rho_{\text{AWGN}}^2}{r_{\min}^2} \cdot \frac{n}{p} \quad (101)$$

Thus, since $\lim_{n\to\infty} \frac{n}{p} = 0$ by (27), we have that

$$f_{N_T}(x+z \mod \mathbb{Z}) = (1+o_n(1)) f_{N_T}(x). \quad (102)$$

Consequently we obtain (97). Substituting (97) into (96), it follows that

$$E_{\text{GRID}}^0(s;N_T) - E^0(s;N_T) = o_n(1) \quad (103)$$

This proves the lemma.

C Calculation of Exponents

The derivation below (as well as Appendix B) is a special case of a calculation carried out in the Appendix of [14]. We include it for completeness.

For any modulo additive noise channel over the basic interval $[0,1)$

$$Y = X + Z \mod \mathbb{Z} \tag{104}$$

the random coding exponent is given by (see, e.g., [13])

$$\begin{aligned} E^r(\delta; Z) &= \max_{0<s\leq 1} s\Big[\ln|[0,1)| - h_{\frac{1}{1+s}}(Z) - R\Big] && (105)\\ &\geq \max_{0<s\leq 1} s\Big[\ln e^{-R} - h_{\frac{1}{1+s}}(Z)\Big] && (106)\\ & && (107)\\ &= \max_{0<s\leq 1} s[-\ln\delta - h_{\frac{1}{1+s}}(Z)] && (108)\\ &= \max_{0<s\leq 1} s h_{\bar{s}}(\delta Z) && (109) \end{aligned}$$

where $h_\gamma(\cdot)$ denotes Rényi entropy, defined for any γ, $0<\gamma<1$,

$$h_\gamma(Z) = \frac{\gamma}{1-\gamma}\ln\left(\int_x f_Z(x)^\gamma dx\right)^{\frac{1}{\gamma}}$$

and

$$\delta = e^R = \frac{1}{[\mathrm{Vol}(\mathcal{V})]^{1/n}} \tag{110}$$

is the normalized density of the lattice points per dimension, i.e., the normalized number of lattice points per unit volume.

Thus, the random coding exponent of the N_T-modulo additive noise channel channel is given by

$$E^r(R; N_T) = \max_{0\leq s\leq 1} s\big[\log|[0,1)| - h_{\bar{s}}(N_T) - R\big] \tag{111}$$

We further have

$$\begin{aligned} h_\gamma(N_T) &= \frac{1}{1-\gamma}\log\int_o^1\left(\frac{f_N(x)}{1-\epsilon_T}\right)^\alpha dx && (112)\\ &< \frac{1}{1-\gamma}\log\int_{-\infty}^{\infty}\left(\frac{f_N(x)}{1-\epsilon_T}\right)^\gamma dx && (113)\\ &= \frac{-\gamma}{1-\gamma}\log(1-\epsilon_T) + \frac{1}{1-\gamma}\log\int_{-\infty}^{\infty} f_N(x)^\gamma dx. && (114) \end{aligned}$$

Taking $\gamma = \bar{s} = \frac{1}{1+s}$ we get

$$s h_{\bar{s}}(Z_T) < s h_{\bar{s}}(N) + \log\frac{1}{1-\epsilon_T}. \tag{115}$$

We therefore have

$$E^r(R;N_T) = \max_{0\le s\le 1} s\big[\log|[0,1)| - h_{\bar{s}}(N) - R\big] - o(1). \tag{116}$$

For a Gaussian random variable $N\sim\mathcal{N}(0,P_N)$ the Rényi entropy is

$$h_\gamma(N) = \frac{\gamma}{1-\gamma}\ln\left(\int_x\left(\frac{1}{\sqrt{2\pi P_N}}e^{-\frac{x^2}{2P_N}}\right)^\gamma dx\right)^{\frac{1}{\gamma}} \tag{117}$$

$$= \frac{1}{2}\ln 2\pi e P_N - \frac{1}{2}\ln e\gamma^{\frac{1}{1-\gamma}} \tag{118}$$

The maximization in (116) is carried out explicitly in Appendix C.1 to yield

$$\max_{0\le s\le 1} s\big[\log|[0,1)| - h_{\bar{s}}(N) - R\big] = \frac{1}{2}\left[(\rho^2_{\mathrm{AWGN}}-1) - \ln\rho^2_{\mathrm{AWGN}}\right]. \tag{119}$$

This yields (90) for ρ_{AWGN} in the random coding region, i.e., corresponding to the first line of (23).

We next treat the expurgated exponent. Define the following generalized Bhattacharyya distances, for $s\ge 1$:

$$D^{Bhatt}_{\mathrm{mod}}(N_T;s) = -\ln\int_0^1\left(\int_0^1\sqrt{f_{N_T}(y)f_{N_T}((y+x)^*)}d\mathbf{y}\right)^{\frac{1}{s}} \tag{120}$$

and

$$D^{Bhatt}(N;s) = -\ln\int_{x\in\mathbb{R}}\left(\int_{y\in\mathbb{R}}\sqrt{f_N(y)f_N(y+x)}dy\right)^{\frac{1}{s}}dx \tag{121}$$

For a modulo additive noise channel the expurgated exponent is achieved by a uniform input [16], giving

$$\begin{aligned}
E^x(R;Z) &= \sup_{s\ge 1} s\left[-\ln\int_0^1\int_0^1\left(\int_0^1\sqrt{f(y|x_1)f(y|x_2)}\,dy\right)^{\frac{1}{s}}dx_1dx_2 - R\right]\\
&= \sup_{s\ge 1} s\left[-\ln\int_0^1\int_0^1\left(\int_0^1\sqrt{f_Z((y-x_1)^*)f_Z((y-x_2)^*)}\,dy\right)^{\frac{1}{s}}dx_1dx_2 - R\right]\\
&= \sup_{s\ge 1} s\left[-\ln\int_0^1\left(\int_0^1\sqrt{f_Z(y)f_Z((y+x)^*)}\,dy\right)^{\frac{1}{s}}dx - R\right]\\
&= \sup_{s\ge 1} s\left[D^{Bhatt}_{\mathrm{mod}}(Z;s) - R\right]
\end{aligned} \tag{122}$$

Similarly to (115) we have

$$sD^{Bhatt}(N_T;s) < sD^{Bhatt}(N;s) + \log\frac{1}{1-\epsilon_T}. \tag{123}$$

Thus

$$\sup_{s\geq 1} s\left[D_{\text{mod}}^{Bhatt}(N_T;s)-R\right] = \sup_{s\geq 1} s\left[D_{\text{mod}}^{Bhatt}(N;s)-R\right]+o(1). \tag{124}$$

In Appendix C.2 the maximization of (124) is explicitly carried out to yield

$$\sup_{s\geq 1} s\left[D^{Bhatt}(N;s)-R\right] = E_P(\rho_{\text{AWGN}}) \tag{125}$$

for $\rho_{\text{AWGN}} \geq 4$.

C.1 Maximization of (116)

For Gaussian X we have

$$\begin{aligned} h_{\bar{s}}(X) &= h(X) - \frac{1}{2}\ln e\bar{s}^{\frac{1}{1-\bar{s}}} && (126)\\ &= \frac{1}{2}\left[\ln 2\pi e P_X - \frac{1}{1-\bar{s}}\ln\bar{s} - 1\right] && (127)\\ &= \frac{1}{2}\left[\ln 2\pi e P_X + \frac{1+s}{s}\ln(1+s) - 1\right] && (128)\end{aligned}$$

We therefore have

$$sh_{\bar{s}}(X) = \frac{s}{2}\ln 2\pi e P_X + \frac{1+s}{2}\ln(1+s) - \frac{s}{2}. \tag{129}$$

Taking the derivative of $sh_{\bar{s}}(X)$ with respect to s we get

$$\frac{d}{ds}\left[sh_{\bar{s}}(X)\right] = \frac{1}{2}\left[\ln 2\pi e P_X + \ln(1+s)\right] \tag{130}$$

Thus an extremum occurs when

$$\ln(1+s) = -\ln 2\pi e P_X \tag{131}$$

or equivalently when

$$s = \frac{1}{2\pi e P_X} - 1. \tag{132}$$

It is easy to verify that this extremum is indeed a maximum. Substituting (131) and (132) into (129) we get

$$\max_s sh_{\bar{s}}(X) = \frac{1}{2}\left[-\frac{1}{2\pi e P_X} - \ln 2\pi e P_X + 1\right]. \tag{133}$$

From the definition of δ (110) we get

$$P_X = \delta^2 P_N = \frac{P_N}{[\mathrm{Vol}(\mathcal{V})]^{2/k}} = \frac{P_N}{\rho^2_{\mathrm{AWGN}} 2\pi e P_N} = \frac{1}{2\pi e \rho^2_{\mathrm{AWGN}}}. \tag{134}$$

Substituting (134) for P_X in (133) we get

$$\begin{aligned} E^r(R) &= \max_s s h_{\bar{\rho}}(X) && (135)\\ &= \frac{1}{2}\left[-\rho^2_{\mathrm{AWGN}} - \ln \rho^2_{\mathrm{AWGN}} + 1\right] && (136)\\ &= \frac{1}{2}(1-\rho^2_{\mathrm{AWGN}}) - \ln \rho^2_{\mathrm{AWGN}}. && (137) \end{aligned}$$

C.2 Maximization of (124)

We use the following property of Gaussian distributions

$$f_N(y+x) f_N(x) = f_{N/\sqrt{2}}(y+x/2) f_{\sqrt{2}N}(x) \tag{138}$$

Let $\sigma^2 = \mathrm{Var}(N)$. Then the above relation holds since

$$\begin{aligned} f_N(y) f_N(y+x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+x)^2}{2\sigma^2}} && (139)\\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2 e^{-\frac{2y^2+2yx+x^2}{2\sigma^2}} && (140)\\ &= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})} \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} e^{-\frac{(\sqrt{2}y+x/\sqrt{2})^2}{2\sigma^2}} \cdot e^{-\frac{(x/\sqrt{2})^2}{2\sigma^2}} && (141)\\ &= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})} e^{-\frac{(y+x/2)^2}{2(\sigma/\sqrt{2})^2}} \cdot \frac{1}{\sqrt{2\pi}(\sqrt{2}\sigma)} e^{-\frac{x^2}{2(\sqrt{2}\sigma)^2}}. && (142) \end{aligned}$$

Using (139) we obtain,

$$\begin{aligned} D_s^{Bhatt}(N;s) &= -\ln \int_{x\in\mathbb{R}} \left(\int_{y\in\mathbb{R}} \sqrt{e^{-\frac{y^2}{2\sigma^2}} e^{-\frac{(y+x)^2}{2\sigma^2}}}\, dy \right)^{\frac{1}{s}} dx && (143)\\ &= -\ln \int_x \left(\int_y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y+x/2)^2}{2\sigma^2}} e^{-\frac{x^2}{2(2\sigma)^2}} dy \right)^{\frac{1}{s}} dx && (144)\\ &= -\ln \int_x e^{-\frac{x^2}{2(2\sqrt{s}\sigma)^2}} dx && (145)\\ &= -\ln \frac{1}{\sqrt{2\pi}\, 2\sqrt{s}\sigma} && (146)\\ &= -\frac{1}{2}\ln 8\pi\sigma^2 s. && (147) \end{aligned}$$

Thus, it is left to show that

$$\sup_{s\geq 1} s[\ln\delta - \frac{1}{2}\ln 8\pi s\sigma^2] = E_P(\rho_{\text{AWGN}}) \tag{148}$$

for $\rho_{\text{AWGN}} \geq 4$. Differentiating the left hand side of (148) we get

$$\frac{d}{ds}s\left[\ln\delta - \frac{1}{2}\ln 8\pi s\sigma^2\right] = \left[\ln\delta - \frac{1}{2}\ln 8\pi s\sigma^2\right] - s\frac{1}{2}\frac{8\pi\sigma^2}{8\pi\sigma^2 s} \tag{149}$$

$$= \ln\delta - \frac{1}{2}\ln 8\pi\sigma^2 s - \frac{1}{2}. \tag{150}$$

Equating to zero we obtain

$$\ln\frac{\delta^2}{8\pi\sigma^2 s} = 1 \tag{151}$$

or equivalently

$$s = \frac{\delta^2}{8\pi e\sigma^2}. \tag{152}$$

Taking into account that $\sigma^2 = P_N$, we get

$$\sup_{s\geq 1} s[\ln\delta - \frac{1}{2}\ln 8\pi s\sigma^2] = \frac{\delta^2}{16\pi e\sigma^2} = \frac{[\text{Vol}(\mathcal{V})]^{2/n}}{16\pi e P_N} = \frac{\rho^2_{\text{AWGN}}}{8}. \tag{153}$$

Finally note that from (152) we have that $s = 1$ indeed corresponds to a rate satisfying

$$\rho^2_{\text{AWGN}} = 4. \tag{154}$$

Finally, the straight line part of $E_P(\rho_{\text{AWGN}})$ is obtained by combining the results for the random coding exponent and the expurgated exponent.

References

[1] A. Barg and G. D. Forney Jr. Random codes: Minimum distances and error exponents. *IEEE Trans. Information Theory*, IT-48:2568–2573, Sept., 2002.

[2] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[3] V. M. Blinovskiĭ. A lower bound on the number of words of a linear code in an arbitrary sphere with given radius in $\mathbb{F}_q^n$. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 23(2):50–53, 1987.

[4] E. L. Blokh and V. V. Zyablov. *Linear Concatenated Codes (in Russian).* Moskva: Nauka, 1982.

[5] G.J. Butler. Simultaneous packing and covering in Euclidean space. *Proc. London Mat. Soc.*, 25:721–735, 1972.

[6] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein. *Covering Codes.* North Holland Publishing, Amsterdam, The Netherlands, 1997.

[7] G. D. Cohen. A nonconstructive upper bound on covering radius. *IEEE Trans. Information Theory*, IT-29:352–353, May 1983.

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

[9] R. de Buda. Some optimal codes have structure. *IEEE Jr. on Selected Areas in Comm.*, 7:893–899, Aug. 1989.

[10] P. Delsarte and P. M. Piret. Do most binary linear codes achieve the Goblick bound on the covering radius? *IEEE Trans. Information Theory*, IT-32:826–828, Nov. 1986,.

[11] R.L. Dobrushin. Asymptotic optimality of group and systematic codes for some channels. *Theor. Probab. Appl.*, 8:52–66, 1963.

[12] P. Elias. Coding for noisy channels. *IRE Concention Record*, Part 4:37–46, 1955, Also appears in *Key Papers in the Development of Information Theory*, Ed. D. Slepian, IEEE Press, 102-111, 1974.

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

[14] U. Erez and R. Zamir. Lattice decoding can achieve $\frac{1}{2}\log(1+SNR)$ on the AWGN channel. *IEEE Trans. Information Theory*, submitted, 2001.

[15] R. G. Gallager. *Low Density Parity Check Codes.* M.I.T Press, Cambridge, Massachusetts, 1963.

[16] R. G. Gallager. *Information Theory and Reliable Communication.* Wiley, New York, N.Y., 1968.

[17] A. Gersho. Asymptotically optimal block quantization. *IEEE Trans. Information Theory*, IT-25:373–380, July 1979.

[18] G.H.Hardy and E. M. Wright. *An introduction to the Theory of Numbers.* Oxford University Press, Oxford, fifth edition, 1979.

[19] E. N. Gilbert. A comparison of signalling alphabets. *Bell Sys. Tech. Jour.*, 31:504–522, 1952.

[20] T. J. Goblick. *Coding for a discrete information source with a distortion measure.* Ph.D. dissertation, Dept. Elec. Eng., MIT, MA, 1962.

[21] G.A. Kabataiansky and V. I. Levenshtein. Bounds for packings on a sphere and in space (in Russian). *Problemy Pered. Inform.*, 14(1):3–25, 1978. English translation: *Problems of Information Transmission*, 14(1):1–17, 1978.

[22] T. Linder, Ch. Schlegel, and K. Zeger. Corrected proof of de Buda's theorem. *IEEE Trans. Information Theory*, IT-39:1735–1737, Sept. 1993.

[23] T. Linder and R. Zamir. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Information Theory*, IT-40:2026–2031, Nov. 1994.

[24] H. A. Loeliger. Averaging bounds for lattices and linear codes. *IEEE Trans. Information Theory*, 43:1767–1773, Nov. 1997.

[25] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error Correcting Codes.* Elsevier, 1977.

[26] H. Minkowski. Dichteste gitterförmige Lagerung kongruenter Körper. *Nachr. Ges. Wiss. Göttingen*, pages 311–355, 1904.

[27] J. N. Pierce. Limit distribution of the minimum distance of random linear codes. *IEEE Trans. Information Theory*, IT-13:595–599, Oct. 1967.

[28] V. S. Pless, W. C. Huffman, and R. A. Brualdi, editors. *Handbook of Coding Theory.* North-Holland, Amsterdam, 1998.

[29] G. Poltyrev. On coding without restrictions for the AWGN channel. *IEEE Trans. Information Theory*, IT–40:409–417, Mar. 94.

[30] C. A. Rogers. A note on coverings. *Mathematica*, 4:1–6, 1957.

[31] C. A. Rogers. Lattice coverings of space. *Mathematica*, 6:33–39, 1959.

[32] C.A. Rogers. *Packing and covering.* Cambridge University Press, Cambridge, 1964.

[33] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, Vol. 27:379–423, July 1948.

[34] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell Syst. Tech. J.*, 38:611–656, May 1959.

[35] N. Shulman and M. Feder. On universal channel priors. *IEEE Trans. Information Theory*, in preparation.

[36] R. Urbanke and B. Rimoldi. Lattice codes can achieve capacity on the AWGN channel. *IEEE Trans. Information Theory*, IT-44:273–278, Jan. 1998.

[37] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Dokl. Akad. Nauk SSSR*, 117(5):739–741, 1957.

[38] P. L. Zador. Topics in the asymptotic quantization of continuous random variables. Bell Lab. Tech. Memo., 1966.

[39] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Information Theory*, IT-42:1152–1159, July 1996.

[40] R. Zamir and M. Feder. On universal quantization by randomized uniform / lattice quantizer. *IEEE Trans. Information Theory*, pages 428–436, March 1992.

[41] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Information Theory*, IT–48:1250 –1276, June, 2002.

[42] J. Ziv. On universal quantization. *IEEE Trans. Information Theory*, IT-31:344–347, May 1985.