Can Structure Beat Random? - The Story of Lattice Codes

Ram Zamir,
EE - Systems, Tel Aviv University, ISRAEL
zamir@eng.tau.ac.il

***Abstract*— At the birth of information theory, Shannon surprised the communication world with the concept of random coding, which he used for proving the ultimate limits of his theory. This powerful tool is, however, non-constructive. In Shannon's words: "An attempt to obtain a good approximation to ideal coding by following the method of the proof is generally impractical... related to the difficulty of giving an explicit construction for a good approximation to a random sequence." A practical substitute to random coding are structured codes (one example of which - the Hamming code - appeared already in Shannon's paper from 1948). Multiterminal information theory provides us now with a new surprise: for some distributed coding problems structured codes seem to be better than random codes! This summary of my ISIT 2010 plenary talk discusses how lattice codes are used in Gaussian multiterminal settings, and the intuition they provide for the question in the title.**

I. Motivation

It is not hard to detect the few differences between the two faces in Fig. 1. Once detected, it is also not too hard to describe them with just a few words. But would a few words be sufficient if the two faces were described by two *separate* observers?

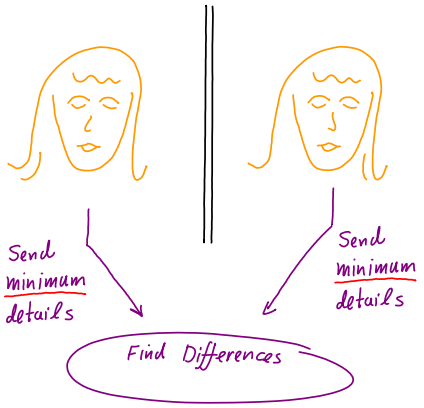

Fig. 1: Find (and communicate) the differences.

An information-theoretic analogue of this question is the "two help one" problem of Fig. 2, which was proposed in a seminal paper from the late 70's by Körner and Marton [25]. They showed that if one wishes to reconstruct the modulo-two sum of two correlated binary sources from their independent encodings, then linear coding seems to be better than random coding.

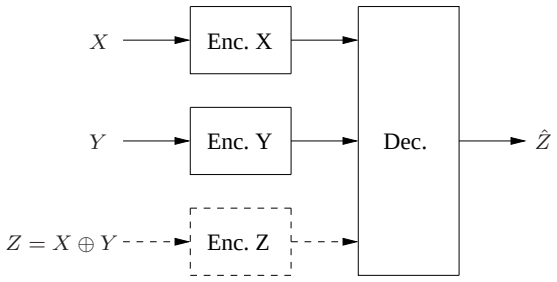

Fig. 2: The Körner-Marton configuration.

Specifically, the Körner-Marton (KM) setup consists of a binary doubly symmetric source (X, Y), and an "error" variable $Z = X \oplus Y$ indicating when X and Y are different, i.e., $\Pr(Z = 1) = \Pr(X \neq Y) = \theta$. The goal is to encode the sources X and Y separately such that Z can be reconstructed losslessly. If coordination between the encoders were allowed, then they could compute the XOR sequence $Z_1, \ldots, Z_n$ and encode it at a rate of $H(Z)$. Via a "genie aided" argument, Körner and Marton showed that in the *un*coordinated case, the sum rate required is at least

$$R_x + R_y \geq 2H(Z). \tag{1}$$

Furthermore, this sum rate can be achieved by a *linear code*: each encoder transmits the syndrome of the observed source relative to a good linear binary code for a BSC with crossover probability θ.

A common technique in proving direct coding theorems in information theory is the use of a *random* code, induced by some *single-letter* formula. In an attempt to find such a formula for the problem in Fig. 2, Körner and Marton examined a "natural" extension for the solution of the "one help one" problem [1], [46]; the resulting achievable rates satisfy, [25, appendix]

$$R_x + R_y \geq H(X, Y). \tag{2}$$

These rates correspond to Slepian-Wolf encoding of X and Y [9],[1] and are clearly strictly contained in (1) (since $H(X, Y) = 1 + H(Z)$ in (2) is greater than $2H(Z)$ for

[1]It can also be derived from the Berger-Tung achievable region [3] for distributed lossy coding of X and Y with one reconstruction $\hat{Z}$ under the distortion measure $d(X, Y, \hat{Z}) \triangleq X \oplus Y \oplus \hat{Z}$.

$\theta \neq \frac{1}{2}$). Thus, the "natural" random binning solution for the "two help one" problem is suboptimal, and inferior to structured (linear) coding.

Does this mean that *any* random coding scheme (i.e., single-letter solution) would be suboptimal for the "two help one" problem? Instead of dealing with that directly, we turn to structured (lattice) coding in the Euclidean space, with the hope to get further intuition about this issue in multi-terminal Gaussian setups.

II. Why Lattices?

Lattices form effective arrangements of points in space for various geometric and coding problems, e.g., sphere covering and packing, quantization, and signaling for the additive white Gaussian-noise (AWGN) channel [6], [16], [10]. The best lattice for each problem may be different. Nevertheless, as the dimension goes to infinity, there exist lattices which tend to be "perfect" for all problems.

In the context of this talk, lattices serve as a bridge from the low dimensions of common modulation techniques (PCM, PAM, QAM) to the large dimensions of coded modulation schemes, or to the infinite dimension of Shannon's theory. They also provide an "algebraic" binning scheme for some Gaussian side information problems [52], [13]. Moreover, recent developments in the area of Gaussian network information theory, [35], [36], [26], [40], [41], [38], indicate that lattices are sometimes even better than their random coding counterparts!

III. Lattice Definitions and Figures of Merit

An n-dimensional lattice Λ is defined by a set of n basis vectors $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_n$ in $\mathbb{R}^n$. The lattice Λ is composed of all integer combinations of the basis vectors, i.e.,

$$\Lambda = \left\{ \lambda = G \cdot \mathbf{i} : \ \mathbf{i} \in \mathbb{Z}^n \right\}, \tag{3}$$

where $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the $n \times n$ *generator matrix* G is given by $G = [\, \boldsymbol{g}_1 \,|\, \boldsymbol{g}_2 | \ldots | \, \boldsymbol{g}_n \,]$. When G is the unit matrix, we obtain the integer lattice $\mathbb{Z}^n$. Thus, Λ in (3) can be written also as $G\mathbb{Z}^n$. Note that the zero vector is always a lattice point, and that G is not unique for a given Λ. See [6] as an excellent background.

A few important notions are associated with a lattice. The nearest neighbor quantizer $Q_\Lambda(\cdot)$ is defined by

$$Q_\Lambda(\mathbf{x}) = \lambda \in \Lambda \text{ if } \ \| \, \mathbf{x} - \lambda \, \| \leq \| \, \mathbf{x} - \lambda' \, \| \ \ \forall \, \lambda' \in \Lambda \tag{4}$$

where $\| \cdot \|$ denotes Euclidean norm, and ties are broken in a systematic manner. The fundamental Voronoi region of Λ is the set of points in $\mathbb{R}^n$ closest to the zero codeword, i.e., $\mathcal{V}_0 = \{\mathbf{x} : \ Q_\Lambda(\mathbf{x}) = \mathbf{0}\}$. The Voronoi region associated with each $\lambda \in \Lambda$ is the set of points $\mathbf{x}$ such that $Q_\Lambda(\mathbf{x}) = \lambda$, and is given by a shift of $\mathcal{V}_0$ by λ.

Other fundamental regions $\mathcal{P}_0$ exist which generate a lattice partition of the form $\{\lambda + \mathcal{P}_0\}_{\lambda \in \Lambda}$, and a corresponding lattice quantizer

$$Q_{\Lambda, \mathcal{P}_0}(\mathbf{x}) = \lambda \ \text{ if } \ \mathbf{x} \in (\lambda + \mathcal{P}_0). \tag{5}$$

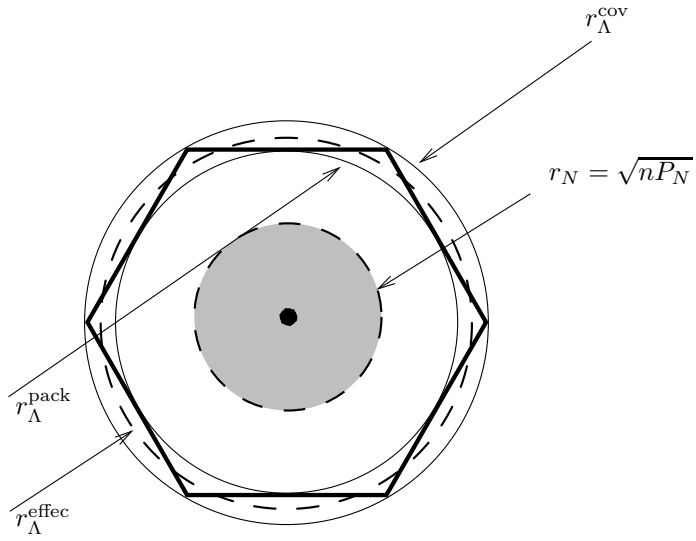

Fig. 3: The fundamental Voronoi region and its packing radius, covering radius and effective radius (radius of the sphere having the same volume). Packing and covering efficiencies are measured by the corresponding ratios.

For example, the fundamental parallelotope $\{G\boldsymbol{\alpha} : \ 0 \leq \alpha_i < 1, \ i = 1 \ldots n\}$ amounts to transforming the unit cube (the fundamental region of $\mathbb{Z}^n$) by the generator matrix G. Nevertheless, the volume of *all* fundamental regions of Λ is the same, and is given by $|\det(G)| \stackrel{\Delta}{=} V_\Lambda$.

The modulo-Λ operation w.r.t. the lattice Λ and some assumed fundamental region $\mathcal{P}_0$ in (5) is defined as

$$\mathbf{x} \ \mathrm{mod}_{\, \mathcal{P}_0} \ \Lambda = \mathbf{x} - Q_{\Lambda, \mathcal{P}_0}(\mathbf{x}) \tag{6}$$

which is also the quantization error of $\mathbf{x}$ with respect to Λ.

The two most well studied figures of merit of a lattice are its packing radius and covering radius, illustrated in Fig. 3. Here we will focus on two other figures of merit which have more of an engineering flavor: the *normalized second moment*, which is a measure of goodness for quantization, and the *volume to noise ratio*, which is a measure of goodness for AWGN channel coding.

Mean-squared error (MSE) quantization: The second moment σ_Λ^2 of a lattice is defined as the second moment per dimension of a uniform distribution over the fundamental Voronoi region $\mathcal{V}_0$,

$$\sigma_\Lambda^2 = \frac{1}{V_\Lambda} \cdot \frac{1}{n} \int_{\mathcal{V}_0} \|\mathbf{x}\|^2 d\mathbf{x}. \tag{7}$$

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

$$G(\Lambda) = \frac{\sigma_\Lambda^2}{V_\Lambda^{2/n}}. \tag{8}$$

The minimum possible value of $G(\Lambda_n)$ over all lattices in $\mathbb{R}^n$ is denoted by G_n. The normalized second moment of a sphere, denoted by G_n^*, approaches $\frac{1}{2\pi e}$ as the dimension

n goes to infinity. The isoperimetric inequality implies that $G_n > G_n^* > \frac{1}{2\pi e}$ for all n. We also have $G_n \le G_1 = G(\mathbb{Z}) = \frac{1}{12}$.

The operational significance of this figure of merit comes from classical results in high-resolution quantization theory. It is also useful in the context of *constellation shaping*, as we shall see in Sec. V. A result due to Poltyrev which appeared in [50] states that the sequence G_n achieves the sphere lower bound, i.e.,

$$\lim_{n\to\infty} G_n = \frac{1}{2\pi e}. \tag{9}$$

Another result in [50] is that the quantization noise of a lattice achieving G_n is "white", i.e., the covariance matrix of a uniform distribution over $\mathcal{V}_0$ is given by $\sigma_\Lambda^2 \cdot I$, where I is the identity matrix.

Coding for the unconstrained AWGN Channel: The AWGN channel is given by the input/output relation

$$Y = X + Z \tag{10}$$

where Z is i.i.d. Gaussian noise of variance σ_z^2. We denote by $\mathbf{Z}$ an i.i.d. vector of length n of noise random variables.

The notion of lattices which are good for AWGN coding may be defined using Poltyrev's [42] definition of capacity per unit volume of *unconstrained* channels, allowing to separate the "granular" properties of the lattice as a good channel code from the issue of shaping (to meet the power constraint). The probability of decoding error in this setup is the probability that the noise leaves the Voronoi region of the transmitted lattice point

$$P_e = \Pr\{\mathbf{Z} \notin \mathcal{V}_0\}. \tag{11}$$

The volume-to-noise ratio (VNR) of a lattice at probability of error P_e is defined as the dimensionless number

$$\mu(\Lambda, P_e) = \frac{V_\Lambda^{2/n}}{\sigma_z^2} \tag{12}$$

where σ_z^2 is such that (11) is satisfied with equality [17]. Note that for fixed P_e, the VNR is invariant to scaling of the lattice. The minimum possible value of $\mu(\Lambda, P_e)$ over all lattices in $\mathbb{R}^n$ is denoted by $\mu_n(P_e)$. The VNR of a sphere is denoted $\mu_n^*(P_e)$. Since a sphere supports the isotropic vector $\mathbf{Z}$ better than any shape of the same volume (see the *sphere bound* of [17]), we have $\mu_n(P_e) > \mu_n^*(P_e) > 2\pi e$, where the second inequality holds for all sufficiently small P_e, and $\mu_n^*(P_e) \to 2\pi e$ as $n \to \infty$, for all $P_e > 0$. It follows from Poltyrev (see also [16], [17]) that the sequence of minimum possible VNRs asymptotically achieve this lower bound:

$$\lim_{n\to\infty} \mu_n(P_e) = 2\pi e, \quad \text{for all } 0 < P_e < 1. \tag{13}$$

In fact, simultaneous goodness in *both* senses (9) and (13) above is asymptotically possible.

Theorem 1. [10] *There exists a sequence Λ_n of lattices of increasing dimension n, which satisfies*

$$G(\Lambda_n) \to \frac{1}{2\pi e} \quad \textit{and} \quad \mu(\Lambda_n, P_e) \to 2\pi e.$$

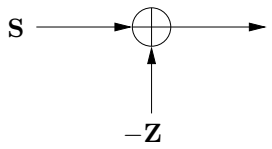

Fig. 4: Equivalent additive-noise channel of a dithered lattice quantizer. ($\mathbf{Z}$ is independent of the input $\mathbf{S}$, and uniform over the fundamental region $\mathcal{P}_0$ of Λ.)

It is also shown in [10] that these lattices achieve the Minkowski and Rogers bounds for sphere packing and covering, and the Poltyrev exponent of the unconstrained AWGN channel.

IV. Dithered Quantization

In quantization theory (as well as in some non-linear processing systems) the term "dithering" corresponds to intentional randomization, aimed to improve the perceptual effect of the quantization, e.g. to reduce "blocking" effects in picture coding. Dithered quantization is also an effective means to guarantee a desired distortion level, independent of the source statistics.

We say that $\mathbf{U}$ is a "subtractive dither" if it is known at both the encoder and the decoder (i.e., it is a common randomness), the encoder adds it to the source vector $\mathbf{s}$ prior to the quantization, while the decoder subtracts it from the quantized value, so the overall reconstruction is $Q_\Lambda(\mathbf{s}+\mathbf{U}) - \mathbf{U}$. Addition and subtraction of $\mathbf{u}$ before and after quantization amounts to shifting the quantizer by $-\mathbf{u}$. Since the lattice quantizer $Q_\Lambda(\cdot)$ is periodic in space, a random shift $\mathbf{U}$ which is uniform over the lattice period makes the quantization error uniform as well.

Theorem 2. [50], [55] *Let $\mathbf{U}$ be uniform over the fundamental region $\mathcal{P}_0$ of the lattice quantizer (5). Then, the quantization error $Q_{\Lambda,\mathcal{P}_0}(\mathbf{s}+\mathbf{U}) - \mathbf{U} - \mathbf{s}$ is uniform over $-\mathcal{P}_0$, the reflection of $\mathcal{P}_0$, independent of the source vector $\mathbf{s}$.*[2]

Equivalently, $(\mathbf{s}+\mathbf{U}) \bmod_{\mathcal{P}_0} \Lambda$ is uniform over $\mathcal{P}_0$ for any $\mathbf{s}$, a result termed the "Crypto Lemma" by Forney [18].

As a corollary of Theorem 2 and (7), the mean-squared distortion of a Voronoi dithered quantizer (4) is equal to the lattice second moment:

$$\frac{1}{n} E\|Q_\Lambda(\mathbf{s}+\mathbf{U}) - \mathbf{U} - \mathbf{s}\|^2 = \sigma_\Lambda^2 \tag{14}$$

independent of the source vector $\mathbf{s}$.

In high-resolution quantization theory it is common to approximate the quantization process as adding (independent) noise to the source. Theorem 2 shows that for dithered quantization this model is exact at *any resolution*. See Fig. 4.

[2]Thm. 2 still holds if $\mathbf{U}$ is replaced by a "generalized dither", i.e., any vector $\tilde{\mathbf{U}}$ such that $(\tilde{\mathbf{U}} \bmod \Lambda)$ is uniform over $\mathcal{P}_0$ [55].

A. Entropy Coded Dithered Quantization

The next theorem makes the connection to an additive-noise channel even stronger. Assume that for given source statistics, the lattice quantizer output is losslessly "entropy" coded, conditioned on the dither value. That is, each lattice point is mapped into a binary word of variable length, such that the average code length is approximately equal to the conditional entropy of the quantizer output. We call such a combination of a lattice quantizer and optimum lossless encoding an Entropy-Coded Dithered Quantizer (ECDQ).

Theorem 3. [48] *The average code length of the ECDQ, i.e., the conditional entropy of the dithered lattice quantizer, is equal to the mutual information in the equivalent additive-noise channel of Fig. 4:*

$$H(Q_\Lambda(\mathbf{S}+\mathbf{U})|\mathbf{U}) = I(\mathbf{S};\mathbf{S}-\mathbf{U}). \qquad (15)$$

The mutual information formula above resembles the expression for Shannon's rate-distortion function [9]: $R(D) = \inf_{\hat{S}:\ E\{(S-\hat{S})^2\}\le D} I(S;\hat{S})$. This formal resemblance leads to a *universal* bound on the loss of the ECDQ.

Theorem 4. [57], [48] *For any source* $\mathbf{S}$*, the redundancy of the ECDQ above the rate-distortion function under a squared error distortion measure is at most*

$$H(Q_\Lambda(\mathbf{S}+\mathbf{U})|\mathbf{U}) - R(D) \le \frac{1}{2} + \frac{1}{2}\log(2\pi e G(\Lambda)) \quad (16)$$

bits, and it is only $\frac{1}{2}\log(2\pi e G(\Lambda))$ *in the limit as* D *goes to zero (i.e., at high-resolution conditions).*

Divergence of dither from Gaussianity: The second term on the right hand side above can be interpreted as the divergence (or "Kullback-Leibler distance") of the dither distribution from AWGN:

$$\frac{1}{2}\log(2\pi e G(\Lambda)) = \frac{1}{n} D(\mathbf{U}\|\mathbf{U}^*) \qquad (17)$$

where $\mathbf{U}^*$ is a zero-mean i.i.d. Gaussian vector with $\mathrm{Var}(U_i) = \sigma_\Lambda^2$ for all i, and where $D(\cdot\|\cdot)$ denotes divergence [9], [50]. Thus, for lattices which are good for quantization, i.e., $\lim_{n\to\infty} G(\Lambda_n) = \frac{1}{2\pi e}$, the divergence of the dither from Gaussianity (17) goes to zero, so the equivalent channel of Fig. 4 becomes an AWGN channel.

B. Filtered ECDQ

Consider the equivalent additive-noise channel model in Fig. 4. As discussed earlier, for any finite dimension the noise of optimal quantization lattices is *white* [50]. If the second order statistics of the source are also known, then we can use Wiener linear estimation principles to reduce the overall MSE in reconstructing the source $\mathbf{S}$. The improvement is most dramatic when the source is Gaussian.

If the source is white, then the Wiener filter is a simple scalar coefficient β at the output of the equivalent channel. For such a source the reconstruction becomes $\hat{\mathbf{S}} = \beta[Q_\Lambda(\mathbf{S}+\mathbf{U}) - \mathbf{U}]$, where $\beta = \frac{\sigma_S^2}{\sigma_S^2+\sigma_\Lambda^2}$, and the overall distortion $D = E\|\hat{\mathbf{S}} - \mathbf{S}\|^2$ decreases from σ_Λ^2 to $D = \left(1/\sigma_S^2 + 1/\sigma_\Lambda^2\right)^{-1}$. This reduction in distortion of the "post-filtered" ECDQ allows us to improve the bound of Thm. 4 in the Gaussian source case.

Theorem 5. [49] *For a Gaussian source with variance* σ_S^2*, the redundancy of the post-filtered ECDQ over the rate-distortion function* $R^*(D) = \frac{1}{2}\log\left(\frac{\sigma_S^2}{D}\right)$ *is at most*

$$H(Q_\Lambda(\mathbf{S}^*+\mathbf{U})|\mathbf{U}) - R^*(D) \le \frac{1}{2}\log(2\pi e G(\Lambda)) \quad (18)$$

for all *distortion levels* $0 < D \le \sigma_S^2$.

See [49] for the extension of this concept to sources with *memory* using pre/post-filters.

Note that the output scaling factor β is smaller than one for the entire distortion range $0 < D \le \sigma_S^2$. Since the reconstruction $\hat{\mathbf{S}}$ belongs to $\beta\Lambda$ (up to a shift due to the dither), it follows that the decoding lattice $\beta\Lambda$ is a "deflated" version of the encoding lattice Λ. More on the meaning of this encoding-decoding "mismatch" in the next section.

V. Voronoi Codebooks

As Information Theory shows us, coding for Gaussian sources and channels should be done using "Gaussian codebooks". That is, the codewords should be selected from a Gaussian generating distribution. The number of codewords is determined by the target rate, while the generating distribution is white, and its variance is equal to the source variance - in source coding, and to the transmitter power - in channel coding. The resulting codebook in $\mathbb{R}^n$ (n being the code dimension) has roughly uniformly distributed codewords over a *sphere*. Can we replace a Gaussian codebook by a lattice code?

In the ECDQ system discussed above, the codebook was the whole (unbounded) lattice and *not shaped* to fit the source variance. The lack of shaping is compensated for by entropy coding, which amounts to "soft" shaping: the lattice points which fall inside the typical (spherical) source region get a shorter binary representation, and dominate the coding rate, while the contribution of the points outside this region is negligible. A similar situation occurs in channel coding with *probabilistic shaping* [28], or alternatively, in unconstrained channels [42]. In fixed-length source coding or power-constrained channel coding, however, the codebook must be bounded.

In this section we describe a lattice codebook, whose codewords and shaping region both have a lattice structure. The construction is based on the notion of nested lattices, [51], [52], [11], [13], which has its roots in de Buda's spherical lattice codes [4], [30] and Forney's Voronoi constellations [14], [15], and owe its development to the search for structured binning schemes for side information problems; see the next section.

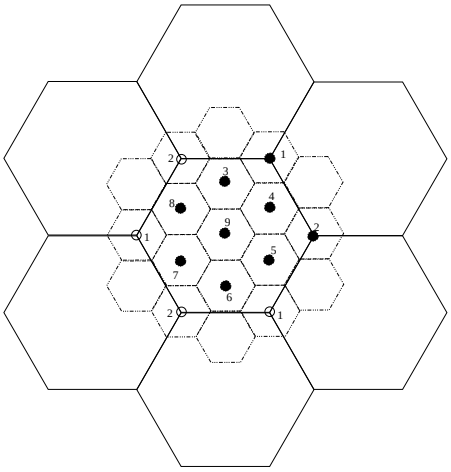

Fig. 5: Nested lattices: special case of self similar lattices.

A. Nested Lattices

A pair of n-dimensional lattices (Λ_1, Λ_2) is called nested if $\Lambda_2 \subset \Lambda_1$, i.e., there exists corresponding generator matrices G_1 and G_2, such that

$$G_2 = G_1 \cdot \mathbf{J} \ ,$$

where $\mathbf{J}$ is an $n \times n$ integer matrix whose determinant is greater than one. We call Λ_1 the *fine lattice* and Λ_2 the *coarse lattice*. The cell volumes of Λ_1 and Λ_2 satisfy

$$V_{\Lambda_2} = |\det(\mathbf{J})| \cdot V_{\Lambda_1}. \tag{19}$$

We call $\sqrt[n]{|\det(\mathbf{J})|} = \sqrt[n]{V_{\Lambda_2}/V_{\Lambda_1}}$ the *nesting ratio*.

Fig. 5 shows nested hexagonal lattices with $\mathbf{J} = 3 \cdot I$, where I is the 2×2 identify matrix. This is an example of the important special case of *self similar lattices*, where Λ_2 is a scaled – and possibly reflected or rotated – version of Λ_1.

For some fundamental region $\mathcal{P}_{0,2}$ of Λ_2, the points of the set

$$\Lambda_1 \bmod \Lambda_2 \stackrel{\triangle}{=} \Lambda_1 \cap \mathcal{P}_{0,2} \tag{20}$$

are called the *coset leaders* of Λ_2 relative to Λ_1; for each $v \in \{\Lambda_1 \mod \Lambda_2\}$ the shifted lattice $\Lambda_{2,v} = v + \Lambda_2$ is called a *coset* of Λ_2 relative to Λ_1. It follows that there are $V_{\Lambda_2}/V_{\Lambda_1} = |\det(J)|$ different cosets.

If $\mathcal{P}_{0,2}$ in (20) is the fundamental Voronoi region $\mathcal{V}_{0,2}$ of Λ_2, then we obtain a Voronoi constellation [14], [15]. In the example of Fig. 5, the Voronoi constellation consists of the bold points. A parallelepiped region $\mathcal{P}_{0,2}$ is preferable, however, if we wish to simplify coset enumeration [56].

Dithered Voronoi codebook: A dithered Voronoi codebook consists of all shifted fine lattice points $\lambda + \mathbf{u}$, for $\lambda \in \Lambda_1$, inside the Voronoi region of the coarse lattice Λ_2, i.e.,

$$(\mathbf{u} + \Lambda_1) \bmod \Lambda_2 \tag{21}$$

where the dither $\mathbf{u}$ is an arbitrary vector in $\mathbb{R}^n$ to be specified later. (For $\mathbf{u} = 0$ this is the set of relative coset leaders in (20).) The size of this codebook is $V_{\Lambda_2}/V_{\Lambda_1}$ (independent of $\mathbf{u}$), so the associated coding rate is

$$R = \frac{1}{n} \log_2(V_{\Lambda_2}/V_{\Lambda_1})$$

bits per dimension.

Existence of good nested lattices: The existence of a sequence of good pairs of nested lattices, where one of the lattices (the fine one or the coarse one) is good for AWGN channel coding, while the other lattice is good for source coding under mean-squared distortion, is addressed in [10]. See [27] for an extension. The key to proving the existence of such lattices is to consider an appropriate *random ensemble* of lattices. An ensemble based on *generalized construction A* was defined in [32], while the Minkowski-Hlawka-Siegel ensemble is considered in [42], [21], [56].

B. Achieving the AWGN Channel Capacity

We now show an efficient coding scheme for the AWGN channel $Y = X + Z$ of (10) using a pair of nested lattices $\Lambda_2 \subset \Lambda_1$. In this scheme Λ_2 (the coarse lattice) is used for *shaping* while Λ_1 (the fine lattice) is used for *coding*.

Let the dither $\mathbf{U}$ be uniform over a fundamental region of Λ_2 (or a generalized dither as mentioned earlier), and let $\mathbf{v}$ be any codeword (or coset leader) in $\Lambda_1 \bmod_{\mathcal{P}_0} \Lambda_2$, with $\bmod \Lambda_2$ w.r.t. a "convenient" enumeration fundamental region $\mathcal{P}_0$.

To transmit the message $\mathbf{v}$, the encoder outputs

$$\mathbf{X} = (\mathbf{v} + \mathbf{U}) \bmod_{\mathcal{V}_0} \Lambda_2$$

with $\bmod \Lambda_2$ now performed w.r.t. the fundamental Voronoi region $\mathcal{V}_0$. By (14) we have that $E\{\|\mathbf{X}\|^2\} = \sigma^2_{\Lambda_2}$. Thus if we chose a lattice with second moment $\sigma^2_{\Lambda_2} = P$, then each codeword satisfies the power constraint (on the average with respect to the dither).

The decoder first linearly estimates the vector $\mathbf{v}$ by

$$\hat{\mathbf{Y}} = \alpha \mathbf{Y} - \mathbf{U} \tag{22}$$

(where $0 < \alpha \leq 1$ is a coefficient to be determined later). Then, it quantizes $\hat{\mathbf{Y}}$ to the nearest codeword modulo the codebook. The decoded message is thus $\hat{\mathbf{V}} = Q_{\Lambda_1}(\hat{\mathbf{Y}}) \bmod_{\mathcal{P}_0} \Lambda_2$. This is equivalent to

$$\hat{\mathbf{V}} = \left[\alpha \cdot Q_{\frac{\Lambda_1}{\alpha}}\left(\mathbf{Y} - \frac{\mathbf{U}}{\alpha}\right)\right] \bmod_{\mathcal{P}_0} \Lambda_2, \tag{23}$$

i.e., to decoding with respect to the *inflated* lattice $\frac{\Lambda_1}{\alpha}$. (Note the resemblance to the deflated lattice $\beta\Lambda$ in Sec. IV-B.)

The equivalent channel from the codeword $\mathbf{v}$ to the modulo estimation vector $\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} \bmod_{\mathcal{P}_0} \Lambda_2$ is called a *modulo-lattice transformation* [11]. The distributive law of the modulo operation[3] and Thm. 2 imply:

[3] $((a \bmod_{\mathcal{V}_0} \Lambda_2) + b) \bmod_{\mathcal{P}_0} \Lambda_2 = (a + b) \bmod_{\mathcal{P}_0} \Lambda_2$.

Theorem 6. (Effective modulo-Λ additive-noise channel) [11] *The channel from* $\mathbf{v}$ *to* $\tilde{\mathbf{Y}}$ *is equivalent in distribution to the modulo additive-noise channel*

$$\tilde{\mathbf{Y}} = \Big(\mathbf{v} + \mathbf{Z}_{\text{eff}}\Big) \bmod \mathcal{P}_0\Lambda_2$$

where the effective noise is given by

$$\mathbf{Z}_{\text{eff}} = [\alpha\mathbf{Z} + (1-\alpha)\mathbf{U}'] \bmod \mathcal{P}_0\Lambda_2 \tag{24}$$

and where $\mathbf{U}'$ *is uniform over* $\mathcal{V}_0$ *and independent of* $\mathbf{v}$ *and* $\mathbf{Z}$*.*

Note that the *effective (additive) noise* $\mathbf{Z}_{\text{eff}}$ is a weighted combination of two components: AWGN and a dither component, where the latter is called "self noise" because it comes from the coarse lattice.

For a modulo additive-noise channel, a uniform input $\mathbf{V} \sim \text{Unif}(\mathcal{P}_0)$ maximizes the mutual information $I(\mathbf{V};\tilde{\mathbf{Y}})$, which becomes $\log(V_{\Lambda_2}) - h(\mathbf{Z}_{\text{eff}})$. The optimum α is thus the one that minimizes the entropy of the effective noise.[4] As the lattice dimension increases, the self noise $\mathbf{U}'$ and therefore the effective noise $\mathbf{Z}_{\text{eff}}$ become closer to a Gaussian distribution (in the divergence sense (17)), in which case minimizing entropy amounts to minimizing variance. Thus the optimum α becomes the Wiener coefficient $\alpha = \sigma^2_{\Lambda_2}/(\sigma^2_{\Lambda_2} + N) = \frac{P}{P+N}$, and the resulting noise variance is the MMSE solution

$$\text{Var}(\mathbf{Z}_{\text{eff}}) = \frac{PN}{P+N}. \tag{25}$$

Due to the dither, the error probability is *identical* for all codewords (as reflected by the equivalent modulo-additive channel of Thm. 6), and is equal to

$$P_e = \Pr\{\mathbf{Z}_{\text{eff}} \notin \mathcal{V}_{0,1}\}. \tag{26}$$

Thus, by the definition of the the VNR (12), if we target some P_e, the volume of the fine lattice cell must be $V_{\Lambda_1} \approx [\mu(\Lambda_1, P_e) \cdot \text{Var}(\mathbf{Z}_{\text{eff}})]^{n/2}$ or larger, where we assumed a Gaussian $\mathbf{Z}_{\text{eff}}$.[5] On the other hand, the power constraint implies that the volume of the coarse cell is $V_{\Lambda_2} = [P/G(\Lambda_2)]^{n/2}$ or smaller. For the MMSE solution (25), we thus get a coding rate of

$$R = \frac{1}{n}\log\left(\frac{V_{\Lambda_2}}{V_{\Lambda_1}}\right) \approx \frac{1}{2}\log\left(\frac{P/G(\Lambda_2)}{\mu(\Lambda_1,P_e)\text{Var}(\mathbf{Z}_{\text{eff}})}\right) \tag{27}$$

$$= C - \frac{1}{2}\log\Big(G(\Lambda_2)\cdot\mu(\Lambda_1,P_e)\Big) \tag{28}$$

where $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ is the AWGN channel capacity.

The capacity loss in (28) is approximately the NSM-VNR cross product of the lattice pair. To reduce this loss, we need the coarse lattice to be a "good" quantizer, while the fine lattice should be a "good" AWGN channel code, both in the sense of Sec. III. For such a good pair of nested lattices $G(\Lambda_2) \to 1/2\pi e$ and $\mu(\Lambda_1, P_e) \to 2\pi e$ as $n \to \infty$, so the system approaches the AWGN channel capacity. An analysis of the error exponent of Voronoi codebooks can be found in [31], [44].

[4]For rates below capacity, a smaller α would give better error performance [31], [44].

[5]This assumption is true for high SNR (implying $\alpha = 1$), or high dimension and a "good" coarse lattice (to make the self-noise component "Gaussian enough"). Furthermore, the effective noise $\mathbf{Z}_{\text{eff}}$ is in fact more favorable than Gaussian noise for sufficiently small P_e, so the Gaussian approximation provides a *lower bound* on the rate of the system.

C. Achieving the Gaussian RDF

A dual construction of a *Voronoi quantizer* achieving the quadratic-Gaussian (QG) rate-distortion function can be designed along similar lines. Again, the NSM-VNR cross product of the lattice pair - now with the roles of Λ_1 and Λ_2 switched relative to (28) - will determine the rate loss of the system. The coarse lattice should therefore be a "good" AWGN channel code, while the fine lattice should be a "good" quantizer [54].

VI. Side-Information Problems

Classical Information Theory deals with point-to-point communication, where a single source is transmitted over a channel to a single destination. In a distributed situation there may be more than one (possibly correlated) sources, hence more than one encoder, and/or more destinations, hence more than one channel output and decoder. The simplest situation, which captures much of the essence in the problem, are sources and channels with side information.

In the source version of the problem - solved by Wyner and Ziv [47] - a source S is encoded knowing that a correlated signal J is available at the decoder (but not at the encoder). In the Gaussian case, we assume that $S = J + Q$, where Q is a white Gaussian source independent of J.

The channel version of the problem was solved by Gelfand and Pinsker in [19]. It assumes that the input to a state-dependent channel is encoded knowing the channel states non-casually. The decoding is done solely based on the channel output, without having access to the channel states. In the special case known as the "dirty paper" channel (DPC), or the *Costa problem*, the input-output relation is $Y = X + I + Z$, where I is an interference signal known at the encoder, and Z (the unknown noise) is AWGN [8].

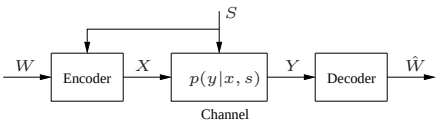

Fig. 6: A channel with side-information at the transmitter.

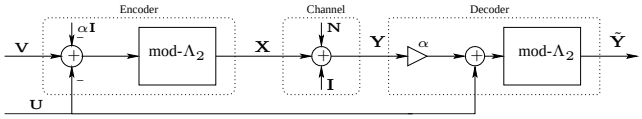

Fig. 7: Lattice-strategies for the dirty-paper channel.

An interesting feature of Gaussian side-information problems is that their information-theoretic solutions

amount to complete elimination of the effect of the partially known signals J and I.

For the DPC problem, a simple variation on the Voronoi modulation and decoding system of Sec. V-B achieves the same coding rate as in (28), where now $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ denotes the "clean" AWGN channel capacity [11], [39]. The main change is the subtraction of the scaled interference αI modulo the coarse lattice - see Fig. 7. (For a scalar-lattice solution for the *causal* DPC problem - see [11], [45].) The Gaussian Wyner-Ziv (WZ) problem is solved by a similar variation on the Voronoi quantization scheme of Sec. V-C [52]. In both DPC and WZ variations, the cosets of Λ_2 relative to Λ_1 (20) replace the *random bins* of the classical solutions of [19], [47].

A nice benefit of the dithered lattice approach is that the known parts (J and I) can be arbitrary signals, i.e., they do not even need to have a stochastic model. Yet, if J and I *are* random, then they can play the role of the dither, so common randomness becomes unnecessary.

See [22] for a *modulo lattice modulation* (MLM) scheme for *joint* source-channel coding with side-information using a *single* shaping lattice.

VII. Gaussian Networks

There are many ways in which side-information paradigms can enter general multi-terminal networks. The obvious cases are the broadcast channel, in which the (joint) encoder may view the transmission to one terminal as side-information for the transmission to the other terminals. Similarly, in multi-terminal coding of correlated sources, the (joint) decoder may view the reconstruction of one source as side information for the reconstruction of the other sources. In both these cases, the side-information is concentrated in the "relevant" terminal in the network. Indeed, in the QG case, it is easy to figure out how to replace the standard information-theoretic "random binning" technique by a lattice-based solution. This solution uses the the lattice-WZ and lattice-DPC schemes of Sec. VI as building blocks [52]. As in section VI, the main motivation for such a lattice scheme is the complexity reduction (and perhaps the intuition) gained by a structured solution.

A more interesting situation, however, occurs when side information is *distributed* among more than one terminal. Surprisingly, it turns out that in some distributed linear network topologies, the lattice-based system *outperforms* the random-binning solution. Moreover, in some cases it is in fact optimal! Apparently, the linearity of the network in these scenarios favors linear (rather than random) binning, as we already saw in the binary Körner-Marton problem.

A. The Gaussian Körner-Marton Problem

Krithivasan and Pradhan [26] extended the Körner-Marton problem of Fig. 2 to the QG case. Suppose X and Y are positively correlated Gaussian sources, say, $Y = X + N$ where N is independent of X, and the decoder wants to reconstruct their difference N with some mean-squared distortion D. As they show, near-optimal performance can be achieved if each source is *lattice-WZ* encoded, where the coarse lattice - tuned to match the variance of the difference N - is *identical* at both encoders. The decoder subtracts the two encodings, modulo the coarse lattice, to isolate the desired (quantized) difference signal.

Unlike the original "lossless" KM setup, however, the lattice scheme does not match the "genie aided" outer bound; for $\sigma_x^2 \gg \sigma_n^2$, it loses 3dB in distortion (one bit in the sum rate) due to the accumulation of two independent quantization noises. Yet, at least for high rates this is still better than a "standard" random binning solution *a la* Berger-Tung [3], which (implicitly) encodes both sources X and Y just to transmit their difference.

B. The Dirty Multiple Access Channel

We next consider what seems to be the "dual" of the Körner-Marton problem: a generalization of the Gaussian dirty-paper problem to a multiple access setup, as illustrated in Fig. 8. There are two additive-interference signals, one known to each transmitter but none to the receiver.

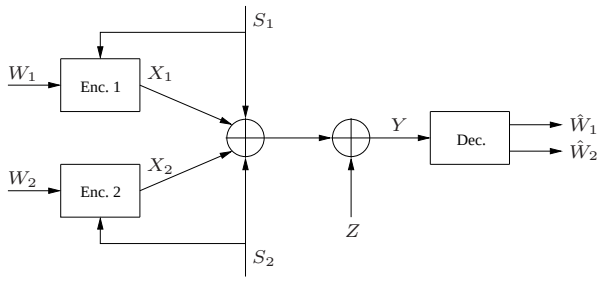

Fig. 8: Doubly dirty MAC.

It is shown in [40] that the rates achievable using Costa's binning scheme (induced by his auxiliary random variables) vanish in the limit when the interference signals are strong. In contrast, if both encoders apply lattice-DPC using the *same* shaping (coarse) lattice Λ_s, then the sum interference is concentrated on Λ_s. The equivalent channel seen by the receiver is thus a MAC version of the modulo-additive channel of Thm. 6, and the sum rate is positive *independent* of the interferences.

Furthermore, [40] gives an outer bound for the capacity region of the dirty MAC for arbitrarily strong interferences, which is strictly smaller than the clean MAC capacity region. Lattice-DPC of large dimension meets this outer bound for some cases, in particular for *imbalanced* power constraints, as well as in the limit of high SNR [40].[6]

C. The Loss of Single-Letter Characterization

Costa's binning scheme is derived from a Gaussian single-letter formula. It fails on the dirty-MAC because,

[6]The loss w.r.t. the outer bound in the *balanced* case is similar to the 3dB loss in the Gaussian KM; it amounts to doubling the "self noise" component in (24), hence the "1" in the AWGN channel capacity formula reduces to some number $1 > \gamma > 1/2$ [40], [37].

unlike for lattice-binning, the sum of two independent bins (from the two users) results in a "bad" codebook. A similar phenomena occurs in the Gaussian Körner-Marton problem: the *difference* of two independent bins, each one generated by a Gaussian single-letter expression, results in a "bad" codebook. Are there better single-letter formulas for these two problems?

We conjecture that the best single-letter formula for the dirty MAC in the limit of strong interference and high SNR is given in terms of a one-dimensional lattice [40], [41]. The resulting rate loss is thus the "shaping gain" $\frac{1}{2}\log(2\pi e/12) \approx 0.254$ bits, i.e., the divergence from Gaussianity of a scalar dither (17). For a binary version of the dirty MAC, it is shown in [41] that the capacity loss of the best known single-letter formula is ~ 0.2 bits.

D. Lattice Network Coding

In a standard packet switching network, nodes act as routers - they wish to find the best route for a packet under the current conditions. If the inflow to a node is higher than its output capacity, then some of the packets need to be discarded. The idea of network coding is that a bottleneck node can "combine" together packets rather than choose which one to pass on and which one to discard. If the final destination gets enough such "combinations" (from different routes), then it can resolve the ambiguity and decode all the transmitted packets reliably.

The focus of most research on network coding has been on *linear* coding schemes [29]. In theory, though, any mapping at the nodes which is overall information preserving would work, as long as the network is lossless. In particular, random binning at the nodes is information preserving with high probability [20]. However, when extending the network coding idea to *noisy* networks, the structure of the code is essential to avoid *noise accumulation* and loss of capacity.

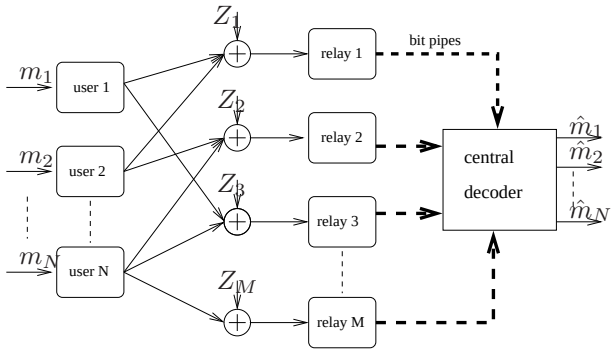

Fig. 9: A multi-relay multi-user network scenario.

Specifically, consider the Gaussian relay network proposed in [36], depicted in Fig. 9, where N users wish to communicate with a destination (central decoder) through a layer of $M \geq N$ relays. Each relay receives some weighted (by the fading coefficients) linear combination of the transmitted signals corrupted by AWGN. Thus, the different signals at the relay input are already "combined" by the network. Relaying this combination as is (say, in some analog or compressed form) means that the noise will be forwarded to the final receiver as well. On the other hand, requiring the relay to decode all its input signals *separately* (as a MAC receiver) means a waste of capacity. See, e.g., [2].

It has been shown recently how to use lattice codes for ("physical-layer") network coding in the presence of Gaussian noise [35], [34], [36], [37]. If all the users use the *same* coding (fine) lattice, then the relay can decode an integer linear combination of the codewords (a lattice point which is close to the received signal), thus removing the channel noise before forwarding the decoded point to the final receiver. A particularly insightful example is that of the *two-way relay*, where each user computes its intended message from its own message and the message-sum it gets from the relay [35], [34].

A framework for treating non-Gaussian noise and non-additive channels is proposed in [12].

E. Interference Alignment

A similar idea applies for the suppression of interference in a multi-node interference channel (IC). One of the interesting observations of the recent years is the idea of interference alignment [5]: a channel aware transmission system can make the effective number of interferers seen by each receiver equal to one. Thus, effectively, the multi-node IC is no worse than the classic double-node IC!

The original idea was to align the interference in the time domain, and it used linear transformations [33]. An alternative approach, based on alignment in the amplitude domain, was proposed in [38]. This approach fits very naturally into the lattice framework.

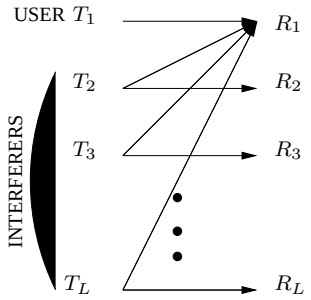

Fig. 10: Many-to-one interference alignment.

Consider the many-to-one interference channel of Fig. 10. Assuming the interference path gains of users 2 to L are identical, and that these users use the *same* coding (fine) lattice Λ_I, the equivalent channel seen by user 1 is similar to that seen in the dirty MAC of Sec. VII-B: the interference signals are all *concentrated* on the points of a *single* lattice Λ_I. Thus, in effect, user 1 experiences a single interferer. Furthermore, using an "estimate-and-modulo" receiver as in Thm. 6, user 1 sees an equivalent

modulo-Λ_I channel. Thus, it can achieve a rate of

$$R_1 = \frac{1}{2}\log\left(\min\left\{\frac{\sigma^2_{\Lambda_I}}{N}, \frac{P+N}{N}\right\}\right)$$

for large lattice dimension, corresponding to a full capacity in the strong interference regime.

VIII. Open Questions

On the practical side, lattice (or alternatively, linear trellis) codes with good performance and low encoding and decoding complexity are essential to make this theory attractive. New design approaches, e.g., [43], may be of interest.

The linear structure of the lattice plays a crucial role in the distributed lattice coding schemes presented in Sec. VII. For a proper operation, we need to align the lattice codes both in time and in amplitude. Yet in all the examples we considered, only one of the component codes of the system - either the shaping or the coding lattice - must be aligned. The other code does not even need to be a lattice! Other examples are of interest.

	Shaping (coarse) lattice	Coding (fine) lattice
Gaussian Korner Marton	aligned	–
dirty MAC	aligned	–
Lattice network coding	–	aligned
Interference alignment	–	aligned

Random coding schemes - based on traditional single-letter (i.i.d.) solutions - seem to fail in these setups. For example, as discussed in Sec. VII-C, the loss of single-letter characterization in the Gaussian dirty MAC setup is conjectured to be $\frac{1}{2}\log(2\pi e/12) \approx 0.254$ bits.

Does structure really beat random? Note that proving the existence of good lattices also requires random coding arguments [4], [42], [32], [10], [21], [56]. Also, our analysis of the lattice coding schemes assumes common randomness in the form of a dither. A question thus remains, if the failure of the traditional random coding approach is due to inappropriate single-letter solutions, or to its inherent weakness. We believe the latter to be true.[7]

The success of lattices in these setups hinges upon a good match between the linearity of the code and the linearity of the source or channel network. Can we go beyond the linear case?

Acknowledgement

My ISIT 10 talk was based on past and present work with Meir Feder, Gregory Poltyrev, Toby Berger, Shlomo Shamai, Uri Erez, Simon Litsyn, Dave Forney, Yuval Kochman and Tal Philosof. Thanks are also due to Bobak Nazer for helpful comments on the manuscript, and to Anatoly Khina for his help with the figures.

[7] Nevertheless, this should not be seen as a discouraging fact, but rather as an indication for new directions and opportunities!

References

[1] R. Ahlswede and J. Körner. Source coding with side information and a converse for the degraded broadcast channel. IEEE Trans. Information Theory, vol. 21, pp. 629637, 1975.

[2] S. Avestimehr, S. Diggavi, and D. Tse. Wireless network information flow: A deterministic approach. to appear in *IEEE Trans. Info. Theory* 2011, see http://arxiv.org/abs/0906.5394.

[3] T. Berger. Multiterminal Source Coding. New York: In G.Longo, editor, the Information Theory Approach to Communications, Springer-Verlag, 1977.

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

[5] , V. R. Cadambe, S. A. Jafar. Interference Alignment and the Degrees of Freedom for the K User Interference Channel. *IEEE Trans. Info. Theory*, IT-54(8):3425–3441, Aug 2008.

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

[7] J. H. Conway and N. J. A. Sloane. Voronoi regions of lattices, second moments of polytopes, and quantization. *IEEE Trans. Info. Theory*, IT-28:211–226, Mar. 1982.

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

[9] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[10] U. Erez, S. Litsyn, and R. Zamir. Lattices which are good for (almost) everything. *IEEE Trans. Info. Theory*, IT-51:3401–3416, Oct. 2005.

[11] U. Erez, S. Shamai, and R. Zamir. Capacity and lattice strategies for cancelling known interference. *IEEE Trans. Info. Theory*, IT-51:3820–3833, Nov. 2005.

[12] U. Erez and R. Zamir. A modulo-lattice transformation for multiple-access channels. In *Electrical and Electronics Engineers in Israel, 2008 IEEE 25th Convention of*, Dec. 2008. Also at ITA, UCSD, February 2009.

[13] U. Erez and R. Zamir. Achieving 1/2 log(1+SNR) on the AWGN channel with lattice encoding and decoding. *IEEE Trans. Info. Theory*, IT-50:2293–2314, Oct. 2004.

[14] G. D. Forney Jr. and L. F. Wei. Multidimensional constellations-Part I: Introduction, figures of merit, and generalized cross constellations. *IEEE Jr. on Selected Areas in Comm.*, 7:877–892, Aug. 1989.

[15] G. D. Forney Jr. and L. F. Wei. Multidimensional constellations-Part II: Voronoi constellations. *IEEE Jr. on Selected Areas in Comm.*, 7:941–958, Aug. 1989.

[16] G. D. Forney Jr. On the duality of coding and quantizing. In *DIMACS Ser. Discr. Math. Theory Comp. Sci.*, volume 14, 1993.

[17] G. D. Forney Jr., M.D.Trott, and S.-Y. Chung. Sphere-bound-achieving coset codes and multilevel coset codes. *IEEE Trans. Info. Theory*, IT-46:820–850, May, 2000.

[18] G. D. Forney Jr. On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener. In *41st Annu. Allerton Conf. Communication, Control, and Computing*, pp. 430 - 439, Oct. 2003.

[19] S.I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.

[20] T. Ho, M. Medard, R. Koetter, D.R. Karger, M. Effros, Jun Shi, B. Leong. A Random Linear Network Coding Approach to Multicast. *IEEE Trans. Information Theory* IT-52(10):4413–4430, Oct. 2006.

[21] A. Ingber, M. Feder, R. Zamir, The rate of convergence to the capacity of an infinite constellation. In preparation.

[22] Y. Kochman and R. Zamir. Joint Wyner-Ziv/dirty-paper coding using analog modulo-lattice modulation. *IEEE Trans. Info. Theory*, IT 55(11): 4878-4889, Nov. 2009.

[23] Y. Kochman and R. Zamir. Analog matching of colored sources to colored channels. *IEEE Trans. Info. Theory*, to appear 2011.

[24] Y. Kochman, A. Khina, U. Erez, and R. Zamir. Rematch and forward for parallel relay networks. In *ISIT-2008, Toronto, ON*, pages 767–771, 2008.

[25] J. Körner and K. Marton. How to encode the modulo-two sum of binary sources. em IEEE Trans. Information Theory, vol. IT-25, pp. 219221, March 1979.

[26] D. Krithivasan and S. S. Pradhan. Lattices for distributed source coding: Jointly Gaussian sources and reconstructions of a linear function. *submitted to IEEE Trans. Inform. Theory*, July 2007, arXiv:0707.3461, e-print.

[27] D. Krithivasan and S. S. Pradhan. A proof of the existence of good nested lattices, [Online]. Available: http://www.eecs.umich.edu/techreports/systems/cspl/cspl-384.pdf

[28] F. R. Kschischang and S. Pasupathy, Optimal nonuniform signaling for Gaussian channels. *IEEE Trans. Info. Theory*, IT-39(3):913-929, May, 1993.

[29] Li, S.-Y.R., Yeung, R.W. and Ning Cai. Linear network coding. *IEEE Trans. Inform. Theory*, 49(2):371–381, Feb. 2003.

[30] T. Linder, C. Schlegel, and K. Zeger Corrected proof of de Budas theorem *IEEE Trans. Inform. Theory*, 39(5):1735-1737, Sep. 1993.

[31] T. Liu, P. Moulin, and R. Koetter. On error exponents of modulo lattice additive noise channels. *IEEE Trans. Inform. Theory*, 52:454-471, Feb. 2006.

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

[33] Maddah-Ali, M.A.; Motahari, A.S.; Khandani, A.K. Communication Over MIMO X Channels: Interference Alignment, Decomposition, and Performance Analysis. *IEEE Trans. Info. Theory*, IT-54(8): 3457–3470, Aug. 2008.

[34] W. Nam, S.-Y. Chung, and Y. H. Lee, Capacity bounds for two-way relay channels, in International Zurich Seminar on Communications. In *IZS 2008, Zurich, Switzerland*, March 2008.

[35] K. Narayanan, M. P. Wilson, and A. Sprintson. Joint physical layer coding and network coding for bi-directional relaying. In *45th Annual Allerton Conference*, Monticello, IL, Sept., 2007.

[36] B. Nazer and M. Gastpar, Compute-and-forward: Harnessing interference with structured codes. In *Proceedings of ISIT 2008*, July 6-11,Toronto, Canada.

[37] B. Nazer and M. Gastpar, Reliable Physical Layer Network Coding to appear in *IEEE Proc.* March 2011.

[38] A. Parekh, G. Bresler, and D. Tse. The approximate capacity of the many-to-one and one-to-many gaussian interference channels. In *Proc. of 45th Allerton Conference (Monticello, IL)*, Sep. 2007.

[39] T. Philosof, U. Erez, and R. Zamir. Combined shaping and precoding for interference cancellation at low SNR. In *Proc. IEEE International Symposium on Information Theory, pp. 68, (Yokohama, Japan)*, June 2003.

[40] T. Philosof, A. Khisti, U. Erez, and R. Zamir, Lattice strategies for the dirty multiple access channel. In *Proc. of IEEE International Symposium on Information Theory*, Nice, France, June 2007. Also accepted for publicaton in the *IEEE Trans. Info. Theory*.

[41] T. Philosof, and R. Zamir, On the loss of single letter characterization: the dirty multiple access channel. *IEEE Trans. Info. Theory*, IT 55(6):2442–2454, June 2009.

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

[43] Sommer, N., Feder, M. and Shalvi, O. Low-Density Lattice Codes. *IEEE Trans. Info. Theory*, IT-54(4):1561–1585, April 2008.

[44] C. Swannack, U. Erez, and G. W. Wornell. Reflecting on the AWGN error exponent. In 43rd Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois, Sept. 28- 30, 2005.

[45] F. Willems. Signalling for the Gaussian channel with side information at the transmitter In *Proc. IEEE International Symposium on Information Theory, p. 348, (Sorrento, Italy)*, June 2000.

[46] A. Wyner. On source coding with side information at the decoder. IEEE Trans. Information Theory, vol. IT-21, pp. 294300, 1975.

[47] A.D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory*, IT-22:1–10, Jan., 1976.

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

[49] R. Zamir and M. Feder. Information rates of pre/post filtered dithered quantizers. *IEEE Trans. Info. Theory*, pages 1340–1353, Sep. 1996.

[50] R. Zamir and M. Feder. On lattice quantization noise. *IEEE Trans. Info. Theory*, pages 1152–1159, July 1996.

[51] R. Zamir and S. Shamai. Nested linear / lattice codes for Wyner-Ziv encoding. In *Proceedings of the Information Theory Workshop, Killarney, Ireland*, pages 92–93, June 1998.

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

[53] R. Zamir, Y. Kochman, and U. Erez. Achieving the Gaussian rate distortion function by prediction. *IEEE Trans. Info. Theory*, IT-54:3354–3364, July 2008.

[54] R. Zamir. Lattices are everywhere. *Proc. of ITA 2009*, pp. 392 - 421, U San Diego CA, 8-13 Feb. 2009.

[55] R. Zamir. How to generate a simple dither. In *IEEE convention*, Eilat, Israel, Nov. 2010.

[56] R. Zamir. *Lattice Coding for Signals and Networks*. In preparation.

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