A Simple Proof for the Existence of "Good" Pairs of Nested Lattices

Or Ordentlich and Uri Erez, *Member, IEEE*

***Abstract*— This paper provides a simplified proof for the existence of nested lattice codebooks allowing to achieve the capacity of the additive white Gaussian noise channel, as well as the optimal rate-distortion tradeoff for a Gaussian source. The proof is self-contained and relies only on basic probabilistic and geometrical arguments. An ensemble of nested lattices that is different, and more elementary, than the one used in the previous proofs is introduced. This ensemble is based on lifting different subcodes of a linear code to the Euclidean space using Construction A. In addition to being simpler, the analysis is less sensitive to the assumption that the additive noise is Gaussian. In particular, for additive ergodic noise channels, it is shown that the achievable rates of the nested lattice coding scheme depend on the noise distribution only via its power. Similarly, the nested lattice source coding scheme attains the same rate-distortion tradeoff for all ergodic sources with the same second moment.**

***Index Terms*— Lattice codes, linear codes, additive noise channels.**

I. Introduction

WHILE lattices are the Euclidean space counterpart of linear codes in Hamming space, the two fields historically developed along quite different paths. From the onset of coding theory, linear codes were treated both using algebraic tools as well as via probabilistic methods. The history of the theory of lattices began much earlier, and with the exception of the Minkowski-Hlawka theorem, its development leaned heavily on purely algebraic constructions until quite recently. This has led to a rather convoluted path for arriving at basic proofs for the existence of lattices possessing "goodness" properties that are central to communication problems. The goal of this work is to provide a simple proof for the existence of lattices with the minimal "goodness" requirements necessary for achieving the capacity of the AWGN channel, as well as the optimal rate-distortion tradeoff for a white Gaussian source.

A major difference between linear codes and lattices is that the former are finite, while the latter are unbounded. As a result, the application of linear codes to communication settings is more straightforward. The application of lattices for communication problems requires intersecting the (infinite) lattice with a finite shaping region, in order to construct a codebook.

For the problem of source coding, it has been recognized early on [1] that the significance of the shaping region becomes less crucial as the quantization resolution grows. Indeed, high resolution is the natural operating point in practical systems, and thus neglecting the shaping region and studying the quantization performance of the lattice is sufficient. Namely, the performance of a lattice quantizer at high resolution, is dictated by its normalized second moment. The asymptotic optimality of lattice quantizers in the latter sense, was established in [2], where the existence of sequences of lattices whose normalized second moment approaches that of a high-dimensional ball, was demonstrated. Such sequences of lattices are called *good for MSE quantization*. A stronger requirement is that the worst-case squared error distortion attained by a sequence of lattices approaches its average. Sequences of lattices that satisfy this property are called *good for covering* and were shown to exist by Rogers [3]. In fact, [2] relied on the result of [3] to establish the existence of lattices that are good for MSE quantization.

When it comes to channel coding, the equivalent of the high resolution regime is that of high transmission rate. However, communication systems supporting a very large number of information bits per dimension are seldom encountered. As a consequence, it was not until the 1970s that lattice codes were considered for the channel coding problem, starting with the works of Blake [4] and de Buda [5], and continuing with [6]–[8]. In these works, the shaping region was naturally taken to be a ball (or a thin spherical shell), which is efficient in terms of power, but results in a codebook with weaker symmetry than the original lattice. Poltyrev [9] bypassed this obstacle, by adopting a path analogous to high resolution quantization, and studied the performance of lattices for the unrestricted additive white Gaussian noise (AWGN) channel. In particular, Poltyrev established the existence of sequences of lattices for which the probability of erroneous detection approaches (in an exponential sense) that of AWGN leaving an effective ball whose volume matches the density of the lattice. Such sequences of lattices are called *Poltyrev good*. As a corollary, it follows that there exist sequences of lattices for which the probability of erroneous detection approaches zero as long as the variance of the AWGN is no greater than the squared radius of the effective ball. Such lattices are called *good for channel coding*. We refer the reader to [10, Ch. 7] for a more comprehensive definition and treatment of asymptotic goodness properties of lattices.

Manuscript received August 7, 2015; revised January 28, 2016; accepted May 14, 2016. Date of publication May 23, 2016; date of current version July 12, 2016. O. Ordentlich was supported by the MIT–Technion Post-Doctoral Fellowship. U. Erez was supported by ISF under Grant 1956/15. The material in this paper was presented in part at the 2012 27th IEEE Convention of Electrical and Electronics Engineers in Israel.

O. Ordentlich is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: ordent@mit.edu).

U. Erez is with the Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv 69978, Israel (e-mail: uri@eng.tau.ac.il).

Communicated by T. Liu, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2571719

An alternative approach [11], [12] to using a spherical shaping region, is using a *nested lattice pair* $\Lambda_c \subset \Lambda_f$, where the Voronoi region $\mathcal{V}_c$ of the lattice Λ_c is used for shaping, such that the codebook is $\mathcal{L} = \Lambda_f \cap \mathcal{V}_c$. This approach has the advantage of retaining the lattice symmetry structure. In particular, it was shown [13] that there exist sequences of such codebooks that can attain any rate below $\frac{1}{2}\log(\mathsf{SNR})$ with *lattice decoding*, i.e., nearest neighbor decoding over the infinite lattice Λ_f. See also [10].

Finally, [14] introduced a coding scheme using nested lattice pairs in conjunction with MMSE estimation and dithering. This scheme was shown to attain capacity, as well as the Poltyrev error exponent, with lattice decoding. This was later modified to achieve a better error exponent in [15]. It is worthwhile noting, that the proof of [14] hinged on the coarse lattice being good for covering, and the fine lattice being Poltyrev good. A similar MMSE estimation approach for the source coding problem, was shown to achieve the rate-distortion function of a Gaussian source [16].

The nested lattice coding scheme of [14], which is described in detail in Section III, transformed the AWGN to a modulo-additive channel, where the additive noise is a linear *mixture* of AWGN and a dither uniformly distributed over the Voronoi region of the coarse lattice. In order to establish that this scheme achieves the capacity of the AWGN channel, the authors first derived its error exponent, and then obtained the capacity result as a corollary. Their error exponent analysis required showing that the probability density function of the mixture noise is upper bounded by that of AWGN with the same second moment, times some term that becomes insignificant as the dimension increases. This in turn, imposed the requirement that the coarse lattice be good for covering. Furthermore, the interest in error exponents led to the requirement that the fine lattice be Poltyrev good.

Consequently, the proof of the error exponent and capacity results in [14] required showing the existence of a sequence of nested lattice pairs where the fine lattice is Poltyrev good, and the coarse lattice is Rogers good. To this end, an ensemble of random Construction A lattices, rotated by the generating matrix of a lattice good for covering, was defined and analyzed. The proof therefore relied on the existence of lattices that are good for covering, which made it indirect, complicated, and overly stringent.

In the last decade lattice codes were found to play a new role in network information theory allowing to obtain new achievable rate regions, that are not achievable using the best known random coding schemes, for many problems [17]–[22]. See [10, Chapter 12], for a comprehensive survey. The scheme of [14], or its variations, plays an important role in many of these new techniques. However, since the capacity region is not known for the majority of problems in network information theory, determining the optimal error exponents is far out of scope. Therefore, it is the capacity result from [14], rather than the error exponent one, that is often used in this context.

This paper relaxes the goodness properties required by a nested lattice pair in order to be capacity achieving. Namely, we show that a pair of nested lattices where the fine lattice is good for coding and the coarse lattice good for MSE quantization, suffices to achieve the capacity of the AWGN channel under the scheme from [14]. In fact we prove a more general result, showing that the scheme from [14] applied with such nested lattice pairs can reliably achieve any rate smaller than $\frac{1}{2}\log(1+\mathsf{SNR})$ over all additive *semi norm-ergodic* noise channels. An analogous result holds for quantization.

The class of semi norm-ergodic processes includes all processes whose empirical variance is with high probability not much greater than the variance. In [23] Lapidoth showed that i.i.d. Gaussian codebooks with nearest neighbor decoding can achieve any rate smaller than $\frac{1}{2}\log(1+\mathsf{SNR})$ over the same class of channels. Our result is therefore the lattice codes analogue of [23]. Moreover, it immediately implies that many nested lattice based coding schemes for Gaussian networks are in fact robust to the exact statistics of the noise, and merely require it to be semi norm-ergodic.

A key result we obtain, is that a dither uniformly distributed over the Voronoi region of a lattice that is good for MSE quantization is semi norm-ergodic, and moreover, any linear combination of such a dither and semi norm-ergodic noise, is itself semi norm-ergodic. This enables to relax the goodness for covering requirement of the coarse lattice, to goodness for MSE quantization.

Our analysis also naturally extends to the more practical case, where the coarse lattice is the simple one-dimensional cubic lattice, whereas the fine lattice is a Construction A lattice based on some p-ary linear code. We show that for large p, the scheme from [14] can reliably achieve any rate smaller than $\frac{1}{2}(1+\mathsf{SNR}) - \frac{1}{2}\log(2\pi e/12)$ with such a coarse lattice. We further explicitly upper bound the loss incurred by using any finite value of p.

Most importantly, we provide a simple, self-contained proof for the existence of nested lattice chains $\Lambda_1^{(n)} \subset \cdots \subset \Lambda_L^{(n)}$, for any finite L, where all lattice sequences $\Lambda_1^{(n)}, \cdots, \Lambda_L^{(n)}$ are good for MSE quantization and for coding. Although this result is not new, and can be obtained as a simple corollary of [24], our proof techniques are quite different and considerably simpler. In particular, we define a novel ensemble of nested lattice chains, based on drawing a random linear p-ary code and using Construction A to lift L of its sub-codes to the Euclidean space. This ensemble, which is a direct extension of the ensemble of nested linear binary codes proposed by Zamir and Shamai in [25], allows for a direct analysis of the goodness figures of merit of its members. Consequently, our existence proof requires only elementary probabilistic and geometrical arguments.

II. Preliminaries on Lattice Codes

A lattice Λ is a discrete subgroup of $\mathbb{R}^n$ which is closed under reflection and real addition. Any lattice Λ in $\mathbb{R}^n$ is spanned by some $n \times n$ matrix $\mathbf{F}$ such that

$$\Lambda = \{\mathbf{t} = \mathbf{F}\mathbf{a} : \mathbf{a} \in \mathbb{Z}^n\}.$$

We denote the nearest neighbor quantizer associated with the lattice Λ by

$$Q_\Lambda(\mathbf{x}) \triangleq \arg\min_{\mathbf{t}\in\Lambda} \|\mathbf{x} - \mathbf{t}\|. \tag{1}$$

The basic Voronoi region of Λ, denoted by $\mathcal{V}$, is the set of all points in $\mathbb{R}^n$ which are quantized to the zero vector, where ties in (1) are broken in a systematic manner. The modulo operation returns the quantization error w.r.t. the lattice,

$$[\mathbf{x}] \bmod \Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x}),$$

and satisfies the distributive law,

$$\big[[\mathbf{x}] \bmod \Lambda + \mathbf{y}\big] \bmod \Lambda = \big[\mathbf{x} + \mathbf{y}\big] \bmod \Lambda.$$

Let $V(\Lambda)$ be the volume of a fundamental cell of Λ, i.e., the volume of $\mathcal{V}$, and let $\mathbf{U}$ be a random variable uniformly distributed over $\mathcal{V}$. We define the second moment per dimension associated with Λ as

$$\sigma^2(\Lambda) \triangleq \frac{1}{n}\mathbb{E}\|\mathbf{U}\|^2 = \frac{1}{n}\frac{\int_{\mathcal{V}} \|\mathbf{x}\|^2 d\mathbf{x}}{V(\Lambda)}.$$

The normalized second moment (NSM) of a lattice Λ is defined by

$$G(\Lambda) \triangleq \frac{\sigma^2(\Lambda)}{V^{\frac{2}{n}}(\Lambda)}.$$

Note that this quantity is invariant to scaling of the lattice Λ.

It is often useful to compare the properties of the Voronoi region $\mathcal{V}$ with those of a ball.

Definition 1: Let

$$\mathcal{B}(\mathbf{s}, r) \triangleq \left\{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x} - \mathbf{s}\| \le r\right\},$$

denote the closed n-dimensional ball with radius r centered at $\mathbf{s}$. We denote the volume of an n-dimensional ball with unit radius by V_n. In general $V\left(\mathcal{B}(\mathbf{s}, r)\right) = V_n r^n$. Note that $nV_n^{\frac{2}{n}}$ is monotonically increasing in n, and satisfies $4 \le nV_n^{\frac{2}{n}} < 2\pi e$ for all n [26], and

$$\lim_{n\to\infty} nV_n^{\frac{2}{n}} = 2\pi e. \tag{2}$$

By the isoperimetric inequality, the ball $\mathcal{B}(\mathbf{0}, r)$ has the smallest second moment per dimension out of all (measurable) sets in $\mathbb{R}^n$ with volume $V_n r^n$, and it is given by

$$\begin{aligned}\sigma^2\left(\mathcal{B}(\mathbf{0}, r)\right) &= \frac{1}{n}\frac{1}{V_n r^n}\int_{\mathbf{x}\in\mathcal{B}(\mathbf{0},r)} \|\mathbf{x}\|^2 d\mathbf{x} \\ &= \frac{1}{n}\frac{1}{V_n r^n}\int_0^r r'^2 d(V_n r'^n) \\ &= \frac{1}{n}\frac{1}{V_n r^n}\frac{nV_n r^{n+2}}{n+2} \\ &= \frac{r^2}{n+2}.\end{aligned} \tag{3}$$

It follows that $\mathcal{B}(\mathbf{0}, r)$ has the smallest possible NSM

$$G\left(\mathcal{B}(\mathbf{0}, r)\right) = \frac{\sigma^2\left(\mathcal{B}(\mathbf{0}, r)\right)}{V^{\frac{2}{n}}\left(\mathcal{B}(\mathbf{0}, r)\right)} = \frac{1}{n+2}V_n^{-\frac{2}{n}}, \tag{4}$$

which approaches $1/(2\pi e)$ from above as $n \to \infty$. Thus, the NSM of any lattice in any dimension satisfies $G(\Lambda) \ge 1/(2\pi e)$.

We define the effective radius $r_{\text{eff}}(\Lambda)$ as the radius of a ball which has the same volume as Λ, i.e.,

$$r_{\text{eff}}^2(\Lambda) \triangleq \frac{V^{\frac{2}{n}}(\Lambda)}{V_n^{\frac{2}{n}}}. \tag{5}$$

Since $\mathcal{B}(\mathbf{0}, r_{\text{eff}}(\Lambda))$ has the smallest second moment of all sets in $\mathbb{R}^n$ with volume $V(\Lambda)$, we have

$$\sigma^2(\Lambda) \ge \sigma^2\left(\mathcal{B}(\mathbf{0}, r_{\text{eff}}(\Lambda))\right) = \frac{r_{\text{eff}}^2(\Lambda)}{n+2}. \tag{6}$$

Thus,

$$r_{\text{eff}}(\Lambda) \le \sqrt{(n+2)\sigma^2(\Lambda)}. \tag{7}$$

Note that for large n we have

$$\frac{r_{\text{eff}}^2(\Lambda)}{n} \approx \frac{V^{\frac{2}{n}}(\Lambda)}{2\pi e}.$$

Definition 2: We say that a sequence in n of random noise vectors $\mathbf{Z}^{(n)}$ of length n with (finite) effective variance $\sigma_{\mathbf{Z}}^2 \triangleq \frac{1}{n}\mathbb{E}\|\mathbf{Z}^{(n)}\|^2$, is *semi norm-ergodic* if for any $\epsilon, \delta > 0$ and n large enough

$$\Pr\left(\mathbf{Z}^{(n)} \notin \mathcal{B}(\mathbf{0}, \sqrt{(1+\delta)n\sigma_{\mathbf{Z}}^2}\right) \le \epsilon. \tag{8}$$

Note that by the law of large numbers, any i.i.d. noise is semi norm-ergodic. However, even for non i.i.d. noise, the requirement (8) is not very restrictive. We will show in Lemma 3 that a vector $\mathbf{U}$ uniformly distributed over the Voronoi region of a lattice that is goof for MSE quantization is semi norm- ergodic. In the sequel we omit the dimension index, and denote the sequence $\mathbf{Z}^{(n)}$ simply by $\mathbf{Z}$.

Definition 3: The *nearest neighbor decoder* with respect to the lattice Λ outputs for every $\mathbf{y} \in \mathbb{R}^n$ the lattice point $Q_\Lambda(\mathbf{y})$.

The following definition is analogous to [10, Definition 7.7.1], where the only difference is that we do not restrict ourselves to AWGN, and consider the more general family of semi norm-ergodic noise.

Definition 4: A sequence of lattices $\Lambda^{(n)}$ with growing dimension, satisfying

$$\lim_{n\to\infty} V^{\frac{2}{n}}(\Lambda^{(n)}) = \Phi$$

for some $\Phi > 0$, is called *good for channel coding in the presence of semi norm-ergodic noise* if for any lattice point $\mathbf{t} \in \Lambda^{(n)}$, and additive semi norm-ergodic noise $\mathbf{Z}$ with effective variance[1] $\sigma_{\mathbf{Z}}^2 = \frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2 < \Phi/2\pi e$

$$\lim_{n\to\infty} \Pr\left(Q_{\Lambda^{(n)}}(\mathbf{t} + \mathbf{Z}) \ne \mathbf{t}\right) = 0,$$

That is, the error probability under nearest neighbor decoding in the presence of semi norm-ergodic additive noise $\mathbf{Z}$ vanishes with n if $\lim_{n\to\infty} r_{\text{eff}}^2(\Lambda^{(n)})/n > \sigma_{\mathbf{Z}}^2$. For brevity, we simply call such sequences of lattices *good for coding* in the sequel.

[1] In [14] the volume-to-noise ratio (VNR) was defined as

$$\mu = \lim_{n\to\infty} V^{\frac{2}{n}}(\Lambda^{(n)})/2\pi e\sigma_{\mathbf{Z}}^2.$$

Thus, the condition $\Phi > 2\pi e\sigma_{\mathbf{Z}}^2$ is equivalent to VNR > 1.

Definition 5: A sequence of lattices $\Lambda^{(n)}$ with growing dimension is called good for mean squared error (MSE) quantization if

$$\lim_{n\to\infty} G\left(\Lambda^{(n)}\right) = \frac{1}{2\pi e}.$$

A lattice Λ_c is said to be nested in Λ_f if $\Lambda_c \subset \Lambda_f$. The lattice Λ_c is referred to as the coarse lattice and Λ_f as the fine lattice. The *nesting ratio* is defined as $\left(V(\Lambda_c)/V(\Lambda_f)\right)^{1/n}$.

Next, we define "good" pairs of nested lattices. Our definition for the "goodness" of nested lattice pairs is different from the one used in [14].

Definition 6: A sequence of pairs of nested lattices $\Lambda_c^{(n)} \subset \Lambda_f^{(n)}$ is called "good" if the sequence of lattices $\Lambda_c^{(n)}$ and $\Lambda_f^{(n)}$ are good for both MSE quantization and for coding.

Remark 1: As we shall see in Section III, for the problem of coding over the AWGN channel (or more generally, any additive semi norm-ergodic noise channel), it suffices that $\Lambda_f^{(n)}$ is good for coding and $\Lambda_c^{(n)}$ is good for MSE quantization. In order to achieve the optimal rate-distortion function of a white Gaussian source, the roles are reversed and $\Lambda_f^{(n)}$ should be good for MSE quantization while $\Lambda_c^{(n)}$ is good for coding. A sequence of pairs $\Lambda_c^{(n)} \subset \Lambda_f^{(n)}$ that is good according to Definition 6 is therefore adequate for both problems.

Our existence proofs are based on Construction A [26], as defined next.

Definition 7 (p-Ary Construction A): Let p be a prime number, and let $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$ be a $k \times n$ matrix whose entries are all members of the finite field $\mathbb{Z}_p$. The matrix $\mathbf{G}$ generates a linear p-ary code

$$\mathcal{C}(\mathbf{G}) \triangleq \left\{\mathbf{x} \in \mathbb{Z}_p^n \ : \ \mathbf{x} = [\mathbf{w}^T\mathbf{G}] \bmod p \quad \mathbf{w} \in \mathbb{Z}_p^k\right\}.$$

The p-ary Construction A lattice induced by the matrix $\mathbf{G}$ is defined as

$$\Lambda(\mathbf{G}) \triangleq p^{-1}\mathcal{C}(\mathbf{G}) + \mathbb{Z}^n.$$

III. Main Results

Our main result is the following.

Theorem 1: For any finite L, $0 < \alpha_1 < \ldots < \alpha_L < \infty$, there exists a sequence of nested lattice chains $\Lambda_1^{(n)} \subset \cdots \subset \Lambda_L^{(n)}$ for which

1) $\Lambda_\ell^{(n)}$ is good for MSE quantization and for coding for all $\ell = 1, . . , L$;
2) $\lim_{n\to\infty} V^{\frac{2}{n}}\left(\Lambda_\ell^{(n)}\right) = 2\pi e 2^{-\alpha_\ell}$ for all $\ell = 1, \ldots, L$.

For the proof of Theorem 1, as given in Section IV, we define a novel ensemble of nested lattice chains. This ensemble is defined in Section IV and is based on drawing a random linear p-ary code and using Construction A to lift L of its sub-codes to the Euclidean space. Theorem 6, stated in Section IV and proved in Section V, shows that with high probability each of these lifted sub-codes possesses the goodness properties. The existence of a sequence of good nested lattice chains then follows from a simple union bound argument.

An immediate corollary of Theorem 1 is the following.

Theorem 2: For any $P_1 > P_2 > \cdots > P_L > 0$ there exists a sequence of nested lattice chains $\Lambda_1^{(n)} \subset \cdots \subset \Lambda_L^{(n)}$ with the following properties

1) $\Lambda_\ell^{(n)}$ is good for MSE quantization and for coding for all $\ell = 1, \ldots, L$;
2) $\lim_{n\to\infty} \sigma^2\left(\Lambda_\ell^{(n)}\right) = P_\ell$ for all $\ell = 1, \ldots, L$;
3) For any $1 \le k < m \le L$ the sequence of nested lattice codebooks $\mathcal{L}_{km}^{(n)} \triangleq \Lambda_m^{(n)} \cap \mathcal{V}_k^{(n)}$ has rate $R_{km}^{(n)} \triangleq \frac{1}{n}\log\left|\mathcal{L}_{km}^{(n)}\right|$ that satisfy[2]

$$\lim_{n\to\infty} R_{km}^{(n)} = \frac{1}{2}\log\left(\frac{P_k}{P_m}\right).$$

Proof: Fix $\alpha_1 > 0$ and, for any $1 < \ell \le L$, set $\alpha_\ell = \alpha_1 + \log\left(\frac{P_1}{P_\ell}\right)$. By Theorem 1 there exists a sequence $\Lambda_1^{(n)} \subset \cdots \subset \Lambda_L^{(n)}$, where all lattices are good for MSE quantization and for coding, and in addition, $\lim_{n\to\infty} V^{\frac{2}{n}}\left(\Lambda_\ell^{(n)}\right) = 2\pi e 2^{-\alpha_1}\left(\frac{P_\ell}{P_1}\right)$. Scaling all lattices in the sequence by $P_1 2^{\alpha_1}$, we get a sequence of lattices that are good for MSE quantization and coding for which $\lim_{n\to\infty} V^{\frac{2}{n}}\left(\Lambda_\ell^{(n)}\right) = 2\pi e P_\ell$. Since $\sigma^2(\Lambda) = G(\Lambda)V^{\frac{2}{n}}(\Lambda)$, the above implies that $\lim_{n\to\infty}\sigma^2\left(\Lambda_\ell^{(n)}\right) = P_\ell$ for all ℓ. In addition,

$$\begin{aligned}\lim_{n\to\infty} R_{km}^{(n)} &= \frac{1}{2}\log\left(\frac{\lim_{n\to\infty} V^{\frac{2}{n}}(\Lambda_k^{(n)})}{\lim_{n\to\infty} V^{\frac{2}{n}}(\Lambda_m^{(n)})}\right)\\ &= \frac{1}{2}\log\left(\frac{P_k}{P_m}\right),\end{aligned}$$

as desired. ∎

It is important to note that Theorem 1 and Theorem 2 can be obtained as a special case of the more general results proved in [18], [24], and [27]. These results showed the existence of chains of nested lattices where all latices in the chain are both good for coding and good for covering. Goodness for covering implies goodness for MSE quantization [2], [10], and is therefore a stronger property. However the existence proofs of such chains are quite complicated, and are not self-contained. In particular, these proofs involve starting with a lattice that is good for covering, whose existence is difficult to establish, and rotating a random Construction A lattice using it. Our main contribution in this paper is in providing a relatively simple, and self-contained proof for Theorem 1, from first principles.

In [14] it was shown that if Λ is good for covering, $\mathbf{U}$ is an independent random vector uniformly distributed over the Voronoi region of Λ, and $\mathbf{Z}$ is AWGN with variance σ^2, then a linear combination $\alpha\mathbf{Z} + \beta\mathbf{U}$ is close in distribution to an AWGN with variance $\alpha^2\sigma^2 + \beta^2\sigma^2(\Lambda)$. This property played an important role in the analysis of the AWGN capacity achieving nested lattice scheme of [14], namely, the mod-Λ scheme.

In order to show that pairs of nested lattices that are good according to Definition 6 achieve the AWGN capacity under

[2]All logarithms in this paper are to the base 2, and therefore all rates are expressed in bits per (real) channel use.

the mod-Λ coding scheme introduced in [14], we need the following theorem that states that any linear combination of semi norm-ergodic noise and a dither from a lattice that is good for MSE quantization is itself semi norm-ergodic.

Theorem 3: Let $\mathbf{Z} = \alpha\mathbf{N} + \beta\mathbf{U}$, where $\alpha, \beta \in \mathbb{R}$, $\mathbf{N}$ is semi norm-ergodic noise, and $\mathbf{U}$ is a dither statistically independent of $\mathbf{N}$, uniformly distributed over the Voronoi region $\mathcal{V}$ of a lattice Λ that is good for MSE quantization. Then, the random vector $\mathbf{Z}$ is semi norm-ergodic.

The proof is given in Section VI. In [14] it was shown that a nested lattice codebook $\mathcal{L} = \Lambda_f \cap \mathcal{V}_c$, based on a pair $\Lambda_c \subset \Lambda_f$ where both lattices are good for covering and Poltyrev good can achieve the capacity (as well as the Poltyrev error exponent) of the AWGN channel under the mod-Λ scheme. Theorem 4, stated below, shows that the capacity result continues to hold even if the two lattices $\Lambda_c \subset \Lambda_f$ are only good for MSE quantization and for coding, i.e., good according to Definition 6. The existence of such good nested lattice pairs is guaranteed by Theorem 2. Theorem 4 further extends the main result of [14] to any additive semi norm-ergodic noise channel.

Theorem 4: Consider an additive noise channel $Y = X+N$, where N is a semi norm-ergodic noise process with effective variance $\sigma^2_{\mathbf{N}} = 1$ and the input is subject to the power constraint $\frac{1}{n}\|\mathbf{X}\|^2 < \mathsf{SNR}$. For any $R < \frac{1}{2}\log(1+\mathsf{SNR})$ there exists a sequence of nested lattice codebooks $\mathcal{L}^{(n)} = \Lambda_f^{(n)} \cap \mathcal{V}_c^{(n)}$ based on a sequence of good nested lattice pairs $\Lambda_c^{(n)} \subset \Lambda_f^{(n)}$, whose rate approaches R and attains a vanishing error probability under the mod-Λ scheme.

Proof: Fix $0 < \epsilon < 1$ and let $\Lambda_c^{(n)} \subset \Lambda_f^{(n)}$ be a sequence of good nested lattice pairs with

$$\lim_{n\to\infty} \sigma^2\left(\Lambda_c^{(n)}\right) = (1-\epsilon)\mathsf{SNR},$$
$$\lim_{n\to\infty} \sigma^2\left(\Lambda_f^{(n)}\right) = (1+\epsilon)\tfrac{\mathsf{SNR}}{1+\mathsf{SNR}},$$

such that the rate of the sequence of codebooks $\mathcal{L}^{(n)} = \Lambda_f^{(n)} \cap \mathcal{V}_c^{(n)}$ satisfies

$$\lim_{n\to\infty} R^{(n)} = \frac{1}{2}\log\left(\frac{1-\epsilon}{1+\epsilon}(1+\mathsf{SNR})\right).$$

The existence of such a sequence of nested lattice pairs is guaranteed by Theorem 2. For brevity, we omit the sequence superscripts in the remainder of the proof, and simply use $\Lambda_c, \mathcal{V}_c, \Lambda_f, \mathcal{L}$ and R.

Next, apply the mod-Λ scheme of [14] with the codebook $\mathcal{L}$. Let $\mathbf{U}$ is a random dither statistically independent of $\mathbf{t}$, known to both the transmitter and the receiver, uniformly distributed over $\mathcal{V}_c$. Each of the 2^{nR} messages is mapped to a codeword in $\mathcal{L}$. To send the message w, corresponding to the codeword $\mathbf{t} \in \mathcal{L}$, the encoder transmits

$$\mathbf{X} = [\mathbf{t} - \mathbf{U}] \bmod \Lambda_c,$$

if $\frac{1}{n}\|\mathbf{X}\|^2 \le \mathsf{SNR}$, and the all-zeros vector otherwise. Due to the Crypto Lemma [14, Lemma 1], $\mathbf{X}$ is also uniformly distributed over $\mathcal{V}_c$ and is statistically independent of $\mathbf{t}$. Since Λ_c is good for MSE quantization, by Theorem 3, we have that $\mathbf{X}$ is semi norm-ergodic. Using this fact, and recalling that $\frac{1}{n}\mathbb{E}\|\mathbf{X}\|^2 = \sigma^2(\Lambda_c) = (1-\epsilon)\mathsf{SNR}$, it follows that $\frac{1}{n}\|\mathbf{X}\|^2 \le \mathsf{SNR}$ with high probability. Thus, the additional error probability incurred by replacing $\mathbf{X}$ (whenever necessary) with the zero-codeword vanishes with n.

The receiver scales its observation by a factor $\alpha > 0$ to be specified later, adds back the dither $\mathbf{U}$ and reduces the result modulo the coarse lattice

$$\begin{aligned}\mathbf{Y}_{\text{eff}} &= [\alpha\mathbf{Y} + \mathbf{U}] \bmod \Lambda_c \\ &= [\mathbf{X} + \mathbf{U} + (\alpha-1)\mathbf{X} + \alpha\mathbf{N}] \bmod \Lambda_c \\ &= [\mathbf{t} + (\alpha-1)\mathbf{X} + \alpha\mathbf{N}] \bmod \Lambda_c \\ &= [\mathbf{t} + \mathbf{Z}_{\text{eff}}] \bmod \Lambda_c, \end{aligned} \tag{9}$$

where

$$\mathbf{Z}_{\text{eff}} = (\alpha-1)\mathbf{X} + \alpha\mathbf{N} \tag{10}$$

is effective noise, that is statistically independent of $\mathbf{t}$, with effective variance

$$\sigma^2_{\text{eff}}(\alpha) \triangleq \frac{1}{n}\mathbb{E}\|\mathbf{Z}_{\text{eff}}\|^2 < \alpha^2 + (1-\alpha)^2\mathsf{SNR}. \tag{11}$$

Since $\mathbf{N}$ is semi norm-ergodic, and $\mathbf{X}$ is uniformly distributed[3] over the Voronoi region of a lattice that is good for MSE quantization, Theorem 3 implies that $\mathbf{Z}_{\text{eff}}$ is semi norm-ergodic with effective variance $\sigma^2_{\text{eff}}(\alpha)$. Setting $\alpha = \mathsf{SNR}/(1+\mathsf{SNR})$, such as to minimize the r.h.s. of (11) results in effective variance $\sigma^2_{\text{eff}} < \mathsf{SNR}/(1+\mathsf{SNR})$.

The receiver next computes

$$\begin{aligned}\hat{\mathbf{t}} &= \left[Q_{\Lambda_f}(\mathbf{Y}_{\text{eff}})\right] \bmod \Lambda_c \\ &= \left[Q_{\Lambda_f}([\mathbf{t} + \mathbf{Z}_{\text{eff}}] \bmod \Lambda_c)\right] \bmod \Lambda_c \\ &= \left[Q_{\Lambda_f}(\mathbf{t} + \mathbf{Z}_{\text{eff}})\right] \bmod \Lambda_c, \end{aligned} \tag{12}$$

and outputs the message corresponding to $\hat{\mathbf{t}}$ as its estimate. Since Λ_f is good for coding, $\mathbf{Z}_{\text{eff}}$ is semi norm-ergodic, and

$$\lim_{n\to\infty} \frac{V^{\frac{2}{n}}(\Lambda_f)}{2\pi e} = (1+\epsilon)\frac{\mathsf{SNR}}{1+\mathsf{SNR}} > \sigma^2_{\text{eff}},$$

we have that $\Pr(\hat{\mathbf{t}} \neq \mathbf{t}) \to 0$ as $n \to \infty$. Taking $\epsilon \to 0$ completes the proof. ∎

Remark 2: We remark that Theorem 4 is analogous to the results of [23] where it is shown that a Gaussian i.i.d. codebook ensemble with nearest neighbor decoding can attain any rate smaller than $\frac{1}{2}\log(1+\mathsf{SNR})$ over an additive semi norm-ergodic noise channel. Our result show that the same rate can be attained using nested lattice codes and the mod-Λ scheme.

Remark 3: We have shown that nested lattice pairs that are good according to Definition 6 suffice to achieve the capacity of the AWGN channel. Similarly, it can be shown that such pairs can attain the optimal rate-distortion tradeoff for the a Gaussian source, as well as the optimal rate-distortion trade-off for the Wyner-Ziv problem, under the scheme from [16] and [25].

Remark 4: In certain applications, chains of nested lattice codes are used in order to convert a Gaussian multiple access

[3] Strictly speaking, $\mathbf{X}$ is not uniform over $\mathcal{V}_c$ due to the possible replacement by the zero-vector. However, since this event has a vanishing probability, it cannot significantly increase $\Pr(\|(\alpha-1)\mathbf{X} + \alpha\mathbf{N}\|^2 < (1+\delta)n\sigma^2_{\text{eff}}(\alpha))$.

channel (MAC) into an effective modulo-lattice channel whose output is a fine lattice point plus effective noise reduced modulo a coarse lattice. Such a situation arises for example in the compute-and-forward framework [18], where a receiver is interested in decoding linear combinations with integer valued coefficients of the codewords transmitted by the different users of the MAC. In such applications, the effective noise is often a linear combination of AWGN and *multiple* statistically independent dithers uniformly distributed over the Voronoi region of the coarse lattice. Corollary 2, stated in Section VI, shows that such an effective noise is semi norm-ergodic regardless of the number of dithers contributing to it, as long as they are all independent and are induced by lattices that are good for MSE quantization. Consequently, nested lattice chains where all lattices are good for MSE quantization and coding, whose existence is guaranteed by Theorem 2, suffice to recover all results from [17]–[22] as well as many other achievable rate regions based on nested lattice coding schemes. Moreover, the analysis in the proof of Theorem 4 assumes that the additive noise is semi norm-ergodic, and not necessarily AWGN. Consequently, using a similar analysis it is possible to extend all the results from [17]–[22] to networks with any semi norm-ergodic additive noise.

Remark 5: The mod-Λ scheme uses common randomness in the form of a random dither vector $\mathbf{U}$, which is known to the encoder and the decoder. A consequence of the use of dither is that the effective noise $\mathbf{Z}_{\text{eff}}$ is statistically independent of the transmitted point $\mathbf{t}$, and therefore the decoding error probability does not depend on the message w that was chosen. Standard arguments show that the random dither can be replaced with a *fixed* one, without degrading the average error probability of the resulting codebook. However, without common randomness the error probability will not be the same for all messages. In [28] it was shown that if $\mathsf{SNR} > 1$, then the mod-Λ scheme can attain the AWGN channel capacity even without using a dither (i.e., $\mathbf{U} = \mathbf{0}$), and in [29] it was shown that a scheme based on a single lattice and probabilistic shaping can attain the AWGN capacity without dithering for all $\mathsf{SNR} > e$. In both cases, however, the error probability is message-dependant.

As evident from the proof of Theorem 4, the main role of the coarse lattice Λ_c in the mod-Λ scheme is to perform shaping. More specifically, the input to the channel is uniformly distributed on $\mathcal{V}_c$ and in order to approach capacity, such distribution must approach an AWGN as the dimension grows.

In practice, shaping is often avoided in order to reduce the implementation complexity. However, one can always use a nested lattice codebook where the coarse lattice is the simple one-dimensional cubic (integer) lattice, which is of course, not good for MSE quantization. In fact, many practical communication systems apply a p-ary linear code, e.g. turbo or LDPC, mapped to a PAM/QAM constellation. The induced constellation in the Euclidean space can be thought of as a nested lattice codebook $\gamma\,\Lambda_f \cap \gamma\,\mathcal{V}_c$, where Λ_f is a Construction A lattice based on the chosen linear code, whereas Λ_c is the integer lattice $\mathbb{Z}^n$.

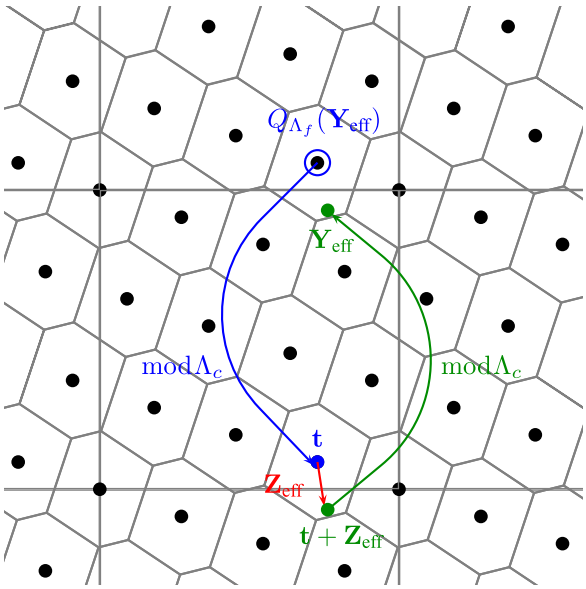

Fig. 1. An illustration of the coset nearest neighbor decoding process. The lattice point $\mathbf{t}$ was transmitted. The output of the induced channel when the mod-Λ transmission scheme is applied is $\mathbf{Y}_{\text{eff}} = [\mathbf{t} + \mathbf{Z}_{\text{eff}}] \bmod \gamma\,\mathbb{Z}^n$. The decoder quantizes $\mathbf{Y}_{\text{eff}}$ to the nearest lattice point in Λ and reduces the quantized output modulo $\gamma\,\mathbb{Z}^n$.

The scaling parameter γ, in this case, is dictated by the power constraint. For example, if the power constraint is $\mathbb{E}(X^2) \le \mathsf{SNR}$ the scaling parameter would be $\gamma = \sqrt{12\mathsf{SNR}}$. Since $\gamma\,\Lambda_c = \gamma\,\mathbb{Z}^n \subset \gamma\,\Lambda_f$, the minimum distance in $\gamma\,\Lambda_f$ cannot exceed $\sqrt{12\mathsf{SNR}}$, and in particular does not grow with the dimension. Thus, $\Pr(Q_{\gamma\,\Lambda_f}(\mathbf{t} + \mathbf{Z}_{\text{eff}}) \neq \mathbf{t})$ cannot vanish with the lattice dimension, and consequently $\gamma\,\Lambda_f$ is not good for coding.[4]

Nevertheless, as evident from (12), an error occurs if and only if the lattice point $Q_{\gamma\,\Lambda_f}(\mathbf{t}+\mathbf{Z}_{\text{eff}})$ is not in the same coset of $\gamma\,\Lambda_f/\gamma\,\Lambda_c$ as $\mathbf{t}$.[5]

Definition 8: The *coset nearest neighbor decoder* with respect to the nested lattice pair $\Lambda_c \subset \Lambda_f$ outputs for every $\mathbf{y} \in \mathbb{R}^n$ the lattice point $\left[Q_{\Lambda_f}(\mathbf{y})\right] \bmod \Lambda_c$.

It follows that the mod-Λ scheme succeeds if the coset nearest neighbor decoder finds the correct coset. For the case where the coarse lattice is $\gamma\,\mathbb{Z}^n$ this corresponds to $Q_{\gamma\,\Lambda_f}(\mathbf{t} + \mathbf{Z}_{\text{eff}}) = \mathbf{t} \bmod \gamma$. See Figure 1 for an illustration. Note that under coset nearest neighbor decoding, the aforementioned pairs of points in $\gamma\,\Lambda_f$, whose distance is γ, do not incur an error. Thus, it may be possible to attain an error probability that vanishes with the dimension using the mod-Λ scheme.

The next theorem, proved in Section VII, shows that this is indeed the case. More specifically, it shows that the mod-Λ scheme with nested lattice codes where $\Lambda_c = \sqrt{12\mathsf{SNR}}\mathbb{Z}^n$ can attain any rate smaller than $\frac{1}{2}\log(1+\mathsf{SNR}) - \frac{1}{2}\log(2\pi e/12)$ with a vanishing error probability, if p is large. For finite p, an explicit upper bound on the additional loss is also specified.

[4]In the next section we specify the ensemble of nested lattices used for the proof of Theorem 1, in which γ grows as $\sqrt{n}$ in order to avoid this problem.

[5]The coset of $\gamma\,\Lambda_c$ to which $\mathbf{t}$ belongs is the discrete set of points $\mathbf{t} + \gamma\,\Lambda_c$.

Let CUBE $\triangleq [-1/2, 1/2)^n$ denote the unit cube centered at the origin.

Theorem 5: Consider an additive noise channel $Y = X+N$, where N is an i.i.d. noise process with unit variance and the input is subject to the power constraint $\frac{1}{n}\mathbb{E}||\mathbf{X}||^2 \leq \mathsf{SNR}$. Let

$$\Gamma(p, \mathsf{SNR}) \triangleq \log\left(1 + \sqrt{\frac{3\mathsf{SNR}}{p^2}}\right).$$

For any

$$R < \frac{1}{2}\log(1+\mathsf{SNR}) - \frac{1}{2}\log\left(\frac{2\pi e}{12}\right) - \Gamma(p, \mathsf{SNR}),$$

there exists a sequence of nested lattice codebooks $\mathcal{L}^{(n)} = \Lambda_f^{(n)} \cap \sqrt{12\mathsf{SNR}} \cdot \mathrm{CUBE}$ with rate R, where $\Lambda_f^{(n)}$ is a sequence of scaled p-ary Construction A lattices, that attains a vanishing error probability under the mod-Λ scheme.

Note that $\Gamma(p, \mathsf{SNR}) \to 0$ as $p \to \infty$, and the gap to capacity is in this case just the standard shaping loss of $\frac{1}{2}\log(2\pi e/12)$. We further note that for any $\epsilon > 0$ the choice

$$\log p > \frac{1}{2}\log(\mathsf{SNR}) + \frac{1}{2}\log(3) - \frac{1}{2}\log\left(2^{\epsilon} - 1\right), \quad (13)$$

guarantees that $\Gamma(p, \mathsf{SNR}) < \epsilon$.

IV. An Ensemble for Nested Lattice Chains

Previous proofs for the existence of capacity achieving pairs of nested lattices used random Construction A, introduced by Loeliger [8], for creating a fine lattice, and then rotated it using a lattice that is good for covering. Here, we take a different approach that is a direct extension of the original approach of [25] to creating nested binary linear codes. We use random Construction A to simultaneously create both the fine and the coarse lattice. Namely, we randomly draw a linear code and lift it to the Euclidean space in order to obtain the fine lattice. The coarse lattice is obtained by lifting a subcode from the same linear code to the Euclidean space.

Let $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$. For any natural number $m \leq k$ we denote by $\mathbf{G}_m$ the $m \times n$ matrix obtained by taking only the first m rows of $\mathbf{G}$. The linear code $\mathcal{C}(\mathbf{G}_m)$ and the lattice $\Lambda(\mathbf{G}_m)$ are defined as in Definition 7.

Clearly, for any $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$, and $k_1 < k$ we have that $\Lambda\left(\mathbf{G}_{k_1}\right) \subset \Lambda(\mathbf{G}_k)$. Thus, we can define an ensemble of nested lattice pairs by fixing k_1, k, n, p and drawing the entries of the matrix $\mathbf{G}$ according to the i.i.d. uniform distribution on $\mathbb{Z}_p$.

Remark 6: We have chosen to specify our ensemble in terms of the linear codes' generating matrices

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{k_1} \\ - - - \\ \mathbf{G}' \end{bmatrix}.$$

We could have equally defined the ensemble using the linear codes' parity check matrices

$$\mathbf{H}_{n-k_1} = \begin{bmatrix} \mathbf{H}_{n-k} \\ - - - \\ \mathbf{H}' \end{bmatrix},$$

as done in [16] and [25] for ensembles of nested binary linear codes.

More generally, for any choice of L natural numbers $k_1 < k_2 < \cdots < k_L < n$ we can define a similar ensemble for a chain of L nested lattices

$$\Lambda\left(\mathbf{G}_{k_1}\right) \subset \Lambda\left(\mathbf{G}_{k_2}\right) \subset \cdots \subset \Lambda\left(\mathbf{G}_{k_L}\right).$$

We now formally define the ensemble of nested lattices we will use in our existence proof.

Definition 9 (Ensemble of Nested Lattice Chains): Let n be a natural number and $0 < \alpha_1 < \ldots < \alpha_L < \log n$. An $(n, \alpha_1, \ldots, \alpha_L)$ ensemble for a chain of L nested lattices is defined as follows. Let $\gamma = 2\sqrt{n}$, and $p = \xi n^{\frac{3}{2}}$, where ξ is chosen as the smallest number in the interval $[1, 2)$ such that p is prime. Let

$$k_\ell \triangleq \frac{n}{2\log p}\left(\log\left(\frac{4}{V_n^{\frac{2}{n}}}\right) + \alpha_\ell\right), \quad \ell = 1, \ldots, L.$$

Draw a matrix $\mathbf{G} \in \mathbb{Z}_p^{k_L\times n}$ whose entries are i.i.d. uniformly distributed over $\mathbb{Z}_p$, and construct the chain $\Lambda_1 \subset \cdots \subset \Lambda_L$ by setting

$$\Lambda_\ell = \gamma\,\Lambda\left(\mathbf{G}_{k_\ell}\right), \quad \ell = 1, \ldots, L.$$

Theorem 1 will follow as a straightforward corollary of the following result.

Theorem 6: Let n be a large natural number, $\gamma = 2\sqrt{n}$, and $p = \xi n^{\frac{3}{2}}$, where ξ is chosen as the smallest number in the interval $[1, 2)$ such that p is prime. Further, let $0 < \alpha < \log n$ and set

$$k \triangleq \frac{n}{2\log p}\left(\log\left(\frac{4}{V_n^{\frac{2}{n}}}\right) + \alpha\right). \quad (14)$$

Let $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$ be a random matrix whose entries are i.i.d. uniformly distributed over $\mathbb{Z}_p$. For any $\epsilon, \delta > 0$, there is an integer $N(\epsilon, \delta)$ such that if $n > N(\epsilon, \delta)$

1) $\Pr\left(\mathrm{rank}(\mathbf{G}) < k\right) < \epsilon$;
2) $\Pr\left(\sigma^2\left(\gamma\,\Lambda(\mathbf{G})\right) > (1+\delta)2^{-\alpha}\right) < \epsilon$;
3) For any additive semi norm-ergodic noise $\mathbf{Z}$ with effective variance $\sigma_{\mathbf{Z}}^2 = \frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2 \leq (1-\delta)2^{-\alpha}$ and any $\mathbf{t} \in \gamma\,\Lambda(\mathbf{G})$, the following holds

$$\Pr\left(\Pr\left(Q_{\gamma\,\Lambda(\mathbf{G})}(\mathbf{t}+\mathbf{Z}) \neq \mathbf{t} \mid \mathbf{G}\right) > \delta\right) < \epsilon.$$

Remark 7: No attempt was made to minimize the value of p that is used, and it was chosen as $\Theta\left(n^{\frac{3}{2}}\right)$ for computational and notational convenience within the proofs to follow. It can be easily verified that all our proceeding arguments remain valid for $p = \Theta\left(n^{\frac{1}{2}+\epsilon}\right)$ for any $\epsilon > 0$, if one restricts α to be smaller that $\log\log n$ (α dictates the range of rates our nested lattice ensemble can support, and therefore upper bounding it by any function of n that grows to infinity is sufficient).

The proof of Theorem 6 is given in Section V. We now prove Theorem 1.

Proof of Theorem 1: Let $\Lambda_1 \subset \cdots \subset \Lambda_L$ be a random lattice chain drawn from the $(n, \alpha_1, \ldots, \alpha_L)$ ensemble and let $\mathbf{G}_{k_1}, \ldots, \mathbf{G}_{k_L}$ be the corresponding linear codes generating

matrices. Set $\epsilon, \delta > 0$ and for all $\ell = 1, \ldots, L$ define the following error events

1) $E_{1\ell}$ is the event that rank $\left(\mathbf{G}_{k_\ell}\right) < k_\ell$;
2) $E_{2\ell}$ is the event that $\sigma^2(\Lambda_\ell) > (1+\delta)2^{-\alpha_\ell}$
3) $E_{3\ell}$ is the event that $\Pr\left(Q_{\Lambda_\ell}(\mathbf{t}+\mathbf{Z}) \neq \mathbf{t}\right) > \delta$ for some $\mathbf{t} \in \Lambda_\ell$ and some additive semi norm-ergodic noise $\mathbf{Z}$ with effective variance $\sigma^2_{\mathbf{Z}} = \frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2 \leq (1-\delta)2^{-\alpha}$.

Further, let

$$E \triangleq \bigcup_{i=1}^{3}\bigcup_{\ell=1}^{L} E_{i\ell}.$$

By the union bound we have that

$$\begin{aligned}\Pr(E) &\leq \sum_{i=1}^{3}\Pr\left(\bigcup_{\ell=1}^{L} E_{i\ell}\right)\\ &= \Pr(E_{1L}) + \Pr\left(\bigcup_{\ell=1}^{L} E_{2\ell}\right) + \Pr\left(\bigcup_{\ell=1}^{L} E_{3\ell}\right)\\ &\leq \Pr(E_{1L}) + \sum_{\ell=1}^{L}\Pr(E_{2\ell}) + \sum_{\ell=1}^{L}\Pr(E_{3\ell}), \qquad (15)\end{aligned}$$

where (15) follows from the fact that if $\mathbf{G}_{k_L}$ has full row rank over $\mathbb{Z}_p$, then so are all the matrices obtained by removing rows from it. Further, since $\mathbf{G}_{k_\ell}$ satisfies the conditions of Theorem 6 for all $\ell = 1, \ldots, L$, then for n large enough $\Pr(E_{1L}) < \epsilon$, $\Pr(E_{2\ell}) < \epsilon$ and $\Pr(E_{3\ell}) < \epsilon$. Thus, $\Pr(E) \leq (2L+1)\epsilon$, and consequently $\Pr(\overline{E}) > 1-(2L+1)\epsilon$, where $\overline{E}$ is the event that E did not occur. Since this holds for any $\epsilon > 0$, we have that for n large enough, the event E does not occur for almost all members in the ensemble.

We now show that any member in the ensemble for which E does not occur, has lattices $\Lambda_1 \subset \cdots \subset \Lambda_L$ whose volumes are close to $2\pi e 2^{-\alpha_\ell}$, whose normalized second moments are close to $1/2\pi e$, and whose error probabilities are small as long as the volume-to-noise ratio is greater than 1.

In particular, if E does not occur, all matrices $\mathbf{G}_{k_1}, \ldots, \mathbf{G}_{k_L}$ have full row rank over $\mathbb{Z}_p$. In this case, we have that $V(\Lambda_\ell) = \gamma^n p^{-k_\ell}$ and therefore

$$\begin{aligned}V^{\frac{2}{n}}(\Lambda_\ell) &= \gamma^2 p^{-\frac{2k_\ell}{n}}\\ &= 4n\frac{V_n^{\frac{2}{n}}}{4}2^{-\alpha_\ell}. \qquad (16)\end{aligned}$$

Since $\lim_{n\to\infty} nV_n^{\frac{2}{n}} = 2\pi e$, we have that $\lim_{n\to\infty} V^{\frac{2}{n}}(\Lambda_\ell) = 2\pi e 2^{-\alpha_\ell}$, as desired. In particular, for n large enough

$$(1-\delta/2)2\pi e 2^{-\alpha_\ell} < V^{\frac{2}{n}}(\Lambda_\ell) < 2\pi e 2^{-\alpha_\ell}. \qquad (17)$$

Now, by Theorem 6

$$\begin{aligned}G(\Lambda_\ell) &= \frac{\sigma^2(\Lambda_\ell)}{V^{\frac{2}{n}}(\Lambda_\ell)}\\ &\leq \frac{(1+\delta)2^{-\alpha_\ell}}{(1-\delta/2)2\pi e 2^{-\alpha_\ell}}\\ &= (1+\delta')\frac{1}{2\pi e},\end{aligned}$$

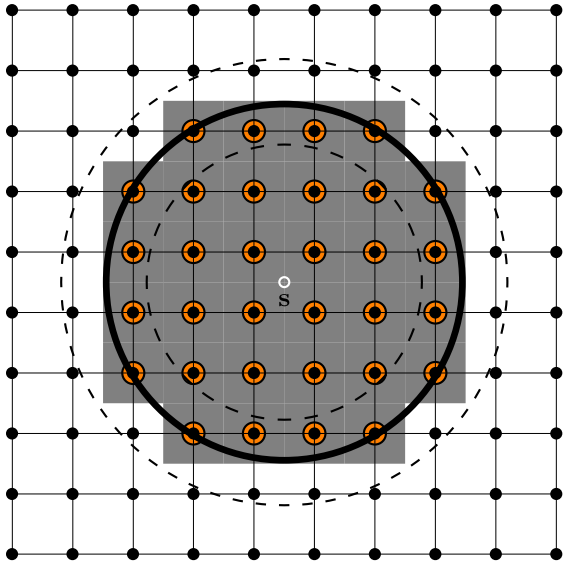

Fig. 2. An illustration of Lemma 1. The solid circle is the boundary of $\mathcal{B}(\mathbf{s}, r)$, and the points inside the small bright circles are the members of the set $\mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r)$. The set $= \mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r) + \text{CUBE}$ is the shaded area, and as the lemma indicates, it contains $\mathcal{B}(\mathbf{s}, r - \frac{\sqrt{n}}{2})$ and is contained in $\mathcal{B}(\mathbf{s}, r + \frac{\sqrt{n}}{2})$, whose boundaries are plotted in dashed circles.

where $\delta' = (1+\delta)/(1-\delta/2)$ can be made as small as desired by increasing n. Thus, the sequence $\Lambda_\ell^{(n)}$ is good for MSE quantization.

In addition (again by Theorem 6, part 3, and (17)), we have that for any semi norm-ergodic noise $\mathbf{Z}_\ell$ with effective variance $\sigma^2_{\mathbf{Z}_\ell} \leq (1-\delta)2^{-\alpha_\ell}$, the probability of error in nearest neighbor decoding is smaller than δ. Thus, for n large enough we have that as long as the ratio $V^{\frac{2}{n}}(\Lambda_\ell)/(2\pi e\sigma^2_{\mathbf{Z}_\ell})$ is greater than $(1-\delta/2)/(1-\delta)$, the error probability in decoding a point from Λ_ℓ in the presence of additive noise $\mathbf{Z}_\ell$ is smaller than δ. Thus, the sequence $\Lambda_\ell^{(n)}$ is good for coding. ■

V. Proof of Theorem 6

Before going into the proof we need to introduce some more notation. Denote the operation of reducing each component of $\mathbf{x} \in \mathbb{R}^n$ modulo γ by $\mathbf{x}^* \triangleq [\mathbf{x}] \bmod \gamma\mathbb{Z}^n$. If $\mathcal{S}$ is a set of points in $\mathbb{R}^n$, $\mathcal{S}^*$ is the set obtained by reducing all points in $\mathcal{S}$ modulo $\gamma\mathbb{Z}^n$. If $\mathcal{S}$ and $\mathcal{T}$ are sets, $\mathcal{S}+\mathcal{T}$ is their Minkowski sum. In the sequel, we use the following lemma, which follows from simple geometric arguments and is illustrated in Figure 2.

Lemma 1: For any $\mathbf{s} \in \mathbb{R}^n$ and $r > 0$, the number of points of $\mathbb{Z}^n$ inside $\mathcal{B}(\mathbf{s}, r)$ can be bounded as

$$\begin{aligned}\left(\max\left\{r - \frac{\sqrt{n}}{2}, 0\right\}\right)^n V_n &\leq \left|\mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r)\right|\\ &\leq \left(r + \frac{\sqrt{n}}{2}\right)^n V_n.\end{aligned}$$

Proof: Let $\mathcal{S} \triangleq (\mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r)) + \text{CUBE}$, and note that $|\mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r)| = \text{Vol}(\mathcal{S})$. We have

$$\mathcal{B}\left(\mathbf{s}, r - \frac{\sqrt{n}}{2}\right) \subset \mathcal{S}. \qquad (18)$$

To see this, note that any $\mathbf{x} \in \mathcal{B}\left(\mathbf{s}, r-\frac{\sqrt{n}}{2}\right)$ lies inside $\mathbf{a}+\text{CUBE}$ for some $\mathbf{a} \in \mathbb{Z}^n$, and for this $\mathbf{a}$ the inequality $\|\mathbf{a}-\mathbf{x}\| \le \sqrt{n}/2$ holds. Applying the triangle inequality gives

$$\|\mathbf{a}-\mathbf{s}\| = \|(\mathbf{a}-\mathbf{x})+(\mathbf{x}-\mathbf{s})\| \le \|(\mathbf{a}-\mathbf{x})\| + \|(\mathbf{x}-\mathbf{s})\| \le r.$$

Thus, $\mathbf{a} \in (\mathbb{Z}^n \cap \mathcal{B}(\mathbf{s}, r))$, and hence $\mathbf{x} \in \mathcal{S}$, which implies (18). On the other hand,

$$\begin{aligned}\mathcal{S} &\subset \mathcal{B}(\mathbf{s}, r) + \text{CUBE} \\ &\subset \mathcal{B}(\mathbf{s}, r) + \mathcal{B}\left(0, \frac{\sqrt{n}}{2}\right) \\ &= \mathcal{B}\left(\mathbf{s}, r+\frac{\sqrt{n}}{2}\right).\end{aligned}$$

Thus,

$$\text{Vol}\left(\mathcal{B}\left(\mathbf{s}, r-\frac{\sqrt{n}}{2}\right)\right) \le \text{Vol}(\mathcal{S}) \le \text{Vol}\left(\mathcal{B}\left(\mathbf{s}, r+\frac{\sqrt{n}}{2}\right)\right).$$

■

A. The Matrix G Is Full Rank With High Probability

The probability that $\mathbf{G}$ is not full-rank was bounded in [30]. We repeat the proof for completeness. The matrix $\mathbf{G}$ is not full rank if and only if there exist some nonzero vector $\mathbf{w} \in \mathbb{Z}_p^k$ such that $\mathbf{w}^T\mathbf{G} = \mathbf{0}$. Thus,

$$\begin{aligned}\Pr(\text{rank}(\mathbf{G}) < k) &= \Pr\left(\bigcup_{\mathbf{w}\in\mathbb{Z}_p^k\setminus\mathbf{0}} (\mathbf{w}^T\mathbf{G} = \mathbf{0})\right) \\ &\le \sum_{\mathbf{w}\in\mathbb{Z}_p^k\setminus\mathbf{0}} \Pr(\mathbf{w}^T\mathbf{G} = \mathbf{0}) \qquad (19) \\ &= (p^k-1)p^{-n} \qquad (20) \\ &< p^{-(n-k)},\end{aligned}$$

where (19) follows from the union bound, and (20) since $\mathbf{w}^T\mathbf{G}$ is uniformly distributed over $\mathbb{Z}_p^n$ for any $\mathbf{w} \ne \mathbf{0}$.

By our definition of p and k, and using the fact that $V_n^{\frac{2}{n}} \ge \frac{4}{n}$ for all n, we have

$$\begin{aligned}k &\le \frac{n}{2\log\xi + 3\log n}(\log n + \alpha) \\ &\le n\left(\frac{1}{3} + \frac{\alpha}{3\log n}\right) \\ &< \frac{2n}{3},\end{aligned}$$

where the last inequality follows from the assumption $\alpha < \log n$. Thus, $\Pr(\text{rank}(\mathbf{G}) < k) < p^{-\frac{n}{3}}$, and can therefore be made smaller than any $\epsilon > 0$, by taking n large enough.

B. Goodness for MSE Quantization

In this subsection we show that for any $\delta, \epsilon > 0$ and n large enough

$$\Pr\left(\sigma^2(\gamma\,\Lambda(\mathbf{G})) > (1+\delta)2^{-\alpha}\right) < \epsilon.$$

Our proof follows the derivation from [31], which dealt with the NSM of Construction A lattices with finite p. In our case p grows with the lattice dimension, and the derivation can be significantly simplified.

We begin by bounding the average MSE distortion attained by the random lattice $\gamma\,\Lambda(\mathbf{G})$ for a source uniformly distributed over $\gamma\,[0,1)^n$. As we shall see, this average MSE distortion is equal to $\mathbb{E}(\sigma^2(\gamma\,\Lambda(\mathbf{G})))$. We then apply Markov's inequality to show that this implies that almost all lattices in the ensemble have a small $\sigma^2(\gamma\,\Lambda(\mathbf{G}))$.

For any (fixed) $\mathbf{x} \in \mathbb{R}^n$, define

$$\begin{aligned}d(\mathbf{x}, \gamma\,\Lambda(\mathbf{G})) &\triangleq \frac{1}{n}\min_{\lambda\in\gamma\,\Lambda(\mathbf{G})} \|\mathbf{x}-\lambda\|^2 \\ &= \frac{1}{n}\min_{\mathbf{a}\in\mathbb{Z}^n, \mathbf{c}\in\mathcal{C}(\mathbf{G})} \|\mathbf{x}-\gamma\,p^{-1}\mathbf{c}-\gamma\,\mathbf{a}\|^2 \\ &= \frac{1}{n}\min_{\mathbf{c}\in\mathcal{C}(\mathbf{G})} \|(\mathbf{x}-\gamma\,p^{-1}\mathbf{c})^*\|^2.\end{aligned}$$

Recall that $\gamma\,\mathbb{Z}^n \subset \gamma\,\Lambda(\mathbf{G})$ and therefore $d(\mathbf{x}, \gamma\,\Lambda(\mathbf{G})) \le \gamma^2/4$ for any $\mathbf{x} \in \mathbb{R}^n$, regardless of $\mathbf{G}$.

Let $0 < \rho < \alpha$. For any $\mathbf{w} \in \mathbb{Z}_p^k \setminus \mathbf{0}$, define the random vector $\mathbf{C}(\mathbf{w}) = [\mathbf{w}^T\mathbf{G}] \bmod p$, and note that $\mathbf{C}(\mathbf{w})$ is uniformly distributed over $\mathbb{Z}_p^n$. For all $\mathbf{w} \in \mathbb{Z}_p^k \setminus \mathbf{0}$ and $\mathbf{x} \in \mathbb{R}^n$, we have

$$\begin{aligned}\varepsilon &\triangleq \Pr\left(\frac{1}{n}\left\|\left(\mathbf{x}-\gamma\,p^{-1}\mathbf{C}(\mathbf{w})\right)^*\right\|^2 \le 2^{-\rho}\right) \\ &= p^{-n}\left|(\gamma\,p^{-1}\mathbb{Z}_p^n) \bigcap \mathcal{B}^*(\mathbf{x}, \sqrt{n2^{-\rho}})\right| \\ &= p^{-n}\left|(\gamma\,p^{-1}\mathbb{Z}^n) \bigcap \mathcal{B}(\mathbf{x}, \sqrt{n2^{-\rho}})\right| \qquad (21) \\ &\ge p^{-n}V_n\left(p\gamma^{-1}\sqrt{n2^{-\rho}} - \frac{\sqrt{n}}{2}\right)^n \qquad (22) \\ &= V_n(\gamma^{-2}n2^{-\rho})^{\frac{n}{2}}\left(1-\frac{\gamma\sqrt{2^\rho}}{2p}\right)^n \\ &= V_n\left(\frac{1}{4}\right)^{\frac{n}{2}}2^{-\frac{\rho n}{2}}\left(1-\frac{\sqrt{n2^\rho}}{p}\right)^n, \qquad (23)\end{aligned}$$

where (21) follows since $\gamma = 2\sqrt{n}$, and hence, for any two distinct points $\mathbf{b}_1, \mathbf{b}_2 \in \mathcal{B}(\mathbf{x}, \sqrt{n2^{-\rho}})$ we have $\mathbf{b}_1^* \ne \mathbf{b}_2^*$ (that is, the ball $\mathcal{B}(\mathbf{x}, \sqrt{n2^{-\rho}})$ is contained in a cube with side γ), and (22) follows from Lemma 1. Substituting $p = \xi n^{\frac{3}{2}}$ and recalling that $\rho < \alpha < \log n$ gives

$$\begin{aligned}\varepsilon &> 2^{-\frac{n}{2}\left(\log\left(\frac{4}{V_n^{n/2}}\right)+\rho\right)}\left(1-\frac{2^{\frac{\rho}{2}}}{\xi n}\right)^n \\ &= 2^{-\frac{n}{2}\left(\log\left(\frac{4}{V_n^{n/2}}\right)+\rho+o(1)\right)}, \qquad (24)\end{aligned}$$

where $o(1) \to 0$ as $n \to \infty$.

Let $M \triangleq p^k - 1$. Label each of the vectors $\mathbf{w} \in \mathbb{Z}_p^k \setminus \mathbf{0}$ by an index $i = 1, \ldots, M$, and refer to its corresponding codeword as $\mathbf{C}_i$. Define the indicator random variable related to the point $\mathbf{x} \in \mathbb{R}^n$

$$\chi_i = \begin{cases} 1 & \text{if } \frac{1}{n}\left|\left(\mathbf{x}-\gamma\,p^{-1}\mathbf{C}_i\right)^*\right|^2 \le 2^{-\rho} \\ 0 & \text{otherwise.}\end{cases}$$

Since each χ_i occurs with probability ε, we have

$$\begin{aligned}\Pr\left(\sum_{i=1}^{M}\chi_i=0\right)&=\Pr\left(\frac{1}{M}\sum_{i=1}^{M}\chi_i-\varepsilon=-\varepsilon\right)\\&\le\Pr\left(\left|\frac{1}{M}\sum_{i=1}^{M}\chi_i-\varepsilon\right|\ge\varepsilon\right)\\&\le\frac{\mathrm{Var}\left(\frac{1}{M}\sum_{i=1}^{M}\chi_i\right)}{\varepsilon^2},\end{aligned}\tag{25}$$

where the last inequality follows from Chebyshev's inequality. In order to further bound the variance term from (25), we note that $\mathbf{C}(\mathbf{w}_1)$ and $\mathbf{C}(\mathbf{w}_2)$ are statistically independent unless $\mathbf{w}_1=[a\mathbf{w}_2] \bmod p$ for some $a\in\mathbb{Z}_p$. Therefore, each χ_i is statistically independent of all but p different χ_j's. Thus,

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{M}\sum_{i=1}^{M}\chi_i\right)&=\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathrm{Cov}(\chi_i,\chi_j)\\&\le\frac{Mp\varepsilon}{M^2}.\end{aligned}$$

Substituting into (25) and using (24), we see that for any $\mathbf{x}\in\mathbb{R}^n$

$$\begin{aligned}\Pr\left(d(\mathbf{x},\gamma\,\Lambda(\mathbf{G}))>2^{-\rho}\right)&\le\Pr\left(\sum_{i=1}^{M}\chi_i=0\right)\\&<\frac{p}{M\varepsilon}\\&<2n^{\frac{3}{2}}\frac{1}{p^k-1}2^{\frac{n}{2}\left(\log\left(\frac{4}{V_n^{n/2}}\right)+\rho+o(1)\right)}\\&<p^{-k}2^{\frac{n}{2}\left(\log\left(\frac{4}{V_n^{n/2}}\right)+\rho+o(1)\right)}\\&=2^{-\frac{n}{2}(\alpha-\rho+o(1))},\end{aligned}$$

where we have used

$$p^{-k}=2^{-\frac{n}{2}\left(\log\left(\frac{4}{V_n^{n/2}}\right)+\alpha\right)}$$

in the last equality.

It follows that for any distribution on $\mathbf{X}$ we have

$$\begin{aligned}\mathbb{E}_{\mathbf{X},\mathbf{G}}\left(d(\mathbf{X},\gamma\,\Lambda(\mathbf{G}))\right)&\le 2^{-\rho}\Pr\left(d(\mathbf{X},\gamma\,\Lambda(\mathbf{G}))\le 2^{-\rho}\right)\\&\quad+\frac{\gamma^2}{4}\Pr\left(d(\mathbf{X},\gamma\,\Lambda(\mathbf{G}))>2^{-\rho}\right)\\&\le 2^{-\rho}\left(1+2^{-\frac{n}{2}(\alpha-\rho+o(1))}\right).\end{aligned}$$

Thus, for any $0<\rho<\alpha$ the upper bound on the distortion averaged over $\mathbf{X}$ and over the ensemble of lattices $\gamma\,\Lambda(\mathbf{G})$ becomes arbitrary close to $2^{-\rho}$ as n increases. Since this is true for all distributions on $\mathbf{X}$, we may take $\mathbf{X}\sim\mathrm{Unif}\,(\gamma\,[0,1)^n)$. Let $\mathbf{U}$ be a random variable uniformly distributed over the Voronoi region $\mathcal{V}_{\mathbf{G}}$ of a lattice $\gamma\,\Lambda(\mathbf{G})$ randomly drawn from the ensemble. By construction, for any lattice $\gamma\,\Lambda(\mathbf{G})$ in the defined ensemble $\left[\gamma\,p^{-1}\mathcal{C}(\mathbf{G})+\mathcal{V}_{\mathbf{G}}\right]^*=\gamma\,[0,1)^n$. Moreover, reducing the set $\gamma\,p^{-1}\mathcal{C}(\mathbf{G})+\mathcal{V}_{\mathbf{G}}$ modulo $\gamma\,\mathbb{Z}^n$ does not change its volume. Therefore,

$$\begin{aligned}\mathbb{E}_{\mathbf{G}}\left(\sigma^2(\gamma\,\Lambda(\mathbf{G}))\right)&=\mathbb{E}_{\mathbf{U},\mathbf{G}}\left(\frac{1}{n}\|\mathbf{U}\|^2\right)\\&=\mathbb{E}_{\mathbf{X},\mathbf{G}}\left(d(\mathbf{X},\gamma\,\Lambda(\mathbf{G}))\right).\end{aligned}$$

It follows that, for any $0<\rho<\alpha$,

$$\mathbb{E}_{\mathbf{G}}\left(\sigma^2(\gamma\,\Lambda(\mathbf{G}))\right)\le 2^{-\rho}\left(1+2^{-\frac{n}{2}(\alpha-\rho+o(1))}\right).$$

Now, define the random variable $T\triangleq\sigma^2(\gamma\,\Lambda(\mathbf{G}))-\frac{n}{n+2}2^{-\alpha}$. We show that the r.v. T is non-negative, or equivalently, that for every $\mathbf{G}$ in the ensemble

$$\sigma^2(\gamma\,\Lambda(\mathbf{G}))\ge\frac{n}{n+2}2^{-\alpha}.\tag{26}$$

To see this, note that $V(\gamma\,\Lambda(\mathbf{G}))\ge\gamma^n p^{-k}$ for all $\mathbf{G}$, with equality if and only if $\mathbf{G}$ has full row rank. Thus, by (16) we have $V^{\frac{2}{n}}(\gamma\,\Lambda(\mathbf{G}))\ge nV_n^{\frac{2}{n}}2^{-\alpha}$, which implies $r_{\mathrm{eff}}^2(\gamma\,\Lambda(\mathbf{G}))=V^{\frac{2}{n}}(\gamma\,\Lambda(\mathbf{G}))/V_n^{\frac{2}{n}}\ge n2^{-\alpha}$ by (5). Using the isoperimetric inequality (6), we get (26).

Since T is non-negative, we can apply Markov's inequality

$$\begin{aligned}\Pr(T>\delta 2^{-\alpha})&\le\frac{E(T)}{\delta}2^{\alpha}\\&=\frac{E(\sigma^2(\gamma\,\Lambda(\mathbf{G})))-\frac{n}{n+2}2^{-\alpha}}{\delta}2^{\alpha}\\&\le\frac{2^{\alpha-\rho}\left(1+2^{-\frac{n}{2}(\alpha-\rho+o(1))}\right)-\frac{n}{n+2}}{\delta}\end{aligned}$$

Setting $\rho=\alpha-\log\left(\frac{n}{n+2}+\frac{\epsilon\delta}{2}\right)$ we get that for n large enough $\Pr(T>\delta 2^{-\alpha})<\epsilon$, and therefore $\Pr\left(\sigma^2(\gamma\,\Lambda(\mathbf{G}))>(1+\delta)2^{-\alpha}\right)<\epsilon$ as desired.

C. Goodness for Coding

In this subsection we show that for any $\delta,\epsilon>0$, and any additive semi norm-ergodic noise $\mathbf{Z}$ with effective variance $\sigma_{\mathbf{Z}}^2=\frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2\le(1-\delta)2^{-\alpha}$, we have that

$$\Pr\left(\Pr\left(Q_{\gamma\,\Lambda(\mathbf{G})}(\mathbf{t}+\mathbf{Z})\ne\mathbf{t}\mid\mathbf{G}\right)>\delta\right)<\epsilon$$

for any $\mathbf{t}\in\gamma\,\Lambda(\mathbf{G})$, provided that n is large enough.

For any $\mathbf{G}$, we upper bound the error probability of the nearest neighbor decoder $Q_{\gamma\,\Lambda(\mathbf{G})}(\cdot)$ using the *bounded distance* decoder, which is inferior. More precisely, we analyze the performance of a decoder that finds all lattice points of $\gamma\,\Lambda(\mathbf{G})$ within Euclidean distance r from $\mathbf{t}+\mathbf{Z}$. If there is a unique codeword in this set, this is the decoded codeword. Otherwise, the decoder declares an error. It is easy to see that regardless of the choice of r, the nearest neighbor decoder makes the correct decision whenever the bounded distance decoder does. Therefore, the error probability of the nearest neighbor decoder is upper bounded by that of the bounded distance decoder.

Given $\mathbf{G}$, an error event E for the bounded distance decoder can be expressed as the union of three events:

1) E_1 - The noise vector $\mathbf{Z}$ falls outside a ball of radius r;

2) E_2 - The ball $\mathcal{B}(\mathbf{t}+\mathbf{Z},r)$ contains a point $\mathbf{t}+\gamma\,\mathbf{a}$ for some $\mathbf{a}\in\mathbb{Z}^n\setminus\mathbf{0}$. This is equivalent to the event $(\mathcal{B}(\mathbf{t}+\mathbf{Z},r)\cap(\mathbf{t}+\gamma\,\mathbb{Z}^n))\setminus\mathbf{t}\neq\emptyset$;
3) E_3 - The ball $\mathcal{B}(\mathbf{t}+\mathbf{Z},r)$ contains a point from $\gamma\,\Lambda(\mathbf{G})$ that does not belong to $\gamma\,\mathbb{Z}^n$. This is equivalent to the event $\mathcal{B}(\mathbf{t}+\mathbf{Z},r)\cap(\mathbf{t}+(\gamma\,\Lambda(\mathbf{G})\setminus\gamma\,\mathbb{Z}^n))\neq\emptyset$;

Note that the first two events E_1 and E_2 depend only on $\mathbf{Z}$, but not on $\mathbf{G}$. Moreover,

$$E_1=\{\mathbf{Z}\notin\mathcal{B}(0,r)\}$$

and for $r<\gamma$ we can write

$$\begin{aligned}E_2&=\left\{(\mathbf{Z}+\mathcal{B}(0,r))\cap\left(\gamma\,\mathbb{Z}^n\setminus\mathbf{0}\right)\neq\emptyset\right\}\\&\subset\{\|\mathbf{Z}\|+r\geq\gamma\}\\&=\{\mathbf{Z}\notin\mathcal{B}(0,\gamma-r)\}.\end{aligned}$$

In particular, if $\gamma>2r$ we have $E_2\subset E_1$. We choose $r^2=n\sqrt{1-\delta}2^{-\alpha}$ such that this condition indeed holds, and we can write

$$\Pr(E\mid\mathbf{G})=\Pr(E_1\cup E_3\mid G)\leq\Pr(E_1)+\Pr(E_3\mid\mathbf{G}).\quad(27)$$

Thus,

$$\begin{aligned}\Pr\left(\Pr(E\mid\mathbf{G})\geq\delta\right)&\leq\Pr\left(\Pr(E_1)+\Pr(E_3\mid\mathbf{G})\geq\delta\right)\\&=\Pr\left(\Pr(E_3\mid\mathbf{G})\geq\delta-\Pr(E_1)\right).\end{aligned}$$

Let $\delta'=\sqrt{\frac{1}{1-\delta}}-1>0$. We have for $\sigma_{\mathbf{Z}}^2\leq(1-\delta)2^{-\alpha}$

$$\begin{aligned}\Pr(E_1)&=\Pr\left(\mathbf{Z}\notin\mathcal{B}(0,r)\right)\\&=\Pr\left(\mathbf{Z}\notin\mathcal{B}\left(0,\sqrt{\frac{r^2}{n\sigma_{\mathbf{Z}}^2}}\sqrt{n\sigma_{\mathbf{Z}}^2}\right)\right)\\&\leq\Pr\left(\mathbf{Z}\notin\mathcal{B}\left(0,\sqrt{\frac{1}{\sqrt{1-\delta}}}\sqrt{n\sigma_{\mathbf{Z}}^2}\right)\right)\\&=\Pr\left(\mathbf{Z}\notin\mathcal{B}\left(0,\sqrt{(1+\delta')n\sigma_{\mathbf{Z}}^2}\right)\right).\end{aligned}$$

Since $\mathbf{Z}$ is semi norm-ergodic, it follows that $\Pr(E_1)<\delta/2$ for n large enough.

Next, we turn to upper bounding $\Pr(E_3\mid\mathbf{G})$. Note that in contrast to E_1 and E_2, this event does depend on $\mathbf{G}$. We therefore first show that $\mathbb{E}_{\mathbf{G}}\left(\Pr(E_3|\mathbf{G})\right)$ is small, and then apply Markov's inequality to show that the probability of drawing a matrix $\mathbf{G}$ for which $\Pr(E_3|\mathbf{G})>\delta/2$ is smaller than ϵ.

Let $\mathbb{1}(\mathcal{A})$ be the indicator function of the event $\mathcal{A}$.

$$\begin{aligned}&\mathbb{E}_{\mathbf{G}}\left(\Pr(E_3|\mathbf{G})\right)\\&=\mathbb{E}_{\mathbf{G}}\left(\Pr\left((\gamma\,\Lambda(\mathbf{G})\setminus\gamma\,\mathbb{Z}^n)\bigcap\mathcal{B}(\mathbf{Z},r)\neq\emptyset\right)\right)\\&=\mathbb{E}_{\mathbf{G}}\mathbb{E}_{\mathbf{Z}}\left(\mathbb{1}\left(\left(\gamma\,p^{-1}\mathcal{C}(\mathbf{G})\setminus\mathbf{0}\right)\bigcap\mathcal{B}^*(\mathbf{Z},r)\neq\emptyset\right)\mid\mathbf{G}\right)\\&=\mathbb{E}_{\mathbf{Z}}\mathbb{E}_{\mathbf{G}}\left(\mathbb{1}\left(\left(\gamma\,p^{-1}\mathcal{C}(\mathbf{G})\setminus\mathbf{0}\right)\bigcap\mathcal{B}^*(\mathbf{Z},r)\neq\emptyset\right)\mid\mathbf{Z}\right)\\&=\mathbb{E}_{\mathbf{Z}}\Pr\left(\left(\gamma\,p^{-1}\mathcal{C}(\mathbf{G})\setminus\mathbf{0}\right)\bigcap\mathcal{B}^*(\mathbf{Z},r)\neq\emptyset\mid\mathbf{Z}\right).\quad(28)\end{aligned}$$

Since each codeword in $\mathcal{C}(\mathbf{G})\setminus\mathbf{0}$ is uniformly distributed over $\mathbb{Z}_p^n$, and there are less than p^k such codewords (i.e., p^k-1), applying the union bound gives

$$\begin{aligned}\mathbb{E}_{\mathbf{G}}\left(\Pr(E_3|\mathbf{G})\right)&\leq\mathbb{E}_{\mathbf{Z}}\left(p^{k-n}\cdot\left|\gamma\,p^{-1}\mathbb{Z}_p^n\bigcap\mathcal{B}^*(\mathbf{Z},r)\right|\;\middle|\;\mathbf{Z}\right)\\&\leq\mathbb{E}_{\mathbf{Z}}\left(p^{k-n}\cdot\left|\gamma\,p^{-1}\mathbb{Z}^n\bigcap\mathcal{B}(\mathbf{Z},r)\right|\;\middle|\;\mathbf{Z}\right)\\&\leq p^{k-n}V_n\left(\frac{p}{\gamma}r+\frac{\sqrt{n}}{2}\right)^n\quad(29)\\&=p^k\gamma^{-n}V_nr^n\left(1+\frac{\gamma\sqrt{n}}{2p\,r}\right)^n\quad(30)\\&=\left(\frac{V_n^{\frac{2}{n}}}{\gamma^2p^{-\frac{2k}{n}}}r^2\right)^{\frac{n}{2}}\left(1+\frac{1}{2p}\frac{\gamma\,2^{\alpha/2}}{(1-\delta)^{1/4}}\right)^n\\&=\left(\frac{r^2}{n2^{-\alpha}}\right)^{\frac{n}{2}}\left(1+\frac{1}{2p}\frac{\gamma\,2^{\alpha/2}}{(1-\delta)^{1/4}}\right)^n\quad(31)\\&\leq(1-\delta)^{\frac{n}{4}}\left(1+\frac{2^{\alpha/2}(1-\delta)^{-1/4}}{n}\right)^n\\&\leq2^{n\left(\frac{1}{4}\log(1-\delta)+o(1)\right)},\quad(32)\end{aligned}$$

where (29) follows from Lemma 1 and (31) follows from (16) and since $\gamma=2\sqrt{n}$. Now, by (32) we have that $\mathbb{E}_{\mathbf{G}}\left(\Pr(E_3|\mathbf{G})\right)<\delta\epsilon/2$ for n large enough. Applying Markov's inequality gives that $\Pr\left(\Pr(E_3|\mathbf{G})>\delta/2\right)<\epsilon$ as desired.

VI. Mixture Noise Is Semi Norm-Ergodic for MSE-Good Coarse Lattices

Our aim is to prove Theorem 3 that states that a mixture noise composed of semi norm-ergodic noise and a dither from a lattice that is good for MSE quantization, is semi norm-ergodic. First, we show that if the sequence $\Lambda^{(n)}$ is good for MSE quantization, i.e., its normalized second moment approaches $1/2\pi e$, then a sequence of random dithers uniformly distributed over $\mathcal{V}^{(n)}$ is semi norm-ergodic. To that end, we first prove the following lemma, which is a simple extension of [32].

Lemma 2: Let $\mathcal{S}\in\mathbb{R}^n$ be a set of points with volume $V(\mathcal{S})$ and normalized second moment

$$G(\mathcal{S})=\frac{1}{nV(\mathcal{S})}\frac{\int_{\mathcal{S}}\|\mathbf{x}\|^2d\mathbf{x}}{V(\mathcal{S})^{\frac{2}{n}}}.$$

Let r_{eff} be the radius of an n-dimensional ball with the same volume as $V(\mathcal{S})$, i.e., $V(\mathcal{S})=V_nr_{\text{eff}}^n$. For any $0<\epsilon<1$ define

$$r_\epsilon\triangleq\sqrt{\frac{2\pi\,eG(\mathcal{S})-\frac{n}{n+2}(1-\epsilon)^{1+\frac{2}{n}}}{\epsilon}}r_{\text{eff}}.$$

Then, the probability that a random variable $\mathbf{U}\sim\text{Unif}(\mathcal{S})$ leaves a ball with radius r_ϵ is upper bounded by

$$\Pr\left(\mathbf{U}\notin\mathcal{B}(\mathbf{0},r_\epsilon)\right)\leq\epsilon.$$

Proof: Let $\tilde{r}_\epsilon$ be the radius of a ball that contains exactly a fraction of $1-\epsilon$ of the volume of $\mathcal{S}$, i.e.,

$$\text{Vol}\left(\mathcal{S}\bigcap\mathcal{B}(\mathbf{0},\tilde{r}_\epsilon)\right)=(1-\epsilon)V(\mathcal{S}).$$

Clearly, $\Pr(\mathbf{U} \notin \mathcal{B}(\mathbf{0}, \tilde{r}_\epsilon)) = \epsilon$. In order to establish the lemma we have to show that $\tilde{r}_\epsilon \le r_\epsilon$. To that end, we write

$$\begin{aligned} nG(\mathcal{S})V^{\frac{2}{n}}(\mathcal{S}) &= \frac{1}{V(\mathcal{S})}\int_{\mathbf{x}\in\mathcal{S}} \|\mathbf{x}\|^2 d\mathbf{x} \\ &= \frac{1}{V(\mathcal{S})}\left(\int_{\mathbf{x}\in(\mathcal{S}\cap\mathcal{B}(\mathbf{0},\tilde{r}_\epsilon))} \|\mathbf{x}\|^2 d\mathbf{x} \right. \\ &\qquad \left. + \int_{\mathbf{x}\in(\mathcal{S}\cap(\mathbb{R}^n\setminus\mathcal{B}(\mathbf{0},\tilde{r}_\epsilon)} \|\mathbf{x}\|^2 d\mathbf{x}\right). \quad (33) \end{aligned}$$

The first integral in (33) may be lower bounded by replacing its integration boundaries with an n-dimensional ball $\mathcal{B}(\mathbf{0}, \rho_\epsilon)$, where

$$\rho_\epsilon^2 = V_n^{-\frac{2}{n}}(1-\epsilon)^{\frac{2}{n}}V^{\frac{2}{n}}(\mathcal{S}) \quad (34)$$

is chosen such that $V_n\rho_\epsilon^n = (1-\epsilon)V(\mathcal{S})$. Thus

$$\begin{aligned} \int_{\mathbf{x}\in(\mathcal{S}\cap\mathcal{B}(\mathbf{0},\tilde{r}_\epsilon)} \|\mathbf{x}\|^2 d\mathbf{x} &\ge \int_{\mathbf{x}\in\mathcal{B}(\mathbf{0},\rho_\epsilon)} \|\mathbf{x}\|^2 d\mathbf{x} \\ &= nV_n\rho_\epsilon^n\sigma^2(\mathcal{B}(\mathbf{0},\rho_\epsilon)) \\ &= \frac{n}{n+2}V_n\rho_\epsilon^n\rho_\epsilon^2 \quad (35) \\ &= \frac{n}{n+2}\frac{V^{1+\frac{2}{n}}(\mathcal{S})(1-\epsilon)^{1+\frac{2}{n}}}{V_n^{\frac{2}{n}}} \\ &= \frac{n}{n+2}V(\mathcal{S})(1-\epsilon)^{1+\frac{2}{n}}r_{\text{eff}}^2, \quad (36) \end{aligned}$$

where we have used (3) to get (35). The second integral in (33) is over a set of points with volume $\epsilon V(\mathcal{S})$ which are all at distance greater than $\tilde{r}_\epsilon$ from the origin. Therefore, it can be bounded as

$$\int_{\mathbf{x}\in(\mathcal{S}\cap(\mathbb{R}^n\setminus\mathcal{B}(\mathbf{0},\tilde{r}_\epsilon)))} \|\mathbf{x}\|^2 d\mathbf{x} \ge \epsilon V(\mathcal{S})\tilde{r}_\epsilon^2. \quad (37)$$

Substituting (36) and (37) into (33) gives

$$nG(\mathcal{S})V^{\frac{2}{n}}(\mathcal{S}) \ge \left(\frac{n}{n+2}(1-\epsilon)^{1+\frac{2}{n}}r_{\text{eff}}^2 + \epsilon\tilde{r}_\epsilon^2\right). \quad (38)$$

Using the fact that $V^{\frac{2}{n}}(\mathcal{S}) = V_n^{\frac{2}{n}}r_{\text{eff}}^2$, (38) reduces to

$$\begin{aligned} \tilde{r}_\epsilon^2 &\le \frac{nV_n^{\frac{2}{n}}G(\mathcal{S}) - \frac{n}{n+2}(1-\epsilon)^{1+\frac{2}{n}}}{\epsilon}r_{\text{eff}}^2 \\ &\le \frac{2\pi e G(\mathcal{S}) - \frac{n}{n+2}(1-\epsilon)^{1+\frac{2}{n}}}{\epsilon}r_{\text{eff}}^2, \end{aligned}$$

as desired. ∎

Using Lemma 2 we can prove the following.

Lemma 3: Let $\Lambda^{(n)}$ be a sequence of lattices that is good for MSE quantization. Then the sequence of random dither vectors $\mathbf{U}^{(n)} \sim \text{Unif}(\mathcal{V}^{(n)})$ is semi norm-ergodic.

Proof: We need to show that for any $\epsilon, \delta > 0$ and n large enough

$$\Pr\left(\mathbf{U}^{(n)} \notin \mathcal{B}(0, \sqrt{(1+\delta)n\sigma^2\left(\Lambda^{(n)}\right)})\right) \le \epsilon.$$

By Lemma 2, it suffices to show that

$$\begin{aligned} &\sqrt{\frac{2\pi e G\left(\Lambda^{(n)}\right) - \frac{n}{n+2}(1-\epsilon)^{1+\frac{2}{n}}}{\epsilon}}r_{\text{eff}}\left(\Lambda(n)\right) \\ &\qquad \le \sqrt{(1+\delta)n\sigma^2\left(\Lambda^{(n)}\right)}. \quad (39) \end{aligned}$$

From (7), we have

$$r_{\text{eff}}\left(\Lambda^{(n)}\right) \le \sqrt{(n+2)\sigma^2\left(\Lambda^{(n)}\right)}. \quad (40)$$

and the l.h.s. of (39) can be therefore upper bounded by

$$\sqrt{n\sigma^2\left(\Lambda^{(n)}\right)}\sqrt{\frac{\frac{n+2}{n}2\pi e G\left(\Lambda^{(n)}\right) - (1-\epsilon)^{1+\frac{2}{n}}}{\epsilon}} \quad (41)$$

The sequence of lattices $\Lambda^{(n)}$ is good for MSE quantization, and therefore for any $\delta_1 > 0$ and n large enough

$$G(\Lambda^{(n)}) < (1+\delta_1)\frac{1}{2\pi e}.$$

Setting $\delta_1 = \delta\epsilon/3$, we have that for n large enough

$$\begin{aligned} &\frac{n+2}{n}2\pi e G\left(\Lambda^{(n)}\right) - (1-\epsilon)^{1+\frac{2}{n}} \\ &\qquad \le \frac{n+2}{n}\left(1+\frac{\delta\epsilon}{3}\right) - (1-\epsilon)^{1+\frac{2}{n}} \\ &\qquad \le \epsilon + \delta\epsilon, \quad (42) \end{aligned}$$

where the last inequality follows since for n large enough $\frac{n+2}{n}(1+\frac{\delta\epsilon}{3}) < 1 + \frac{2\delta\epsilon}{3}$ and $(1-\epsilon)^{1+\frac{2}{n}} > 1 - \epsilon - \frac{\delta\epsilon}{3}$. Combining (41) and (42) establishes (39). ∎

We are now ready to prove Theorem 3.

Proof of Theorem 3: Since $\mathbf{N}$ and $\mathbf{U}$ are statistically independent, the effective variance of $\mathbf{Z}$ is

$$\sigma_{\mathbf{Z}}^2 = \frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2 = \alpha^2\sigma_{\mathbf{N}}^2 + \beta^2\sigma_{\mathbf{U}}^2.$$

We have to prove that for any $\epsilon > 0$, $\delta > 0$ and n large enough

$$\Pr\left(\mathbf{Z} \notin \mathcal{B}(\mathbf{0}, \sqrt{(1+\delta)n\sigma_{\mathbf{Z}}^2})\right) < \epsilon.$$

For any $\epsilon > 0$, $\delta > 0$ and n large enough we have

$$\begin{aligned} &\Pr\left(\mathbf{Z} \notin \mathcal{B}(\mathbf{0}, \sqrt{(1+\delta)n\sigma_{\mathbf{Z}}^2})\right) \\ &= \Pr\left(\|\mathbf{Z}\|^2 > (1+\delta)n\sigma_{\mathbf{Z}}^2\right) \\ &= \Pr\left(\|\mathbf{N}\|^2 > (1+\delta)n\sigma_{\mathbf{N}}^2\right) \\ &\quad \cdot \Pr\left(\|\mathbf{Z}\|^2 > (1+\delta)n\sigma_{\mathbf{Z}}^2 \mid \|\mathbf{N}\|^2 > (1+\delta)n\sigma_{\mathbf{N}}^2\right) \\ &\quad + \Pr\left(\|\mathbf{N}\|^2 \le (1+\delta)n\sigma_{\mathbf{N}}^2\right) \\ &\quad \cdot \Pr\left(\|\mathbf{Z}\|^2 > (1+\delta)n\sigma_{\mathbf{Z}}^2 \mid \|\mathbf{N}\|^2 \le (1+\delta)n\sigma_{\mathbf{N}}^2\right) \\ &\le \frac{\epsilon}{3} + \Pr\left(\beta^2\|\mathbf{U}\|^2 + 2\alpha\beta\mathbf{N}^T\mathbf{U}\right. \\ &\qquad \left. > (1+\delta)n\beta^2\sigma_{\mathbf{U}}^2 \mid \|\mathbf{N}\|^2 \le (1+\delta)n\sigma_{\mathbf{N}}^2\right) \quad (43) \end{aligned}$$

$$\leq \frac{\epsilon}{3} + \Pr\left(\beta^2\|\mathbf{U}\|^2 > n\beta^2\sigma_{\mathbf{U}}^2(1+\delta/2)\right) + \Pr\left(2\alpha\beta\mathbf{N}^T\mathbf{U} > n\beta^2\sigma_{\tilde{\mathbf{U}}}^2\delta/2 \;\middle|\; \|\mathbf{N}\|^2 \leq (1+\delta)n\sigma_{\mathbf{N}}^2\right) \tag{44}$$

$$\leq \frac{2\epsilon}{3} + \Pr\left(2\alpha\beta\mathbf{N}^T\mathbf{U} > n\beta^2\sigma_{\mathbf{U}}^2\delta/2 \;\middle|\; \|\mathbf{N}\|^2 \leq (1+\delta)n\sigma_{\mathbf{N}}^2\right), \tag{45}$$

where (43) follows from the fact that $\mathbf{N}$ is semi norm-ergodic, (44) from the union bound and (45) from the fact that $\mathbf{U}$ is semi norm-ergodic due to Lemma 3. We are left with the task of showing that the last probability in (45) can be made smaller than $\epsilon/3$ for n large enough. This requires some more work.

Since $\mathbf{U}$ is semi norm-ergodic noise, then for any $\epsilon_2 > 0$, $\delta_2 > 0$ and n large enough

$$\Pr\left(\|\mathbf{U}\| > \sqrt{(1+\delta_2)n\sigma_{\mathbf{U}}^2}\right) < \epsilon_2.$$

Let $r_{\mathbf{U}} = \sqrt{(1+\delta_2)n\sigma_{\mathbf{U}}^2}$, and $f_{\mathbf{U}}(\mathbf{u})$ be the probability density function (pdf) of $\mathbf{U}$. For any $r > 0$ we have

$$\begin{aligned}
&\Pr\left(\mathbf{N}^T\mathbf{U} > r \;\middle|\mathbf{N} = \mathbf{n}\right)\\
&= \int_{|\mathbf{u}|\leq r_{\mathbf{u}}} f_{\mathbf{U}}(\mathbf{u})\mathbb{1}(\mathbf{n}^T\mathbf{u} > r)d\mathbf{u} + \int_{|\mathbf{u}|> r_{\mathbf{u}}} f_{\mathbf{U}}(\mathbf{u})\mathbb{1}(\mathbf{n}^T\mathbf{u} > r)d\mathbf{u}\\
&\leq \int_{|\mathbf{u}|\leq r_{\mathbf{u}}} \frac{1}{V(\Lambda)}\mathbb{1}(\mathbf{n}^T\mathbf{u} > r)d\mathbf{u} + \epsilon_2\\
&= \frac{V(\mathcal{B}(\mathbf{0}, r_{\mathbf{U}}))}{V(\Lambda)} \int_{|\mathbf{u}|\leq r_{\mathbf{u}}} \frac{1}{V(\mathcal{B}(\mathbf{0}, r_{\mathbf{U}}))}\mathbb{1}(\mathbf{n}^T\mathbf{u} > r)d\mathbf{u} + \epsilon_2.
\end{aligned}$$

Using the fact that Λ is good for MSE quantization we have $V(\Lambda)^{\frac{2}{n}} \to 2\pi e\sigma_{\mathbf{U}}^2$, and hence, for n large enough,

$$\left(\frac{V(\mathcal{B}(\mathbf{0}, r_{\mathbf{U}}))}{V(\Lambda)}\right)^{\frac{2}{n}} < (1+2\delta_2).$$

Let $\tilde{\mathbf{U}}$ be a random vector uniformly distributed over $\mathcal{B}(\mathbf{0}, r_{\mathbf{U}})$. We have

$$\Pr\left(\mathbf{N}^T\mathbf{U} > r \;\middle|\mathbf{N} = \mathbf{n}\right) < \epsilon_2 + (1+2\delta_2)^{\frac{n}{2}}\Pr(\mathbf{n}^T\tilde{\mathbf{U}} > r). \tag{46}$$

Let $\tilde{\mathbf{Z}}$ be AWGN with zero mean and variance $r_{\mathbf{U}}^2/n$. Using a similar approach to that taken in [14, Lemma 11], we would now like to upper bound the pdf of $\tilde{\mathbf{U}}$ using that of $\tilde{\mathbf{Z}}$. For any $\mathbf{x} \in \mathbb{R}^n$ we have

$$\frac{f_{\tilde{\mathbf{U}}}(\mathbf{x})}{f_{\tilde{\mathbf{Z}}}(\mathbf{x})} = \frac{f_{\tilde{\mathbf{U}}}(\|\mathbf{x}\|)}{f_{\tilde{\mathbf{Z}}}(\|\mathbf{x}\|)} \leq \frac{f_{\tilde{\mathbf{U}}}(r_{\mathbf{U}})}{f_{\tilde{\mathbf{Z}}}(r_{\mathbf{U}})} = \left(\frac{2\pi e}{nV_n^{\frac{2}{n}}}\right)^{\frac{n}{2}}.$$

Thus, for any $\mathbf{x} \in \mathbb{R}^n$

$$f_{\tilde{\mathbf{U}}}(\mathbf{x}) \leq 2^{\frac{n}{2}\log\left(\frac{2\pi e}{n}V_n^{-\frac{2}{n}}\right)} f_{\tilde{\mathbf{Z}}}(\mathbf{x}).$$

We can further bound (46) for large enough n as

$$\begin{aligned}
&\Pr\left(\mathbf{N}^T\mathbf{U} > r \;\middle|\mathbf{N} = \mathbf{n}\right)\\
&\leq \epsilon_2 + 2^{\frac{n}{2}\log\left((1+2\delta_2)\frac{2\pi e}{n}V_n^{-\frac{2}{n}}\right)}\Pr(\mathbf{n}^T\tilde{\mathbf{Z}} > r)\\
&= \epsilon_2 + 2^{\frac{n}{2}\log\left((1+2\delta_2)\frac{2\pi e}{n}V_n^{-\frac{2}{n}}\right)} Q\left(\frac{\sqrt{n}r}{\|\mathbf{n}\|r_{\mathbf{U}}}\right),
\end{aligned}$$

where $Q(\cdot)$ is the standard Q-function, which satisfies $Q(x) < e^{-x^2/2}$. It follows that

$$\begin{aligned}
&\Pr\left(2\alpha\beta\mathbf{N}^T\mathbf{U} > n\beta^2\sigma_{\mathbf{U}}^2\delta/2 \;\middle|\; \|\mathbf{N}\|^2 \leq (1+\delta)n\sigma_{\mathbf{N}}^2\right)\\
&\leq \epsilon_2 + 2^{\frac{n}{2}\log\left((1+2\delta_2)\frac{2\pi e}{n}V_n^{-\frac{2}{n}}\right)} Q\left(\frac{\sqrt{n}\beta\sigma_{\mathbf{U}}\delta/2}{2\alpha\sigma_{\mathbf{N}}\sqrt{(1+\delta)(1+2\delta_2)}}\right).
\end{aligned}$$

Taking δ_2 sufficiently smaller than δ and $\epsilon_2 < \epsilon/6$, for n large enough we have

$$\Pr\left(2\alpha\beta\mathbf{N}^T\mathbf{U} > n\beta^2\sigma_{\mathbf{U}}^2\delta/2 \;\middle|\; \|\mathbf{N}\|^2 \leq (1+\delta)n\sigma_{\mathbf{N}}^2\right) < \frac{\epsilon}{3}.$$

■

We end this section with two simple corollaries of Theorem 3. The first follows since any i.i.d. noise is semi norm-ergodic, and the second follows by iterating over Theorem 3.

Corollary 1: Let $\mathbf{Z} = \alpha\mathbf{N} + \beta\mathbf{U}$, where $\alpha, \beta \in \mathbb{R}$, $\mathbf{N}$ is an i.i.d. noise vector, and $\mathbf{U}$ is a dither statistically independent of $\mathbf{N}$, uniformly distributed over the Voronoi region $\mathcal{V}$ of a lattice Λ that is good for MSE quantization. Then, the random vector $\mathbf{Z}$ is semi norm-ergodic.

Corollary 2: Let $\mathbf{U}_1, \cdots, \mathbf{U}_K$ be statistically independent dither random vectors, each uniformly distributed over the Voronoi region $\mathcal{V}_k$ of Λ_k, $k = 1, \ldots, K$, that are all good for MSE quantization. Let $\mathbf{N}$ be a semi norm-ergodic random vector statistically independent of $\{\mathbf{U}_1, \cdots, \mathbf{U}_K\}$. For any $\alpha, \beta_1, \cdots, \beta_K \in \mathbb{R}$ the random vector $\mathbf{Z} = \alpha\mathbf{N} + \sum_{k=1}^K \beta_k\mathbf{U}_k$ is semi norm-ergodic.

VII. Nested Lattice Codes With a Cubic Coarse Lattice

In this section we prove Theorem 5. As before, we consider an ensemble of p-ary random Construction A lattices. More precisely, we draw a matrix $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$ with i.i.d. entries uniformly distributed over $\mathbb{Z}_p$, and construct the (random) lattice $\gamma\,\Lambda(\mathbf{G})$ as in Definition 7, with $\gamma = \sqrt{12\mathsf{SNR}}$. We take $\gamma\,\Lambda(\mathbf{G})$ as a fine lattice and $\gamma\,\mathbb{Z}^n \subset \gamma\,\Lambda(\mathbf{G})$ as a coarse lattice, to construct the nested lattice codebook $\mathcal{L} = \gamma\,\Lambda(\mathbf{G}) \cap \gamma$ CUBE. Clearly, $\sigma^2(\gamma\,\mathbb{Z}^n) = \mathsf{SNR}$ and the rate of all codebooks in the ensemble is $R = \frac{k}{n}\log p$.

Applying the mod-Λ scheme with the codebook $\mathcal{L}$, as described in the proof of Theorem 4, gives rise to the effective channel (9), where $\mathbf{Z}_{\text{eff}}$ is as defined in (10). Note that for the coarse lattice $\gamma\,\mathbb{Z}^n$ which is used, the random vector $\mathbf{X}$ is i.i.d. with each component uniformly distributed over $[-\gamma/2, \gamma/2)$. Thus, $\mathbf{Z}_{\text{eff}}$ is i.i.d., and in particular semi norm-ergodic, with variance $\sigma_{\mathbf{Z}_{\text{eff}}}^2(\alpha) = \alpha^2 + (1-\alpha)^2\mathsf{SNR}$. As in the proof of Theorem 4, we choose $\alpha = \mathsf{SNR}/(1+\mathsf{SNR})$ such as to

minimize $\sigma^2_{\mathbf{Z}_{\text{eff}}}(\alpha)$, which gives $\sigma^2_{\mathbf{Z}_{\text{eff}}} = \mathsf{SNR}/(1+\mathsf{SNR})$. As in (12), the decoder finds

$$\hat{\mathbf{t}} = \left[Q_{\gamma\Lambda(\mathbf{G})}(\mathbf{Y}_{\text{eff}})\right] \bmod \Lambda_c = \left[Q_{\gamma\Lambda(\mathbf{G})}(\mathbf{t}+\mathbf{Z}_{\text{eff}})\right] \bmod \gamma\mathbb{Z}^n,$$

and outputs the message corresponding to $\hat{\mathbf{t}}$. In order to complete the proof we will need the following lemma.

Lemma 4: Let n be a natural number, p a prime number, $R>0$, $k = nR/\log p$ and $\gamma > 0$. Let $\mathbf{G} \in \mathbb{Z}_p^{k\times n}$ be a random matrix with i.i.d. entries uniformly distributed over $\mathbb{Z}_p$, and $\Lambda(\mathbf{G})$ be constructed as in Definition 7. Let $\mathbf{Z}$ be an additive semi norm-ergodic noise with effective variance $\sigma^2_{\mathbf{Z}} = \frac{1}{n}\mathbb{E}\|\mathbf{Z}\|^2$ and define $\Gamma(p,\gamma^2/\sigma^2_{\mathbf{Z}}) \triangleq \log\left(1+\sqrt{\frac{\gamma^2}{4p^2\sigma^2_{\mathbf{Z}}}}\right)$. For any $\epsilon,\delta>0$ and n large enough, if $R < \frac{1}{2}\log\left(\frac{\gamma^2}{(1+\delta)2\pi e\sigma^2_{\mathbf{Z}}}\right) - \Gamma(p,\gamma^2/\sigma^2_{\mathbf{Z}})$, then

$$\Pr\left(\Pr\left(Q_{\gamma\Lambda(\mathbf{G})}(\mathbf{t}+\mathbf{Z}) \neq \mathbf{t} \bmod \gamma\mathbb{Z}^n \mid \mathbf{G}\right) > \delta\right) < \epsilon. \quad (47)$$

for any $\mathbf{t} \in \gamma\Lambda(\mathbf{G})$.

Note that in (47), the error probability in *coset nearest neighbor decoding* is required to be smaller than δ. In other words, the decoder is only required to find the correct coset $\gamma\Lambda(\mathbf{G})/\gamma\mathbb{Z}^n$ to which $\mathbf{t}$ belongs, and not the exact point $\mathbf{t}$ that was transmitted. See Figure 1 for an illustration of coset nearest neighbor decoding.

Theorem 5 now follows by applying Lemma 4 with $\gamma = \sqrt{12\mathsf{SNR}}$, $\sigma^2_{\mathbf{Z}} = \mathsf{SNR}/(1+\mathsf{SNR})$ and taking δ to zero. This shows that for every $\delta>0$, for almost every $\mathbf{G}$ and n large enough, the error probability of the mod-Λ scheme with codebook $\mathcal{L} = \gamma\Lambda(\mathbf{G}) \cap \sqrt{12\mathsf{SNR}}\cdot\text{CUBE}$ is smaller than δ. In particular, there exists a sequence of such codebooks with vanishing error probability.

It now only remains to prove Lemma 4.

Proof of Lemma 4: The proof is similar to that of Theorem 6, part 3, with a few differences we now specify.

We upper bound the error probability of the coset nearest neighbor decoder with that of a bounded distance coset decoder. The latter finds all points of $\gamma\Lambda(\mathbf{G})$ in a ball of radius r around the output $\mathbf{t}+\mathbf{Z}_{\text{eff}}$ and outputs the list of all these points reduced modulo $\gamma\mathbb{Z}^n$. If the list of cosets does not contain exactly one point, an error is declared. It can be verified that an error event E of this decoder is the union of E_1 and E_3, defined in Section V-C. The event E_2 that was defined there, corresponds to decoding a point different than $\mathbf{t}$ inside the same coset as $\mathbf{t}$. This event does not incur an error for coset nearest neighbor decoding. Thus, equation (27) continues to hold here.

We take the decoding radius as $r^2 = n(1+\delta)\sigma^2_{\mathbf{Z}}$, such that by the semi norm-ergodicity of $\mathbf{Z}_{\text{eff}}$, it follows that for n large enough $\Pr(E_1) < \delta/2$. In order to upper bound $\Pr(\Pr(E_3 \mid \mathbf{G}))$ we upper bound $\mathbb{E}_{\mathbf{G}}(\Pr(E_3 \mid \mathbf{G}))$ and then apply Markov's inequality. By (30) we have

$$\begin{aligned}
&\mathbb{E}_{\mathbf{G}}(\Pr(E_3 \mid \mathbf{G})) \\
&\quad\le p^k\gamma^{-n}V_n r^n\left(1+\frac{\gamma\sqrt{n}}{2p\,r}\right)^n \\
&\quad= 2^{n\left(R+\frac{1}{2}\log\left(V_n^{\frac{2}{n}}\frac{r^2}{\gamma^2}\right)+\log\left(1+\sqrt{\frac{n}{4p^2}\frac{\gamma^2}{r^2}}\right)\right)} \\
&\quad\le 2^{n\left(R+\frac{1}{2}\log\left(\frac{2\pi e}{n}\frac{n(1+\delta)\sigma^2_{\mathbf{Z}}}{\gamma^2}\right)+\log\left(1+\sqrt{\frac{n}{4p^2}\frac{\gamma^2}{n(1+\delta)\sigma^2_{\mathbf{Z}}}}\right)\right)} \\
&\quad= 2^{-n\left(\frac{1}{2}\log\left(\frac{\gamma^2}{2\pi e(1+\delta)\sigma^2_{\mathbf{Z}}}\right)-\log\left(1+\sqrt{\frac{\gamma^2}{4p^2\sigma^2_{\mathbf{Z}}}}\right)-R\right)},
\end{aligned}$$

Thus, for any $R < \frac{1}{2}\log\left(\frac{\gamma^2}{2\pi e(1+\delta)\sigma^2_{\mathbf{Z}}}\right) - \Gamma(p,\gamma^2/\sigma^2)$, we have that $\mathbb{E}_{\mathbf{G}}(\Pr(E_3|\mathbf{G}) < \delta\epsilon/2)$, for n large enough. Applying Markov's inequality gives that $\Pr(\Pr(E_3|\mathbf{G}) > \delta/2) < \epsilon$ as desired. ∎

VIII. Discussion

We have presented a novel proof for the existence of chains of nested lattices, where all lattices in the chain are good for coding and for MSE quantization. Our analysis of the coding schemes based on such chains revealed that these relaxed "goodness" requirements are sufficient for achieving the capacity of the AWGN channel (and the rate-distortion function of a white Gaussian source). In fact, our analysis did not require the assumption that the additive noise (or the source) is AWGN. Instead, it was only assumed that the additive noise (or source) is semi norm-ergodic. Consequently, our results show that lattice-based coding schemes are robust to variations in the distribution of the additive noise (or source), and their (first-order) performance, essentially depend only on its second moment.

Our analysis required the cardinality p of the finite-field over which the underlying linear codes are constructed to grow with n. In order to achieve capacity, the density of the grid from which the lattice points are chosen must be small. When using Construction A for lifting a p-ary linear code to a cube of side γ, this density is given by γp^{-1}. Thus, the size of γ in our construction, formed the main constraint on the minimal possible size of p. We took $\gamma = \Theta(\sqrt{n})$ for two reasons. The first was to justify equation (21), which requires that a cube of side γ contains an n-dimensional ball with radius $\Theta(\sqrt{n})$. The second, was to ensure that the corresponding lattice has a low error probability under lattice decoding, and not only under coset nearest neighbor decoding. For channel coding problems, it suffices to ensure that the latter decoding rule has a small error probability. For source coding applications, however, the lattice decoder must succeed with high probability. An interesting direction for future research is to quantify the performance loss of nested lattice codes based on linear p-ary code with a *finite* value of p. The results of [31] (see also [10, Sec. 7.9.5]) show that for the quantization problem the loss decreases rather fast with p. Our analysis in Section VII reveals a similar phenomenon for the coding problem. It is thus expected that nested lattice codes constructed from our ensemble with finite values for p and γ would also perform well.

Acknowledgment

The authors thank Bobak Nazer, Yair Yona and Ram Zamir for discussions that helped prompt this work, and the anonymous reviewers for their careful reading of the manuscript and their thoughtful comments.

REFERENCES

[1] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 139–149, Mar. 1982.

[2] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[3] C. A. Rogers, "Lattice coverings of space," *Mathematika*, vol. 6, no. 1, pp. 33–39, 1959.

[4] I. F. Blake, "The Leech lattice as a code for the Gaussian channel," *Inf. Control*, vol. 19, no. 1, pp. 66–74, 1971.

[5] R. de Buda, "The upper error bound of a new near-optimal code," *IEEE Trans. Inf. Theory*, vol. 21, no. 4, pp. 441–445, Jul. 1975.

[6] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, Jan. 1998.

[7] T. Linder, C. Schlegal, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1735–1737, Sep. 1993.

[8] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

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

[10] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[11] J. Conway and N. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, pp. 820–824, Nov. 1983.

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

[13] G. D. Forney, Jr., M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[14] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[15] T. Liu, P. Moulin, and R. Koetter, "On error exponents of modulo lattice additive noise channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 454–471, Feb. 2006.

[16] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[17] T. Philosof, R. Zamir, U. Erez, and A. J. Khisti, "Lattice strategies for the dirty multiple access channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5006–5035, Aug. 2011.

[18] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[19] W. Nam, S.-Y. Chung, and Y. H. Lee, "Capacity of the Gaussian two-way relay channel to within $1/2$ bit," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5488–5494, Nov. 2010.

[20] G. Bresler, A. Parekh, and D. N. C. Tse, "The approximate capacity of the many-to-one and one-to-many Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4566–4592, Sep. 2010.

[21] O. Ordentlich, U. Erez, and B. Nazer, "The approximate sum capacity of the symmetric Gaussian K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3450–3482, Jun. 2014.

[22] B. Nazer, V. R. Cadambe, V. Ntranos, and G. Caire, "Expanding the compute-and-forward framework: Unequal powers, signal levels, and multiple linear combinations," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: http://arxiv.org/abs/1504.01690

[23] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.

[24] D. Krithivasan and S. S. Pradhan, "A proof of the existence of good nested lattices," *UM CSPL Tech. Rep. Ser.*, 2007.

[25] R. Zamir and S. Shamai (Shitz), "Nested linear/lattice codes for Wyner–Ziv encoding," in *Proc. Inf. Theory Workshop*, Jun. 1998, pp. 92–93.

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

[27] W. Nam, S.-Y. Chung, and Y. H. Lee, "Nested lattice codes for Gaussian relay networks with interference," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7733–7745, Dec. 2011.

[28] N. Di Pietro, "On infinite and finite lattice constellations for the additive white Gaussian noise channel," Ph.D. dissertation, Inst. Math. Bordeaux, Univ. Bordeaux, Pessac, France, 2014.

[29] C. Ling and J.-C. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[30] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[31] B. D. Kudryashov and K. V. Yurkov, "Random coding bound for the second moment of multidimensional lattices," *Problems Inf. Transmiss.*, vol. 43, no. 1, pp. 57–68, 2007.

[32] U. Erez and R. Zamir, "Bounds on the ϵ-covering radius of linear codes with applications to self-noise in nested Wyner–Ziv coding," Dept. Elect. Eng. Syst., Tel Aviv Univ., Tel Aviv, Israel, Tech. Rep., 2002. [Online]. Available: http://www.eng.tau.ac.il/~zamir/techreport/selfnoiseTR.pdf

Or Ordentlich received the B.Sc. degree (*cum laude*) in 2010, M.Sc. degree (*summa cum laude*) in 2011, and his PhD degree in 2016, all in electrical engineering in Tel Aviv University, Israel. He is currently a postdoctoral fellow in the Laboratory for Information and Decision Systems at the Massachusetts Institute of Technology (MIT), Cambridge. Or is the recipient of the MIT - Technion Postdoctoral Fellowship, the Adams Fellowship awarded by the Israel Academy of Sciences and Humanities, the Thalheimer Scholarship for graduate students, the Advanced Communication Center (ACC) Feder Family Award for outstanding research work in the field of communication technologies (2011,2014), and the Weinstein Prize for research in signal processing (2011,2013,2014).

Uri Erez (M'09) was born in Tel-Aviv, Israel, on October 27, 1971. He received the B.Sc. degree in mathematics and physics and the M.Sc. and Ph.D. degrees in electrical engineering from Tel-Aviv University in 1996, 1999, and 2003, respectively During 2003-2004, he was a Postdoctoral Associate at the Signals, Information and Algorithms Laboratory at the Massachusetts Institute of Technology (MIT), Cambridge. Since 2005, he has been with the Department of Electrical EngineeringâĂŞ-Systems at Tel-Aviv University. His research interests are in the general areas of information theory and digital communication. He served in the years 2009-2011 as Associate Editor for Coding Techniques for the IEEE TRANSACTIONS ON INFORMATION THEORY.