Lattices are Everywhere

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

Abstract

As bees and crystals (and people selling oranges in the market) know it for many years, lattices provide efficient structures for packing, covering, quantization and channel coding. In the recent years, interesting links were found between lattices and coding schemes for multi-terminal networks. This tutorial paper covers close to 20 years of my research in the area; of enjoying the beauty of lattice codes, and discovering their power in dithered quantization, dirty paper coding, Wyner-Ziv DPCM, modulo-lattice modulation, distributed interference cancelation, and more.

I. Introduction

Lattice codes 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 [11], [21]. The effectiveness (as well as the complexity of the solution or the coding effort) usually increases with the spatial dimension; good lattices tend to be "perfect" in all aspects as the dimension goes to infinity. However, for a given dimension, the problems are not equivalent. Figures of merit like thickness, density, normalized second moment (NSM) and volume-to-noise ratio (VNR), characterize how good a given lattice is with respect to each of the various aspects.

Recent developments in the area of Gaussian network information theory generated new solutions based on lattices, and hence new figures of merit [66], [18]. For example, for side-information problems known as the "Wyner-Ziv" source and the "dirty-paper" channel, a nested pair of lattices is needed where one component lattice forms a good channel code while the other component lattice forms a good source code. For joint source-channel coding problems, lattices with a good NSM-VNR product are desired [34].

We review these results, and re-examine the theory of lattice figures of merit in the context of multi-user systems and linear Gaussian networks. We hope that this overview will motivate and guide future research on efficient lattice codes construction.

II. LATTICES AND FIGURES OF MERIT

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

$$\Lambda = \left\{ \boldsymbol{\ell} = G \cdot \mathbf{i} \, : \, \mathbf{i} \in \mathbb{Z}^n \right\} , \tag{1}$$

where $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the $n \times n$ *generator matrix* G is given by $G = [\, \boldsymbol{g}_1 \,|\, \boldsymbol{g}_2 | \, \ldots \, | \, \boldsymbol{g}_n]$. Note that the zero vector is always a lattice point, and that G is not unique for a given Λ. See [11].

A few important notions are associated with a lattice, see [27], [30]. The nearest neighbor quantizer $Q(\cdot)$ associated with Λ is defined by

$$Q(\mathbf{x}) = \boldsymbol{\ell} \in \Lambda \quad \text{ if } \; \| \, \mathbf{x} - \boldsymbol{\ell} \, \| \leq \| \, \mathbf{x} - \boldsymbol{\ell}' \, \| \;\; \forall \, \boldsymbol{\ell}' \in \Lambda \, , \tag{2}$$

where $\| \cdot \|$ denotes Euclidean norm, and ties are broken in a systematic manner. The basic Voronoi cell of Λ is the set of points in $\mathbb{R}^n$ closest to the zero codeword, i.e., $\mathcal{V}_0 = \{ \mathbf{x} \, : \; Q(\mathbf{x}) = \mathbf{0} \}$. The Voronoi cell associated with each $\boldsymbol{\ell} \in \Lambda$ is the set of points $\mathbf{x}$ such that $Q(\mathbf{x}) = \boldsymbol{\ell}$ and it is given by a shift of $\mathcal{V}_0$ by $\boldsymbol{\ell}$. The modulo-Λ operation w.r.t. the lattice is defined as

$$\mathbf{x} \bmod \Lambda = \mathbf{x} - Q(\mathbf{x}) \tag{3}$$

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

The use of high dimensional lattice codes is justified by the existence of asymptotically "good" lattice codes. Lattice "goodness" may take one of several forms [11], [21]. Below we consider four such forms. It is interesting to note that a lattice which is good in one sense need not necessarily be good in the other. Nevertheless, it is shown in [15] that a sequence of lattices exists which is simultaneously good in all four aspects.

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

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

where $\mathcal{B}$ denotes the unit ball. That is, the spheres do not intersect. Define the packing radius r_Λ^{pack} of the lattice by

$$r_\Lambda^{\text{pack}} = \sup\{ r : \;\; \Lambda + r\mathcal{B} \text{ is a packing} \}. \tag{4}$$

Note that r_Λ^{pack} is the radius of the largest n-dimensional ball contained in the Voronoi cell $\mathcal{V}_0$. Denote by r_Λ^{effec} the "effective radius" of the Voronoi region, meaning the radius of a sphere having the same volume, so that r_Λ^{effec} is defined by

$$V_{\mathcal{B}}(r_\Lambda^{\text{effec}}) = \text{Vol}(\mathcal{V}) \stackrel{\triangle}{=} V \tag{5}$$

where $V_{\mathcal{B}}(r)$ denotes the volume of a sphere of radius r. Figure 1 gives the geometric picture of r_{pack} and r_Λ^{effec} with respect to the Voronoi region, as well as the other radii to be defined below. Define the packing efficiency

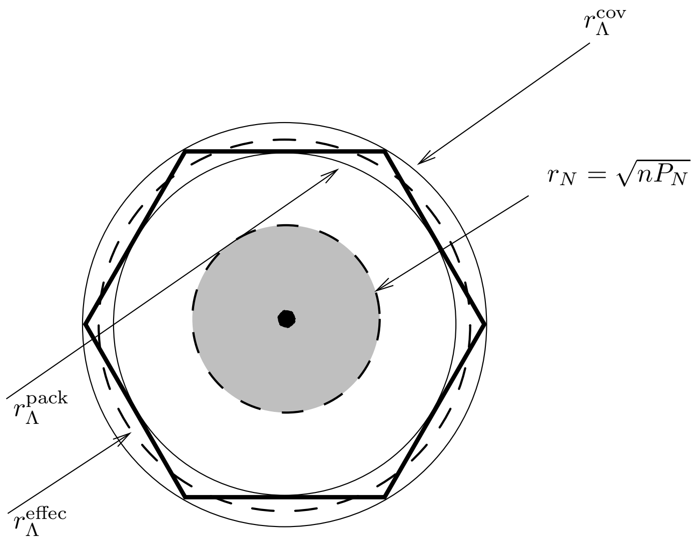

Fig. 1: Geometric picture

ρ_{pack} of a lattice Λ by

$$\rho_{\text{pack}}(\Lambda) = \frac{r_\Lambda^{\text{pack}}}{r_\Lambda^{\text{effec}}}. \tag{6}$$

We note that the packing efficiency $\rho_{\text{pack}}(\Lambda)$ is by definition no greater than one. We wish $\rho_{\text{pack}}(\Lambda)$ to be as large as possible. For a sequence of lattices Λ_n, the best known asymptotic lower bound for $\rho_{\text{pack}}(\Lambda)$ is equal to $\frac{1}{2}$, a result known as the Minkowski-Hlawka theorem [53].

Covering problem: The associated notions for the covering problem are defined similarly to their packing counterparts. The set $\Lambda + r\mathcal{B}$, composed of spheres centered around the lattice points, is a covering of Euclidean space if

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

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

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

This is also the minimum radius of a ball containing $\mathcal{V}_0$. Define the covering efficiency $\rho_{\text{cov}}(\Lambda)$ of a lattice by

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

We note that the covering efficiency $\rho_{\text{cov}}(\Lambda)$ is by definition not less than one. We wish $\rho_{\text{cov}}(\Lambda)$ to be as small as possible. It is a result of Rogers [52] that there exist a sequence of lattices such that $\rho_{\text{cov}}(\Lambda_n) \to 1$ as $n \to \infty$. This means that covering (in contrast to packing) may be asymptotically efficient, i.e., every point in space can be covered (for a good lattice covering) by at most a sub-exponential number of spheres.

See standard textbooks on packing and covering such as Rogers [53] and Conway and Sloane [11].

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 $\mathcal{V}$,

$$\sigma_\Lambda^2 = \frac{1}{\mathrm{Vol}(\mathcal{V})} \cdot \frac{1}{n} \int_{\mathcal{V}} \|\mathbf{x}\|^2 d\mathbf{x}. \tag{7}$$

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

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

where $V = \mathrm{Vol}(\mathcal{V})$. 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 $G_n \leq G_1 = \frac{1}{12}$.

The operational significance of this figure of merit comes from classical results in high resolution quantization theory; see e.g. [28]. A result due to Poltyrev in [64] states that

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

i.e., that there exist a sequence of "good" lattice quantizers Λ_n^* such that $G(\Lambda_n^*) = G_n \to \frac{1}{2\pi e}$. Another result in [64] is that the quantization noise of a good lattice (e.g., a lattice achieving G_n) is "white", i.e., the covariance matrix of a uniform distribution over $\mathcal{V}$ is given by $\sigma_\Lambda^2 \cdot I$, where I is the identity matrix.

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

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

where Z is i.i.d. Gaussian noise of variance N. 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 [50] definition of coding for the *unconstrained* AWGN channel, allowing to separate the "granular" properties of the lattice as a good channel code from the issue of shaping (to meet the power constarint). In this scenario any point of a lattice may be transmitted, corresponding to infinite power and transmission rate. For a given lattice the ML decoder will search for the lattice point that is nearest to the received vector. Therefore, the probability of decoding error is the probability that the noise 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

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

where N is such that (11) is satisfied with equality [22].[1] 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

[1]We omit the $2\pi e$ from the original definition to keep the symmetry with the definition of the NSM.

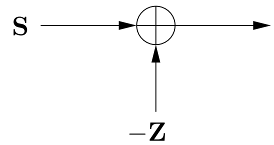

Fig. 2: Equivalent additive noise channel of the dithered lattice quantizer.

of a sphere, denoted $\mu_n^*(P_e)$, approaches $2\pi e$ for any $1 > P_e > 0$ as $n \to \infty$.[2] Furthermore, since a sphere supports the isotropic vector $\mathbf{Z}$ better than any shape of the same volume (see the *sphere bound* of [22]), we have $\mu_n(P_e) > \mu_n^*(P_e) > 2\pi e$, where the second inequality holds for all sufficiently small P_e. It follows from Poltyrev (see also [21], [22], [15]) that

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

III. 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 “blockiness” in picture coding. Dithered quantization is also an effective mean to guarantee a desired distortion level, independent of the source statistics.

Specifically, let the vector $\mathbf{Z}$ be uniform over the fundamental Voronoi region $\mathcal{V}_0$ of the lattice Λ, and independent of the source. We say that $\mathbf{Z}$ is “subtractive dither” if it is known at both the encoder and the decoder, and we reconstruct the source vector $\mathbf{s}$ as $Q_\Lambda(\mathbf{s}+\mathbf{Z})-\mathbf{Z}$. Addition and subtraction of a vector $\mathbf{z}$ before and after lattice quantization amounts to shifting the lattice quantizer $Q_\Lambda(\cdot)$ by the vector $-\mathbf{z}$. Since the lattice quantizer is periodic in space, a random uniform shift over the lattice period makes the quantization error uniform as well.

Theorem 1. [64] *The quantization error $Q_\Lambda(\mathbf{s}+\mathbf{Z})-\mathbf{Z}-\mathbf{s}$ is uniform over $-\mathcal{V}_0$, the reflection of the fundamental Voronoi region $\mathcal{V}_0$, independent of the source vector $\mathbf{s}$.*

Equivalently, $(\mathbf{s}+\mathbf{Z}) \bmod \Lambda$ is uniform over $-\mathcal{V}_0$ for any $\mathbf{s}$, a result termed “Crypto Lemma” by Forney [23]. As a corollary from Theorem 1 and (7), the mean squared distortion of the dithered quantizer is equal to the lattice second moment:

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

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

[2]Note that by the Law of Large Numbers $\frac{1}{n}\|\mathbf{Z}\|^2 \to P_N$ as $n \to \infty$, thus the typical set of the noise $\mathbf{Z}$ coincides with a sphere of radius $\sqrt{nP_N}$, and therefore the NSM-VNR product for a sphere $G_n^*\mu_n^*(P_e) \to 1$.

In high resolution quantization theory it is common to model the quantization process as adding (independent) noise to the source [27]. While this is just an approximation (which holds under the assumption that the quantizer cells are small compared to the variations in the source p.d.f.), Theorem 1 shows that for dithered quantization the *equivalent additive noise* model, shown in Figure 2, is accurate at *any resolution*.

The next theorem makes the connection to an additive noise channel even stronger. Assume that for a 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

$$H(Q_\Lambda(\mathbf{S}+\mathbf{Z})|\mathbf{Z}) = \frac{1}{V}\int_{\mathcal{V}_0} d\mathbf{z} \sum_i -p_i(\mathbf{z}) \log p_i(\mathbf{z}) \tag{15}$$

where $p_i(\mathbf{z}) = \int_{\mathcal{V}_i} f_S(\mathbf{x}-\mathbf{z})d\mathbf{x}$ is the probability that the source vector falls in the ith cell of the shifted lattice. If the source statistics are known a priori, then this entropy is achieved by making the binary word length close to $-\log p_i(\mathbf{z})$ [14]. Otherwise, there exist universal coding schemes (e.g., the Lempel-Ziv compression algorithm and others) which can sequentially approach the entropy of any stationary and ergodic source [14]. We call such a combination of a lattice quantizer and optimum lossless encoding an Entropy Coded Dithered Quantizer (ECDQ).

Theorem 2. [62] *The rate 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 Figure 2:*

$$H(Q_\Lambda(\mathbf{S}+\mathbf{Z})|\mathbf{Z}) = I(\mathbf{S};\mathbf{S}-\mathbf{Z}) \tag{16}$$

where I denoted mutual information [14].

The mutual information formula above resembles the expression for Shannon's rate-distortion function:

$$R(D) = \inf_{\hat{S}:\ E\{d(S,\hat{S})\}\le D} I(S;\hat{S}) \tag{17}$$

where the distortion measure d can be, for example, squared error $d(s,\hat{s}) = (s-\hat{s})^2$. This function characterizes the minimum achievable rate in lossy compression of a memoryless source S, and can be extended to sources with memory [14]. The formal resemblance between the two formulas leads to a *universal* bound on the loss of the entropy-coded dithered lattice quantizer.

Theorem 3. [69], [62] *For any source* $\mathbf{S}$*, the redundancy of the ECDQ above the rate-distortion function under a squared error distortion measure is at most*

$$H(Q_\Lambda(\mathbf{S}+\mathbf{Z})|\mathbf{Z}) - R(D) \le \frac{1}{2} + \frac{1}{2}\log(2\pi e G(\Lambda)) \quad \textit{bit.} \tag{18}$$

See [62] for general difference distortion measures. Any "good" (squared-error) lattice quantizer satisfies Lloyd's centroid condition, [27], implying that the dither vector has zero mean. For such a lattice quantizer, the second term in the right hand side above can be interpreted as the divergence (or "Kullback-Leibler distance") of the dither distribution from AWGN:

$$\frac{1}{2}\log(2\pi e G(\Lambda)) = D(\mathbf{Z}\|\mathbf{Z}^*) \tag{19}$$

where Z^* is a zero-mean i.i.d. Gaussian vector with $\text{Var}(Z_i) = \sigma_\Lambda^2$ for all i, and where $D(\cdot\|\cdot)$ denotes divergence [14], [64].

Interestingly, the channel of Figure 2 could realize the rate-distortion function (17) of any continuous source in the limit of high resolution quantization ($D \to 0$), if the dither $\mathbf{Z}$ was replaced by Gaussian noise [4]. The divergence $D(\mathbf{Z}\|\mathbf{Z}^*)$ measures the information distance of the dither from being Gaussian. Thus, at high resolution conditions, we can obtain a tighter characterization for the rate loss of the ECDQ.

Theorem 4. *For any continuous source* $\mathbf{S}$ *(i.e., a source with a p.d.f.), the redundancy of the ECDQ above the rate-distortion function under squared-error distortion measure satisfies*

$$\lim_{D\to 0} H(Q_\Lambda(\mathbf{S}+\mathbf{Z})|\mathbf{Z}) - R(D) = \frac{1}{2}\log(2\pi e G(\Lambda)). \tag{20}$$

Note that this rate loss is half a bit smaller than the universal bound of Theorem 3, which holds at any quantization resolution.

To close the discussion of universal quantization, we recall from the previous section that for a sequence of lattices ($\Lambda_n^* \in \mathcal{R}^n$) which are good for quantization, $\lim_{n\to\infty} G(\Lambda_n^*) = \frac{1}{2\pi e}$. Thus, for such lattices the divergence of the dither from Gaussianity (19) is going to zero. As a consequence, the bounds above on the redundancy of the ECDQ go to half a bit - at general resolution, and to zero - at high resolution.

Note that the redundancy of the best lossy compression scheme goes to zero with the dimension like $\log(n)/n$, which is the same asymptotic behavior as $\log(2\pi e G_n)$. See [63]. Does the lattice Λ_n^* represent the best arrangement of code points at *any* finite dimension n? This is, in fact, an open question at high-resolution vector quantization theory.

A. Filtered ECDQ: the "test-channel simulator"

Consider the equivalent additive noise channel model in Figure 2. If the second order statistics of the source are 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.

Note first that following the discussion in Section II, the dither vector of an optimum (squared-error) lattice quantizer is "white", i.e., the dither components are uncorrelated and have equal variance. Thus, the noise in the equivalent ECDQ channel of Figure 2 is white, although not quite Gaussian (unless for Λ_n^* as $n \to \infty$). If also the source is white (i.e., memoryless), then the Wiener filter is a simple scalar coefficient β at the output of the equivalent channel, [23]. For such a source the reconstruction becomes $\hat{\mathbf{S}} = \beta[Q_\Lambda(\mathbf{S}+\mathbf{Z}) - \mathbf{Z}]$, 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 $D = \frac{\sigma_S^2\sigma_\Lambda^2}{\sigma_S^2+\sigma_\Lambda^2}$. If we further assume that the source is Gaussian, $S^* \sim N(0, \sigma_S^2)$, then the rate-distortion function (17) becomes

$$R^*(D) = \frac{1}{2}\log\left(\frac{\sigma_S^2}{D}\right) \quad 0 < D \le \sigma_S^2. \tag{21}$$

On the other hand, the mutual information in (16) can be written as a difference of differential entropies $I(\mathbf{S}^*; \mathbf{S}^* - \mathbf{Z}) = h(\mathbf{S}^* - \mathbf{Z}) - h(\mathbf{Z})$, where the first term is bounded by the Gaussian entropy (corresponding to variance of $\sigma_S^2 + \sigma_\Lambda^2$), while the second term is a uniform entropy, which in view of (8) is equal to

$$h(\mathbf{Z}) = \log(V) = \frac{1}{2}\log\left(\frac{\sigma_\Lambda^2}{G(\Lambda)}\right) \tag{22}$$

where V is the lattice cell volume. Combining together, it follows that the redundancy of the ECDQ above (21) is at most

$$H(Q_\Lambda(\mathbf{S}^* + \mathbf{Z})|\mathbf{Z}) - R^*(D) \le \frac{1}{2}\log(2\pi e G(\Lambda)) \ , \tag{23}$$

now not only for small but for *all* distortion levels. See the proof in [63], where it is also shown that for a *non-Gaussian* source, the rate loss of the “Wiener filtered” ECDQ increases by at most $D(S\|S^*)$, the divergence of the source from Gaussianity.

We can write the output coefficient directly in terms of the target distortion as

$$\beta = 1 - \frac{D}{\sigma_S^2}, \tag{24}$$

and the lattice second moment is then chosen as $\sigma_\Lambda^2 = D/\beta$. We observe that β is smaller than one for all the range $0 < D \le \sigma_S^2$. Thus, interestingly, the reproduction lattice $\beta\Lambda$ is a “deflated” version of the encoding lattice Λ. More on the meaning of that - in the next chapter.

To extend this concept to a source with *memory*, we shall first need to assume that the entropy coding is done *jointly*, to take advantage of the dependence between consecutive outputs of the lattice quantizer.[3] This is equivalent to conditioning the probabilities $p_i(\mathbf{z})$ in (15) on past outputs of the lattice quantizer. An extension of Theorem 2 above shows that the resulting *entropy rate*, denoted $\bar{H}$, is equal to the *mutual information <u>rate</u>* of the source over the equivalent additive noise channel:

$$\bar{H}(Q_\Lambda(\mathbf{S} + \mathbf{Z})|\mathbf{Z}) = \bar{I}(\mathbf{S}; \mathbf{S} - \mathbf{Z}). \tag{25}$$

Now, recall that the rate-distortion function of a general stationary Gaussian source, with power spectrum $S(e^{j2\pi f})$, is given by the “water-pouring” parametric equations [4], [14]:

$$\begin{aligned} R(D) &= \int_{-1/2}^{1/2} \frac{1}{2}\log\left(\frac{S(e^{j2\pi f})}{D(e^{j2\pi f})}\right) df \\ &= \int_{f: S(e^{j2\pi f}) > \theta} \frac{1}{2}\log\left(\frac{S(e^{j2\pi f})}{\theta}\right) df \end{aligned} \tag{26}$$

[3]We shall see a linear prediction approach later, in Section VI.

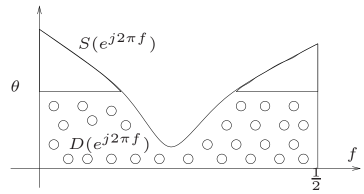

Fig. 3: The water filling solution.

where the *distortion spectrum* is given by (See Figure 3)

$$D(e^{j2\pi f}) = \begin{cases} \theta, & \text{if } S(e^{j2\pi f}) > \theta \\ S(e^{j2\pi f}), & \text{otherwise,} \end{cases} \tag{27}$$

and where we choose the *water level* parameter θ so that the total distortion is D:

$$D = \int_{-1/2}^{1/2} D(e^{j2\pi f}) df. \tag{28}$$

This function is realized by the “forward test-channel realization” [4]

$$Y_n = h_{2,n} * (h_{1,n} * X_n + N_n) \tag{29}$$

where $N_n \sim N(0, \theta)$ is AWGN with $\theta = \theta(D) =$ the water level, $*$ denotes convolution, and $h_{1,n}$ and $h_{2,n}$ are the impulse responses of a pre-filter and a post-filter, respectively, whose absolute squared frequency responses are given by

$$|H_1(e^{j2\pi f})|^2 = |H_2(e^{j2\pi f})|^2 = 1 - \frac{D(e^{j2\pi f})}{S(e^{j2\pi f})} \tag{30}$$

and their phase responses are chosen so that $H_2(e^{j2\pi f}) = H_1^*(e^{j2\pi f})$. Since as we saw above the ECDQ simulates, in an information sense, an additive channel with noise $-\mathbf{Z}$, we can combine the same pre- and post-filters as in (30) with the ECDQ, where entropy-coding is done sequentially over consecutive outputs of the lattice quantizer. Since the system is linear, the resulting MSE would be the same as if the noise $-\mathbf{Z}$ was white Gaussian (as in the forward test channel realization), while the mutual information rate would increase by at most the divergence $D(\mathbf{Z}\|\mathbf{Z}^*)$. Thus, the resulting scheme achieves the R-D function (26) - (28) up-to $D(\mathbf{Z}\|\mathbf{Z}^*) = \frac{1}{2}\log(2\pi e G(\Lambda))$.

Note that if the source spectrum is bounded away from zero, then in the limit of high resolution ($D \ll \min_f S(e^{j2\pi f})$) the pre and post filters degenerate, and can be replaced by shortcuts. Note also that in the memoryless source case, the system reduces to a scalar form which is equivalent to the one discussed in the beginning of this section: the source is first multiplied by $\sqrt{\beta}$, then quantized with a lattice with $\sigma_\Lambda^2 = D$, and then multiplied again by $\sqrt{\beta}$ for reproduction.

As a final conclusion, if we use lattices from the sequence Λ_n^* (lattices which are “good” for quantization), then the rate loss vanishes in the limit as the lattice dimension goes to infinity.

Theorem 5. [63] *For any stationary Gaussian source, the entropy rate of the pre/post filtered ECDQ system satisfies*

$$\bar{H}(Q_{\Lambda_n^*}(\mathbf{S}_{pre}^* + \mathbf{Z})|\mathbf{Z}) \to R^*(D) \quad \text{as } n \to \infty \tag{31}$$

where $\mathbf{S}_{pre}^*$ *denotes the pre-filtered source and* D *is the overall distortion.*

IV. Voronoi Codebooks

As information theory shows us, Gaussian sources and channels should be encoded 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. [4] The resulting codebook in $\mathcal{R}^n$ (n being the code dimension) has a Gaussian - or equivalently, a *spherical* - shape, with roughly evenly spaced points as codewords. 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 un-constrained channel coding [50]. In fixed-length source coding or power-constrained channel coding, however, the codebook must be bounded.

In this Chapter we show how to construct a lattice codebook, that is rate-distortion / capacity achieving for Gaussian sources / channels, and whose codewords and shaping region both have a lattice structure. Our construction is based on the notion of nested lattices, which have its roots in De-Buda's spherical lattice codes [6] and Forney's Voronoi constellations [20], [21], and owe its development to the search for structured binning schemes for side information problems; see the next Chapter.

A. Nested Lattices

We introduce a "double lattice" construction which provides a structured solution for side-information problems [66], [16], [18]. 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. The volumes of the Voronoi cells of Λ_1 and Λ_2 satisfy

$$V_2 = \det\{\mathbf{J}\} \cdot V_1 \tag{32}$$

[4] If the source, or the "power spectral density mask" of the transmitter output, are colored, then the generating distribution must be colored too.

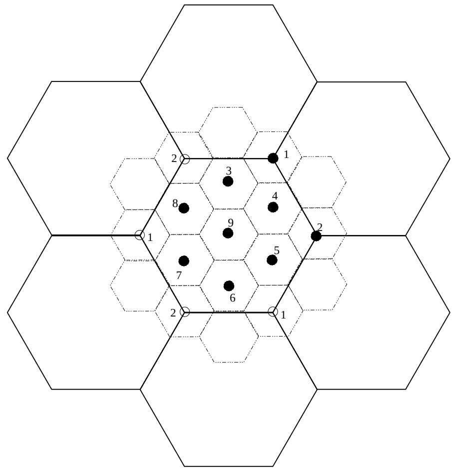

Fig. 4: Nested lattices: special case of self similar lattices.

where $V_2 = \text{Vol}\,(\mathcal{V}_{0,2})$ and $V_1 = \text{Vol}\,(\mathcal{V}_{0,1})$. We call $\sqrt[n]{\det\{\mathbf{J}\}} = \sqrt[n]{V_2/V_1}$ the *nesting ratio*.

Figure 4 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 rotated – version of Λ_1 [10].

The points of the set

$$\Lambda_1 \bmod \Lambda_2 \stackrel{\triangle}{=} \Lambda_1 \cap \mathcal{V}_{0,2} \tag{33}$$

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. Mapping of border points in (33) (i.e., points of Λ_1 who fall on the envelop of the Voronoi region $\mathcal{V}_{0,2}$) to the coset leader set is done in a systematic fashion, so that the cosets $\Lambda_{2,v}\,,\ v \in \{\Lambda_1 \mod \Lambda_2\}$ are disjoint. It follows that there are $V_2/V_1 = \det\{J\}$ different cosets. Enumeration of the cosets can be obtained using a parity-check-like matrix [11]. See also the formulation of Voronoi constellations [19], [20].

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 by Erez *et al* in [15]. The key to proving the existence of such lattices is to consider an appropriate *ensemble* of lattices. Such an ensemble was defined in the seminal work of Loeliger [41] who also demonstrated how random coding arguments can be used to establish “goodness” properties for most members of the lattice ensemble.

B. MMSE Estimation and Lattice Inflation

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

$$(\mathbf{u} + \Lambda_1) \bmod \Lambda_2 \tag{34}$$

where the dither $\mathbf{u}$ is an arbitrary vector in $\mathcal{R}^n$ to be specified later. (For $\mathbf{u} = 0$ this is the set of relative coset leaders in (33).) The size of this codebbok is V_2/V_1 (independent of $\mathbf{u}$), so the associated coding rate is

$$R = 1/n \log_2(V_2/V_1)$$

bit per dimension.

If we use this codebook for Gaussian source coding, then the fine lattice should be a "good quantizer". The coarse lattice, on the other hand, should minimize the cell volume V_2 (for best compression performance) while keeping the *overload probability* [32] (the probability a source vector falls outside the Voronoi region $\mathcal{V}_{0,2}$) low. This means that the coarse lattice should act like a "good AWGN channel code".

If we use this codebook for AWGN channel coding, i.e., as a Voronoi constellation [20], then the roles are reversed. The fine lattice should be a "good AWGN channel code", while the coarse lattice should maximize the cell volume V_2 (for maximum capacity) while keeping the transmit power constraint. Thus the coarse lattice should act as a "good quantizer".

If both component lattices are good, i.e., with NSM $\rightarrow 1/2\pi e$ and VNR $\rightarrow 2\pi e$ as the lattice dimension goes to infinity, then the coding rate is roughly given by

$$R \approx \frac{1}{2} \log_2 \left(\frac{\sigma^2_{\Lambda_2}}{\sigma^2_{\Lambda_1}} \right).$$

This rate, however, corresponds to some rate loss in both problems of interest. Specifically, for an AWGN channel with noise power N and power constraint P, we get $R \approx 1/2 \log(P/N)$, i.e., loss of "1" inside the $\log$ with respect to the AWGN channel capacity

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N} \right). \tag{35}$$

For a Gaussian source with variance σ_S^2 encoded at distortion level D, we get $R \approx 1/2 \log((\sigma_S^2 + D)/D)$, because (as we saw in the previous chapter for dithered quantization) the variance of the quantizer output is equal to the source variance plus the second moment of the (fine) lattice (which is equal to D). Thus we get an extra "1" inside the $\log$, corresponding to some rate redundancy above the QG rate-distortion function (21).

Note that from an information theoretic point of view, these rate losses can be avoided by using joint-typicality arguments in the decoding/quantizing operations. However, this means breaking away with the lattice partition structure of the decision cells[5], hence increasing the decoding/quantizing complexity.

We shall show below that by combining dithering and linear operations, it is possible to achieve capacity and rate-distortion function while still preserving the lattice structure at the encoding and decoding stages. We already used this technique in the filtered ECDQ system of Section III-A. There, linear processing amounted to minimum mean-squared error (MMSE) estimation while dithering was responsible for de-correlating the signal from the

[5]Vectors outside the codebook shaping region must be projected onto the surface of the region rather than quantized to the nearest lattice point.

quantization error. Here we shall give this technique also the interpretation of *lattice inflation*. In particular, we shall use the identities

$$[\alpha Y] \bmod \Lambda = \alpha \Big[Y \bmod \frac{\Lambda}{\alpha} \Big] \tag{36}$$

and

$$Q_\Lambda(\beta S) = \beta \; Q_{\frac{\Lambda}{\beta}}(S), \tag{37}$$

which imply that scaling down / up a vector before a modulo lattice operation or quantization, is equivalent to taking the modulo or the quantization operations with respect to a scaled-up / -down version of the lattice.

C. Achieving the AWGN Channel Capacity

Consider the AWGN channel $Y = X + Z$ of (10). Let the dither $\mathbf{U}$ be uniform over the coarse lattice cell $\mathcal{V}_{0,2}$, and let $\mathbf{v}$ be any codeword in $\Lambda_1 \bmod \Lambda_2$. To transmit the message $\mathbf{v}$, the encoder outputs $X = (\mathbf{v} + \mathbf{U}) \bmod \Lambda_2$. 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). At the decoder we calculate the "decision vector"

$$\tilde{\mathbf{Y}} = [\alpha \mathbf{Y} - \mathbf{U}] \bmod \Lambda_2 \tag{38}$$

(where α is a coefficient to be determined later), and decode to the nearest codeword, i.e., $\hat{\mathbf{V}} = Q_{\Lambda_1}(\tilde{\mathbf{Y}})$. By the identities (36) - (37) above, this is equivalent to

$$\hat{\mathbf{V}} = Q_{\frac{\Lambda_1}{\alpha}} \left(\left[\mathbf{Y} - \frac{\mathbf{U}}{\alpha} \right] \bmod \frac{\Lambda_2}{\alpha} \right), \tag{39}$$

i.e., to decoding with respect to the *inflated codebook* $\frac{\Lambda_1}{\alpha} \bmod \frac{\Lambda_2}{\alpha}$. The equivalent channel from the codeword $\mathbf{v}$ to the decision vector

$$\tilde{\mathbf{Y}} = [\alpha(\mathbf{v} + \mathbf{U} \bmod \Lambda_2 + \mathbf{N}) - \mathbf{U}] \bmod \Lambda_2 \tag{40}$$

is called a *modulo-lattice transformation* [16]. The distributive law of the modulo operation implies:

Lemma 1. (Effective modulo-Λ additive noise channel) [16] *The channel from $\mathbf{v}$ to $\tilde{\mathbf{Y}}$ is equivalent in distribution to the modulo additive noise channel*

$$\tilde{\mathbf{Y}} = \Big(\mathbf{v} + [\alpha \mathbf{N} + (\alpha - 1) U'] \Big) \; mod \, \Lambda_2 \tag{41}$$

where $\mathbf{U}'$ is uniform over $\mathcal{V}_{0,2}$ and independent of $\mathbf{v}$ and N.

The *effective (additive) noise* $N_{eff} = [\alpha \mathbf{N} + (\alpha - 1) U'] \bmod \Lambda_2$ is a weighted combination of two components: AWGN and a dither component called "self noise" because it comes from the coarse lattice. For a modulo additive noise channel a uniform input $\mathbf{V} \sim \text{Unif}(\mathcal{V}_{0,2})$ maximizes the mutual information $I(\mathbf{V}; \tilde{\mathbf{Y}})$, which becomes $\log(V_2) - h(N_{eff})$.

The optimum α from a mutual-information viewpoint is the one that minimizes the entropy of the effective noise [23]. As the lattice dimension increases, the dither U' and therefore the effective noise N_{eff} become closer to a Gaussian distribution (in the divergence sense (19)), in which case minimizing entropy amounts to minimizing variance. Having this asymptote in mind, the information-wise optimum α becomes the Wiener coefficient $\alpha = \frac{\sigma^2_{\Lambda_2}}{\sigma^2_{\Lambda_2}+\sigma^2_N} = \frac{P}{P+\sigma^2_N}$, and the resulting noise variance is the MMSE solution

$$\text{Var}(N_{eff}) = \frac{P\sigma_N^2}{P+\sigma_N^2}. \tag{42}$$

We want to approach the corresponding mutual information using the nested lattice code described above. Note that by the modulo-additivity of the equivalent channel, the error probability is identical for all codewords and is equal to

$$P_e = \Pr\{\mathbf{N}_{eff} \notin \mathcal{V}_{0,1}\}. \tag{43}$$

Thus, by the definition of the the VNR (12), if we target some P_e the volume of the fine lattice cell must be $V_1 = [\mu(\Lambda_1, P_e) \cdot \text{Var}(N_{eff})]^{n/2}$ or larger (where we assumed Gaussian N_{eff}, which is true for high SNR (implying $\alpha = 1$), or high dimension and “good” lattice to make the self-noise component “Gaussian enough”). On the other hand, the power constraint implies that the volume of the coarse cell is $V_2 = [P/G(\Lambda_2)]^{n/2}$ or smaller. For the MMSE solution (42), we thus get a coding rate of

$$R = \frac{1}{n}\log\left(\frac{V_2}{V_1}\right) = \frac{1}{2}\log\left(\frac{P/G(\Lambda_2)}{\mu(\Lambda_1,P_e)\text{Var}(N_{eff})}\right) \tag{44}$$

$$= C - \frac{1}{2}\log\Big(G(\Lambda_2)\cdot\mu(\Lambda_1,P_e)\Big) \tag{45}$$

where C is the AWGN channel capacity (35).

If we now assume a sequence of good nested lattice pairs where $G(\Lambda_2) \to 1/2\pi e$ and $\mu(\Lambda_1, P_e) \to 2\pi e$, then the system approaches the AWGN channel capacity. An analysis of the error exponent of Voronoi codebooks can be found in [40].

D. Achieving the QG Rate-Distortion Function

To encode the source $\mathbf{S}$ using a Voronoi codebook $\Lambda_1 \bmod \Lambda_2$, we first quantize a scaled dithered version of the source using the fine lattice, and then send an index identifying the codeword modulo the coarse lattice. At the decoder, we subtract the dither to obtain the reconstructed vector

$$\hat{\mathbf{S}} = Q_{\Lambda_1}(\beta\mathbf{S} + \mathbf{U}) \bmod \Lambda_2 - \mathbf{U} \tag{46}$$

where the dither $\mathbf{U}$ is uniform over the fine cell $\mathcal{V}_{0,1}$. By the identities (36)-(37) above, this procedure is equivalent to

$$\hat{\mathbf{S}} = \beta\Big[Q_{\frac{\Lambda_1}{\beta}}(\mathbf{S} + \tilde{\mathbf{U}}) \bmod \frac{\Lambda_2}{\beta} - \tilde{\mathbf{U}}\Big] \tag{47}$$

where $\tilde{\mathbf{U}} = \mathbf{U}/\beta$, i.e., to encoding with respect to the *inflated codebook* $\frac{\Lambda_1}{\beta} \bmod \frac{\Lambda_2}{\beta}$ and then re-scaling.

Note that if we ignore the shaping by Λ_2, then the second form (47) is the same as the "filtered ECDQ" scheme of Section III-A. We can thus take the coefficient $\beta = 1 - D/\sigma_S^2$ as in (24), and set the fine lattice second moment to $\sigma_{\Lambda_1}^2 = \beta D$ (or equivalently the second moment of the inflated fine lattice to be D/β). This guarantees that the resulting reconstruction mean-squared error $1/nE\|\hat{\mathbf{S}} - \mathbf{S}\|^2$ would be D. The choice of the coarse lattice determines the coding rate $R = 1/n \log(V_2/V_1)$ and the overload probability

$$P_e = \Pr\Big\{Q_{\Lambda_1}(\beta\mathbf{S} + \mathbf{U}) \notin \mathcal{V}_{0,2}\Big\} \leq \Pr\Big\{\beta\mathbf{S} - \mathbf{U}' \notin \mathcal{V}_{0,2}\Big\} \tag{48}$$

where the upper bound follows by taking a $\bmod \Lambda_2$ operation also at the decoder (which can only make the overload probability worse), and then applying Theorem 1. To guarantee some target P_e, we adjust the coarse lattice to the variance of the equivalent source $\beta\mathbf{S} - \mathbf{U}'$ which for β and $\sigma_{\Lambda_1}^2$ above is $\sigma_S^2 - D$. Specifically we choose $V_2 = [\mu(\Lambda_2, P_e) \cdot (\sigma_S^2 - D)]^{n/2}$. And since $V_1 = [\beta D/G(\Lambda_1)]^{n/2}$, we get

$$R = \frac{1}{2} \log\left(\frac{\mu(\Lambda_2, P_e) \cdot \sigma_S^2}{D/G(\Lambda_1)}\right) \tag{49}$$

thus we achieve the rate distortion function (21) up to a redundancy term of

$$Redundancy = \frac{1}{2} \log\Big(\mu(\Lambda_2, P_e) \cdot G(\Lambda_1)\Big) \text{ bit per sample.} \tag{50}$$

As discussed above, we can find a sequence of pairs of nested lattices such that $\mu(\Lambda_{2,n}, P_e) \to 2\pi e$ and $G(\Lambda_{1,n}) \to 1/2\pi e$, as $n \to \infty$. Thus using large dimensional lattices we can make the redundancy term as small as desired. Thus again the NSM-VNR cross product of the lattice pair (with the roles of Λ_1 and Λ_2 switched relative to (45)) determines the rate loss of the system.

V. 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 [59] - a source S is encoded knowing that a correlated signal J is available non-causally 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 [26]. Here a state-dependent channel is encoded knowing the channel states non-casually, however decoding is done solely from the channel output without having access to the channel states. In the special case known as the "dirty paper" channel, the input-output relation is $Y = X + I + Z$, where I, the interference, is known at the encoder but not at the decoder, and Z (the unknown noise) is AWGN [13].

An interesting feature of the Gaussian side-information problems is that their information theoretic solution amounts to complete elimination of the effect of the known parts J and I. Below we show that the lattice encoding and decoding schemes of Section IV give us almost "for free" a simple structured solution for these two Gaussian side-information problems.

A. Lattice Wyner-Ziv Coding

The Wyner-Ziv function of the source $S = J + Q$, with J known at the decoder, at MSE distortion level D, is given by $0.5 \log(\sigma_Q^2/D)$. See [14]. To approach this optimum performance, we use a dithered Voronoi codebook based on a nested lattice pair $(\Lambda_2 \in \Lambda_1)$, as in Section IV. The encoder is identical to that in Section IV, with the lattices tuned according to the variance of the unknown source part Q and the distortion level D. The decoder subtracts the known part J (properly scaled) prior to the modulo lattice operation, and adds it back after the post scaling by β. The final system is shown in Figure 5.

Like in Section IV, the distributive law of the modulo operation implies that the mapping between S, J and $\hat{S}$ is equivalent to the channel in Figure 6. Note that the effect of the J component is removed prior to the equivalent modulo operation. Thus, for "correct decoding", we are left with the same condition as in the no side-information case (48)-(49), only now with respect just to the unknown part of the source Q. It follows that after adding the side information, the overall distortion in S is D, as desired. Furthermore, the coding rate R is equal to the WZ function $0.5 \log(\sigma_Q^2/D)$, up to a loss factor $\frac{1}{2} \log\left(G(\Lambda_1) \cdot \mu(\Lambda_2, P_e)\right)$, as in (50). Thus, for a sequence of good nested lattice pairs ($G \to 1/2\pi e$, $\mu \to 2\pi e$) the system becomes optimal.

It is interesting to compare the lattice-WZ system to the standard "random binning" technique [14]. Note that all source vectors $\mathbf{s}$ which are associated with the same fine lattice point modulo the coarse lattice (i.e., with $Q_{\Lambda_1}(\mathbf{s}+\mathbf{u}) \bmod \Lambda_2 = \ell$ for some $\ell \in \Lambda_1 \bmod \Lambda_2$) are mapped to the same channel index. Hence, in the lattice-WZ system, a "bin" is equivalent to a coset of Λ_1 relative to Λ_2, as defined in Section IV-A. Unlike random bins, all such cosets are equivalent, i.e., identical in size and in average distortion. Furthermore, we don't need to make any statistical assumptions regarding the side information signal.

B. Lattice Dirty-Paper Coding

Very similar ideas apply to the dirty-paper channel

$$Y = X + I + Z,$$

where the interference I is known at the encoder, the unknown noise Z is Gaussian with variance N, and the encoder satisfies the power constraint P. We use a pair of nested lattices, with the roles of quantization (here in the sense of shaping) and channel coding reversed: the coarse lattice satisfies $\sigma_{\Lambda_2}^2 = P$, while the fine lattice satisfies $\Pr\{\mathbf{Z} \notin \mathcal{V}_{0,1}\} < P_e$ for sufficiently small decoding error probability P_e. The full description of the lattice

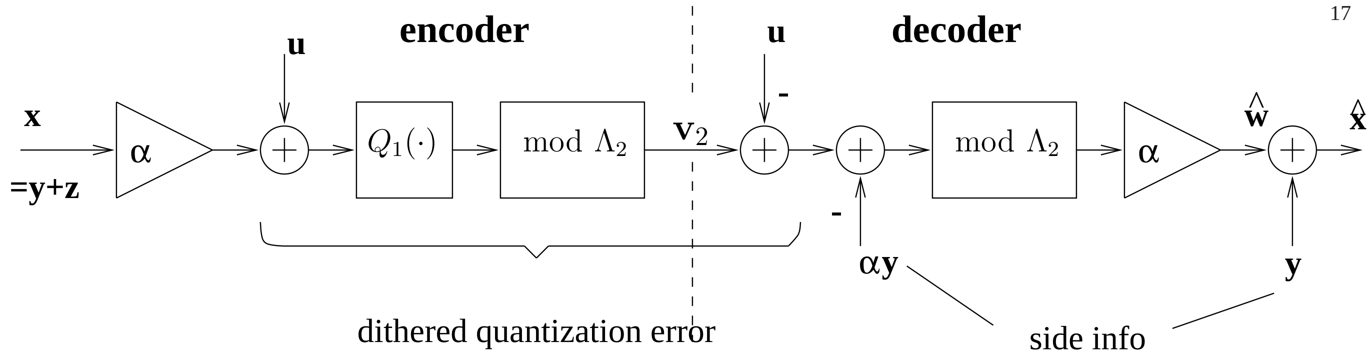

Fig. 5: Wyner-Ziv encoding of a jointly Gaussian source using nested lattice codes. At high resolution $\alpha = 1$.

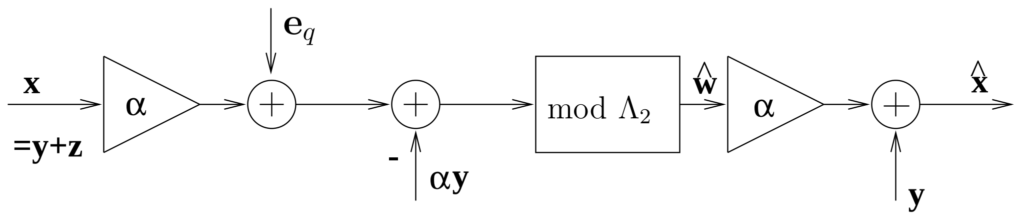

Fig. 6: Equivalent channel for the scheme of Figure 5.

DPC scheme can be found in [66] (also [16]). For the high signal-to-noise (SNR) case the analysis simplifies, and the resulting coding rate (which like above it is the logarithm of the nesting ratio) becomes

$$\frac{1}{2}\log\left(\frac{P/G(\Lambda_2)}{\mu(\Lambda_1,P_e)\cdot\sigma_z^2}\right)\approx C-\frac{1}{2}\log\Big(G(\Lambda_2)\cdot\mu(\Lambda_1,P_e)\Big) \tag{51}$$

where $C=\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ denotes the AWGN channel capacity. We see again that in order to make the capacity loss term small, we need nested lattices with a small cross NSM-VNR product, but in reversed order w.r.t. the Wyner-Ziv problem above.

More delicate analysis and lattice properties are required at the *non-* high SNR regime; see e.g. [40] for the error exponent in lattice decoding.

VI. Waveform Sources and Channels

We shall now apply the lattice coding techniques developed so far to the efficient encoding of signals and channels with memory. One of the interesting observations we shall make is that memory can be treated as "side-information". This gives rise to "reversed" forms of common prediction and equalization techniques in source and channel coding.

A. Predictive Quantization and Wyner-Ziv DPCM

Linear prediction is an effective mean to exploit memory in source coding. In differential pulse code modulation (DPCM), [32], the current source sample is predicted from the past reconstruction - a procedure called *backward*

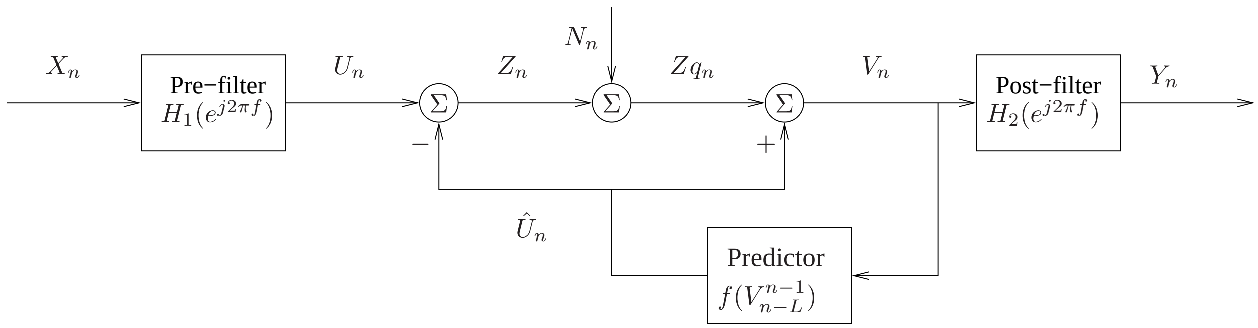

Fig. 7: Predictive Test Channel

prediction - and the prediction error is encoded by a simple scalar quantizer. The error reduction due to prediction translates into rate saving for the same distortion in quantization. Can we replace the complex *sequence* entropy coding in the ECDQ of Section III-A by linear prediction followed by *per-sample* quantization?

The "predictive test channel" of Figure 7 provides a first step towards this goal. This channel has the same pre- and post-filters (30) as in the forward channel realization of Section III-A, while the inner AWGN channel was replaced by a sequential block:

$$J_n = f(V_{n-1}, V_{n-2}, \ldots, V_{n-L}) \tag{52}$$

$$Q_n = U_n - J_n \tag{53}$$

$$Qq_n = Q_n + N_n \tag{54}$$

$$V_n = J_n + Qq_n. \tag{55}$$

where $N_n \sim \mathcal{N}(0, \theta)$ is i.i.d. and $\theta = \theta(D)$ is the "water level" parameter (26).

Theorem 6. *[67] The system of Figure 7, satisfies*

$$E(Y_n - X_n)^2 = D. \tag{56}$$

Furthermore, if the source X_n is Gaussian and $f = f(V_n^-)$ is the optimum infinite order predictor then

$$I(Q_n; Q_n + N_n) = R(D). \tag{57}$$

Thus, the scalar *mutual information over the channel (54) is equal to the RDF (26) - (28).*

The next step is to replace the AWGN channel inside the prediction loop by a Voronoi codebook (as in Section IV-D) or a lattice ECDQ (as in Section III). These coding schemes approximate the information-distortion performance of the channel $Qq_n = Q_n + N_n$, up-to a rate loss of $\frac{1}{2}\log\left(G(\Lambda_1)\cdot\mu(\Lambda_2, P_e)\right)$ (for a Voronoi codebook), or $\frac{1}{2}\log\left(2\pi e G(\Lambda)\right)$ (for a lattice ECDQ). The combination of the encoder section (from S_n to the codeword associated with Qq_n) and the decoder section (mapping back to Qq_n and then to $\hat{S}_n$) results in a lattice-DPCM system.

At first sight, however, the dimensionality of the lattice codebook seems to be in conflict with the temporal sequentiality of the system. Nevertheless, it is shown in [67] (see [29]) how to use the n dimensional lattice over

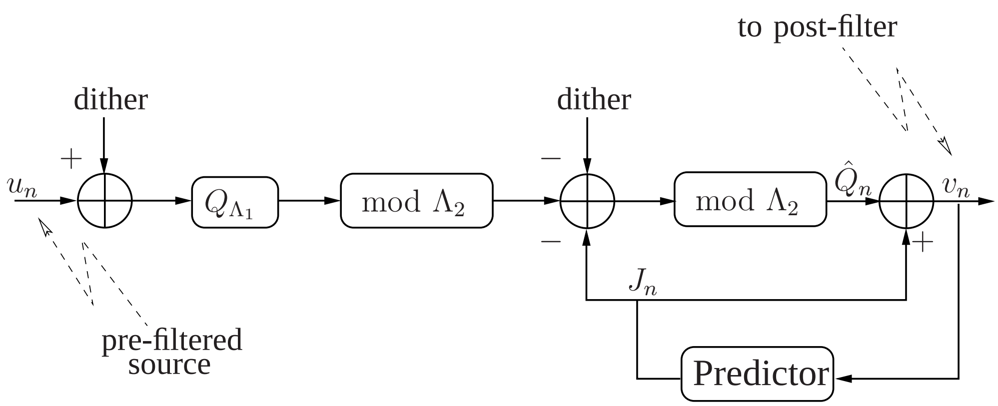

Fig. 8: Wyner-Ziv DPCM.

a virtual "spatial" dimension, by encoding in parallel n different sources or n interleaved long segments of the same source.

Finally, we can view the predicted process J_n as side information which is available at the decoder [57]. We can thus use the lattice-WZ system to encode the filtered source U_n, avoiding the prediction at the encoder all together. This "Wyner-Ziv DPCM" system is shown in Figure 8.

B. Decision Feedback Equalization (DFE) and Lattice Pre-coding

The same ideas can be used to extend the Voronoi lattice coding scheme of Section IV-C to channels with inter-symbol interference (ISI). The channel model is like the interference channel of Section V-B, only here the interference signal is a linear combination of previous channel inputs

$$I_n = \sum_{i=1}^{\infty} h_i X_i.$$

As shown by Cioffi *et al*, [9], if we assume that the receiver decides correctly on previous channel inputs (correct decoding), then the equivalent ISI at the receiver input can be canceled perfectly, implying that the *scalar* mutual information over the decision device ("slicer") coincides with the mutual-information-*rate* over the entire channel. If the input spectrum is capacity achieving, then so is the scalar mutual information at the slicer. Like for colored sources above, a Voronoi codebook can be used, with a lattice decoder (quantizer) as a slicing device, to approach this scalar (Gaussian) mutual information, up-to a loss term of $\frac{1}{2}\log\left(G(\Lambda_2)\cdot\mu(\Lambda_1,P_e)\right)$.[6] Again, to allow sequential operation, the lattice is encoded over a virtual "spatial" dimension of n parallel channels, generated by interleaving long channel segments. See [29], [67].

Finally, we can transfer the channel interference compensation from the decoder to the encoder, by viewing the ISI term I_n above as side information known to the encoder. We can then use the lattice DPC system of Section V-B to cancel the interference at the encoder. The resulting system, illustrated in Figure 9, provides a lattice form of the well known Tomlinson-Harashima pre-coder [9], [66], [47].

[6]Since the equivalent slicer channel is a *reversed* AWGN channel [67], the theoretically achievable rate is half the logarithm of the SNR at the slicer input (i.e., with no additional "1" inside the logarithm [9]).

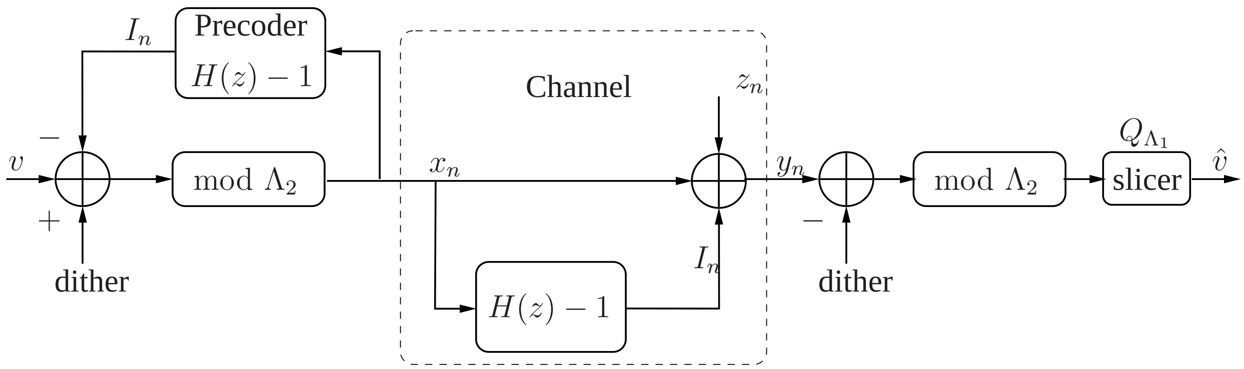

Fig. 9: Lattice Precoding (high SNR version).

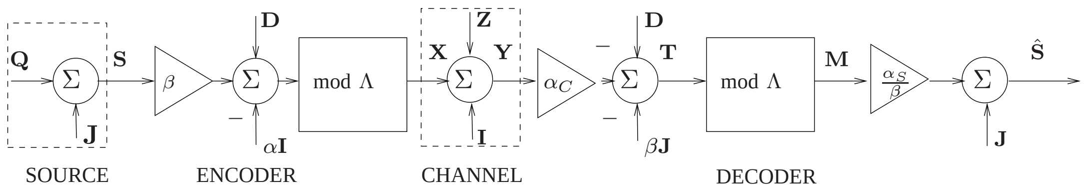

Fig. 10: Analog Wyner-Ziv / dirty-paper coding scheme

VII. Modulo-Lattice Modulation

So far we discussed either source or channel coding problems. In this section we address the combination of the two. We shall present a "semi-analog" joint source-channel modulation technique, which is based on lattice codes. This technique - called "modulo-lattice modulation" (MLM) - enjoys the benefits of both analog and digital communication. It is robust to channel uncertainty and saves complexity - similar to analog modulation, yet it optimally matches sources and channels of general bandwidth and spectra - like in digital communication. See [34], [33].

The presentation starts with an analog version of the lattice side-information schemes of Section V, which forms the basis for the joint source-channel lattice modulation scheme.

A. Joint WZ/DPC Lattice Coding

Suppose that the composite source $S = J + Q$ of Section V-A needs to be transmitted over the power-constrained interference channel $Y = X + I + Z$ of Section V-B. Like in these two Sections, suppose that the decoder knows the J component of the source, while the encoder knows the I component ("interference") of the channel noise. The I and J signals are known non-causally.

We can merge the two systems of Sections V-A and V-B to construct a joint-source channel modulation scheme, as shown in Figure 10. This MLM scheme consists of only one lattice - the former coarse lattice of the WZ/DPC schemes - which is used to take account of the side-information signals I and J. The fine quantization lattice of the WZ system, and the fine lattice codebook of the DPC system, are replaced by direct mapping of the (scaled) source vector to the channel input. Thus, in a sense, we saved the complexity of 3 out the 4 lattices in a full digital implementation, where a lattice WZ system is concatenated with a lattice DPC system.

How good is this combined system? Since the separation principle holds in this setting [43], it follows by combining the QG WZ-RD function $1/2\log(\sigma_Q^2/D)$ and the DPC channel capacity $1/2\log(1+SNR)$ (where $SNR = P/N$), that the minimum theoretically attainable distortion (OPTA) is

$$D^{opt} = \frac{\sigma_Q^2}{1+SNR} \tag{58}$$

Note that this D^{opt} is independent of the power of the interference I as well as of the variance of the known part of the source J. Hence (setting $I = J = 0$), it is the same as if the unknown source part Q was transmitted over a zero-interference AWGN channel $Y = X + Z$. To compare that with the performance of the MLM system of Figure 10, we first use the modulo distributive law to arrive at an equivalent channel with a single modulo operation. See [34]. It can be verified that the signal at the input of the equivalent modulo lattice operation is $Y'' = \beta Q + Z_{eq}$, where $Z_{eq} = \alpha Z + (\alpha - 1)U$, and U denotes the dither. The variance of this (nearly Gaussian) signal determines the size of the coarse lattice cell. Thus, as shown in [34], if we follow the analysis of Sections V-A and V-B, we obtain the distortion

$$D = D^{opt} \cdot \mu(\Lambda, P_e) \cdot G(\Lambda) \tag{59}$$

where P_e is the overload probability as in (43). That is, we suffer a distortion amplification equal to the NSM-VNR product of the lattice - an interesting new figure of merit - provided an overload event did not occur. Thus, if we use lattices with a small NSM-VNR product (i.e., $G(\Lambda) \to 1/2\pi e$ and $\mu(\Lambda, P_e) \to 2\pi e$), then asymptotically for large dimension the non-overload distortion (59) arbitrarily approaches the OPTA of (58). Furthermore, for such lattices the contribution of the overload distortion can be made negligible [34].

The system above is not only asymptotically optimal (for large lattice dimension), but unlike in a digital solution, it is robust to the accuracy of knowing the signal-to-noise ratio SNR=P/N at the encoder, provided the SNR is high. Specifically, if we know in advance that the true SNR is greater than some minimum level γ_0, and tune the encoder parameter β to that value, then the decoder can reconstruct at distortion level which is bounded above by

$$D \leq \frac{\sigma_Q^2}{SNR}\frac{\gamma_0}{\gamma_0 - 1}. \tag{60}$$

This is only slightly worse than the optimum (58) provided γ_0 is sufficiently large. This is in contrast to the digital solution in this situation, where the distortion $D_{digital} = \sigma_Q^2/(1+\gamma_0)$ is fixed for all $SNR \geq \gamma_0$, i.e., determined solely by the minimum SNR and does not improve if the SNR gets better.

Bandwidth Expansion: The MLM system can be used for "bandwidth expansion" - that is, there is no source or channel side-information, but there are ρ channel uses per each source sample, ρ being an integer. Suppose, first, MLM with a scalar lattice ($n = 1$). The first channel input X_1 is the source sample S, scaled to match the input power constraint Then, the received signal $Y_1 = \beta S + Z$ is considered as side-information "J" known to the decoder, and S is transmitted again, now using the MLM principle. (Note that we can always write S as

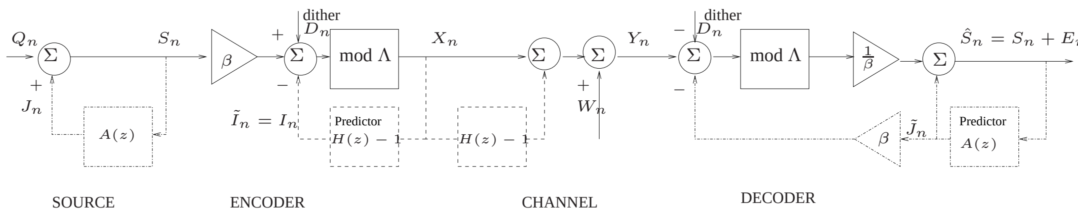

Fig. 11: Workings of the High-SNR Scheme. For illustration purposes, we put together the source, encoder, channel and decoder. Dashed lines show the channel ISI, canceled by the channel predictor. Dotted lines show the source memory component, subtracted and then added again using the source predictor.

$\alpha Y_1 + Q_2$, where αY_1 is the MMSE estimate of S, and Q_2 is the estimation error which is is independent of Y_1.) This “zooming” procedure is repeated $\rho - 1$ times, where every time the previous received samples are considered as side-information [7], [51]. In the general n-dimensional lattice case, the first n-vector input is a scaled version of the source vector, and in each of the $\rho - 1$ iterations the source vector is WZ-encoded treating the previous received vectors as side information. As the lattice dimension goes to infinity, the distortion approaches

$$D = \frac{\sigma_S^2}{(1 + P/N)^\rho} \tag{61}$$

which is the OPTA in this case [51].

In the next section we present a general framework for matching sources and channels of arbitrary bandwidth and spectra using MLM.

B. Analog Matching of Colored Sources and Channels

Suppose a stationary Gaussian source with an arbitrary spectrum needs to be transmitted over a Gaussian channel with an arbitrary noise spectrum, or equivalently, over a filter plus AWGN channel. It is known that, unless the source and the channel spectra are flat over the *same* band, analog transmission - i.e., a simple scalar gain (power matching) at the encoder - cannot achieve Shannon’s OPTA $R(D) = C$. Nevertheless, like we did for the joint source-channel side-information problem, we can merge the lattice precoding and the lattice “reversed” DPCM schemes of Section VI to obtain a “semi-analog” system which is asymptotically optimal for large lattice dimension. Furthermore, unlike the separation-based digital solution, this *Analog Matching* system is robust to channel SNR uncertainty for high SNR.

Consider the system in Figure 11, which consists of inter-symbol interference cancelation at the encoder (a lattice precoder) and “reversed” source prediction at the decoder (lattice WZ of a colored source). Again, we keep the coarse lattice components of the systems of Section VI, and replace the fine quantization / constellation lattices by direct mapping of the (scaled) source vector to the channel input. To simplify the exposition, we shall restrict attention to the high-SNR / small distortion case. In this case the pre-coder subtracts the exact value of the ISI modulo the coarse lattice, and transmits the result as is over the channel. The decoder predicts the next source sample from the past reconstruction as if it was a clean replica of the source (i.e., neglecting the effect of the accumulated channel noise), and use it as side-information for WZ decoding. Like in Section VI, we assume parallel processing or interleaving to allow for a multi-dimensional lattice inside the ISI/source prediction loops.

Under the high resolution assumption (almost "clean" source prediction at the decoder), it follows that the resulting distortion (under the no-overload event) is as if the source innovation Q was transmitted over the zero ISI channel $Y = X + Z$. That is, the distortion is given by $D = \sigma_Q^2/(1 + P/N)$, as in the MLM system for the joint-WZ/DPC setup in Section VII-A. Not surprisingly, this is also Shannon's OPTA in this case. See Kochman *et al* [33] for the complete derivation, as well as for the general resulution/SNR case, and for the robustness property. See [39] for the use of *companding* to overcome the (severe) effect of overload for small lattice dimension and high resolution conditions.

Why don't we take the opposite route, i.e., merge source prediction (DPCM) at the encoder and channel equalization (DFE) at the decoder? We cannot use forward-DPCM, because the encoder must use the "noisy" source samples for prediction [32], [67], but here the noise comes from nature (the physical channel). Likewise, we cannot equalize the channel at the decoder like in a DFE-based system, because the transmitted signal is not digital so the decoder cannot have an exact replica of past transmissions.

Bandwidth conversion: A case of special interest is that when the source and channel bandwidths do not match. In this case we cannot carry on the "high resolution / high SNR" assumption above, because either the source or the channel is sampled above its Nyquist rate.[7] For example, if the source bandwidth is narrower than the channel bandwidth, then at the decoder we must take into account the out-of-band (channel) noise seen by the (strictly causal) predictor. Either way, the optimality of the Analog Matching system implies that it performs *bandwidth conversion* - matching the source bandwidth BW_s to the channel bandwidth BW_c - while preserving mutual information. This fact can be written as

$$(1 + SDR)^{BW_s} = (1 + SNR)^{BW_c} \tag{62}$$

where $SDR = \sigma_s^2/D$ is the unbiased signal to distortion ratio of the source, while $SNR = P/N$ is the channel's signal to noise ratio. Note that here the bandwidth expansion ratio $\rho = BW_c/BW_s$ is not necessarily an integer.

VIII. 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, [14], [5], by a lattice-based solution. This solution uses the the lattice-WZ and lattice-DPC

[7]Without loss of generality, we assume here that the narrower band is contained inside the wider one.

schemes above as building blocks [66]. As in Chapter V, 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 binning.

A. The Korner-Marton Problem

We start with an interesting binary sources setup, the "two help one" problem of Figure 12, which motivates our discussion. In a seminal paper from the late 70's, Korner and Marton [37] 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. Specifically, the Korner-Marton setup consists of three binary sources X, Y, Z, where $Z = X \oplus Y$, and the joint distribution of X and Y is symmetric with $P(X \neq Y) = \theta$. The goal is to encode the sources X and Y separately such that Z can be reconstructed losslessly. Korner and Marton showed that the rate sum required is at least

$$R_x + R_y \geq 2H(Z), \tag{63}$$

and furthermore, this rate sum 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.

In contrast, the "one help one" problem [2], [58] has a closed single-letter expression for the rate region, which corresponds to a random binning coding scheme. Korner and Marton [37] generalize the expression of [2], [58] to the "two help one" problem, and show that the minimal rate sum required using this expression is given by

$$R_x + R_y \geq H(X, Y). \tag{64}$$

The region (64) corresponds to Slepian-Wolf encoding of X and Y, and it can also be derived from the Burger-Tung achievable region [5] for distributed coding of X and Y with one reconstruction $\hat{Z}$ under the distortion measure $d(X, Y, \hat{Z}) \triangleq X \oplus Y \oplus \hat{Z}$. Clearly, the region (64) is strictly contained in the Korner-Marton region $R_x + R_y \geq 2H(Z)$ (63) (since $H(X, Y) = 1 + H(Z) > 2H(Z)$ for $Z \sim$ Bernoulli(θ), where $\theta \neq \frac{1}{2}$).

Krithivasan and Pradhan [38] extended the Korner-Marton problem 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, the lattice scheme does not match the "gini aided" outer bound; it looses 3dB in distortion (half a bit in the rate sum) due to the accumulation of the two (independent) quantization

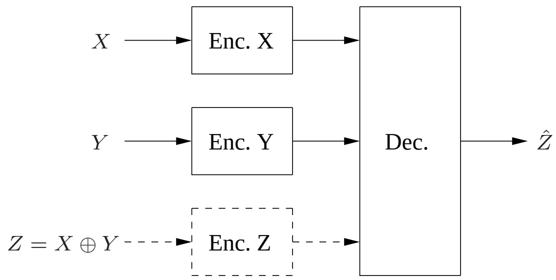

Fig. 12: The Korner-Marton configuration.

noises[8]. Yet this is still much better than a "standard" random binning solution, 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 Korner-Marton problem; a generalization of the Gaussian dirty-paper problem to a multiple access setup [48]. There are two additive interference signals, one known to each transmitter but none to the receiver. See Figure 13. The rates achievable using Costas strategies (i.e. by a random binning scheme induced by Costas auxiliary random variables) vanish in the limit when the interference signals are strong. In contrast, it is shown by Philosof *et al* [48] that lattice strategies (lattice precoding) can achieve positive rates independent of the interferences. Furthermore, they derive an outer bound for the capacity region for arbitrary strong interferences, which is strictly smaller than the clean MAC capacity region. Lattice strategies meet this outer bound for some combinations of noise variance and power constraints. In particular, they are optimal in the limit of high SNR. Thus, the dirty MAC is another instance of a network setup, like the Korner-Marton modulo-two sum problem, where linear coding is potentially better than random binning. See also [49].

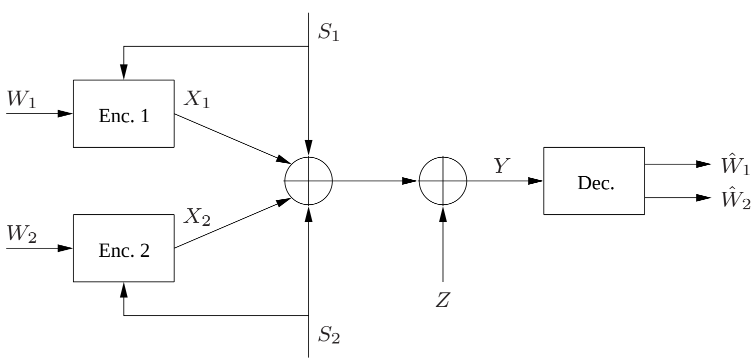

Fig. 13: Doubly dirty MAC.

[8]This is assuming independent dithering at the two terminals. A common dithering scheme is complicated to analize, but may reduce this loss [36].

C. *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 the 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 and which one to discard. If the final destination gets enough "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 was on *linear* coding schemes. In theory, though, any information preserving mapping at the nodes would work, as long as the network is lossless. 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.

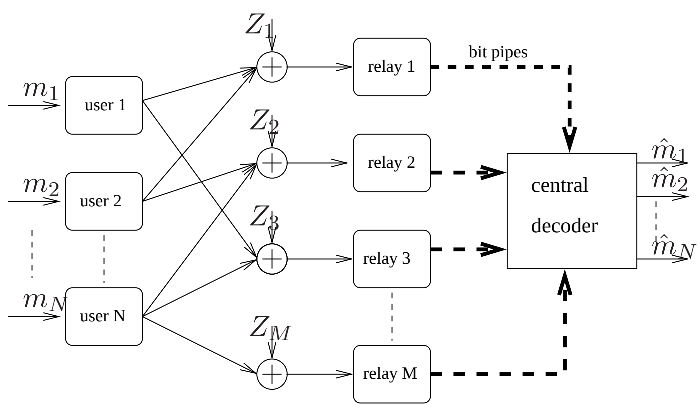

Fig. 14: A multi-relay multi-user network scenario.

Specifically, consider the Gaussian relay network proposed in [45], depicted in Figure 14, 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, i.e., each relay is a Gaussian MAC channel. Thus, the different signals at the relay input are already "combined" by the network. However, unlike in a clean network, if the relay treats its input as an "analog signal" and digitize it, then the noise will be forwarded to the final receiver as well.

It has been shown recently how to use nested lattice codes for network coding in the presence of Gaussian noise [56], [8], [45], [44]. The nested lattice structure allows the relay to decode an integer linear combination of the codewords (a lattice point which is close to the received signal), thus remove the channel noise before forwarding the decoded point to the final receiver. In [17] a general framework is presented, which allows to treat non-integer combinations, as well as non-Gaussian noise and non-additive channels.

D. *Back to Analog: Coherency Gain in Parallel Relaying*

If the relays could *coordinate*, then we could get effectively a multi-antenna relay system. If the number of relays is larger than the number of users, then such a multi-antenna relay could enhance the SNR by a factor of

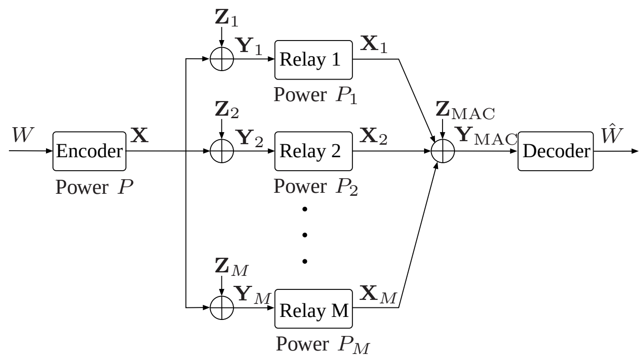

Fig. 15: The Parallel Relay Network.

M/N on the average. This enhancement - due to the coherent combining of transmitted signals at the relays - is known as “array gain”. It becomes significant (even more than noise accumulation avoidance) if $M \gg N$ or if the SNR at the relays is low. Can we achieve this enhancement in a distributed setup?

Consider now the “fully physical” single-user parallel-relay network shown in Figure 15, where both the section from the user to the relays and the section from the relays to the destination are Gaussian channels. An interesting alternative for digital relaying in this setup is *amplify and forward* (A&F) [54], [25]: each relay sends a scaled version of its received signal. Due to the linearity of the network, the transmitted signals are coherently combined at the final receiver. Thus, A&F allows to achieve array gain, in spite of the the lack of relay coordination, at the price of noise accumulation.

This method is, however, limited to the case of simple AWGN channels of identical bandwidth. In a recent work, [35], the concept of A&F was extended to the case where the bandwidth (BW_1) at the user-relays section is different than the bandwidth (BW_2) at the relays-destination section. The new technique - called *rematch and forward* - is based on using an Analog Matching scheme (as in Section VII-B) at the transmitter to match a codeword of bandwidth BW_2 to a channel of bandwidth BW_1. At the relay, the Analog Matching decoder reconstructs the codeword of bandwidth BW_2, while satisfying the mutual information preservation law (62). The reconstructed codeword is then sent in an analog manner to the destination. This procedure exploits the full bandwidth of both sections, while preserving the array gain as in A&F.

IX. Acknowledgement

I would like to thank Uri Erez, Yuval Kochman and Tal Philosof for constant help throughout this work, and to Anatoly Khina for the help in the figures.

References

[1] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger. Closest point search in lattices. *IEEE Trans. Info. Theory*, IT-48:2201–2214, Aug. 2002.

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

[3] R. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Info. Theory*, IT-49:1159–1180, 2003.

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

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

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

[7] B. Chen, and G. Wornell. Analog error-correcting codes based on chaotic dynamical systems. *IEEE Trans. Communications*, IT-46, pp. 881-890, July 1998.

[8] 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.

[9] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. J. Forney. MMSE decision-feedback equalizers and coding - Part I: Equalization results. *IEEE Trans. Communications*, COM-43:2582–2594, Oct. 1995.

[10] J. H. Conway, E. M. Rains, and N. J. A. Sloane. On the existence of similar sublattices. *Canad. J. Math*, 51:1300–1306, 1999.

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

[12] 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.

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

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] G. D. Forney, Jr. Shannon meets Wiener II: On MMSE estimation in successive decoding schemes. In *42nd Annual Allerton Conference on Communication, Control, and Computing, Allerton House, Monticello, Illinois*, Oct. 2004.

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

[25] M. Gastpar. Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network. In *ITA-2007, San Diego, CA*, pages 177–182, 2007.

[26] 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.

[27] A. Gersho and R. M. Gray *Vector Quantization and Signal Compression*. Kluwer Academic Pub., Boston, 1992.

[28] R.M. Gray, and T. Linder. Mismatch in high-rate entropy-constrained vector quantization. *IEEE Trans. Inform. Theory*, 49(5): 1204–1217, May 2003.

[29] T. Guess and M. Varanasi. An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels. *IEEE Trans. Info. Theory*, IT-51:173–187, Jan. 2005.

[30] R. M. Gray and D. L. Neuhoff, *Quantization*. IEEE Trans. Information Theory. IT-44, pp. 2325–2383, Oct. 1998.

[31] R. M. Gray and JR.T. J. Stockham, *Dithered Quantizers*. IEEE Trans. Information Theory. IT-39, pp. 805–812, May 1993.

[32] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

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

[34] Y. Kochman and R. Zamir. Joint Wyner-Ziv/dirty-paper coding using analog modulo-lattice modulation. *IEEE Trans. Info. Theory*, to appear in 2009.

[35] 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.

[36] Y. Kochman and R. Zamir. The joint Gaussian CEO/MAC problem and relaying. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control and Computing*, 2008.

[37] J. Korner 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.

[38] 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.*

[39] I. Leibowitz. The Ziv-Zakai bound at high fidelity, analog matching, and companding. Master's thesis, Tel Aviv University, Nov. 2007.

[40] 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.

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

[42] E. Martinian, G. W. Wornell, and R. Zamir. Source coding with distortion side information. *IEEE Trans. Info. Theory*, 54(10):4638–4665, Oct. 2008.

[43] N. Merhav and S. Shamai. On joint source-channel coding for the Wyner-Ziv source and the Gel'fand-Pinsker channel. *IEEE Trans. Info. Theory*, IT-40:2844–2855, Nov. 2003.

[44] B. Nazer and M. Gastpar. The case for structured random codes in network capacity theorems. *European Trans. on Comm, Special Issue on New Directions in Inform. Theory*, 19:455474, June 2008.

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

[46] 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.

[47] 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.

[48] T. Philosof, A. Khisti, U. Erez, and R. Zamir, Lattice strategies for the dirty multiple access channel. In *Proceedings of IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[49] T. Philosof, and R. Zamir, On the loss of single letter characterization: the dirty multiple access channel. *IEEE Trans. Info. Theory*, to appear in 2009.

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

[51] Z. Reznic, M. Feder, and R. Zamir. Distortion bounds for broadcasting with bandwidth expansion. *IEEE Trans. Info. Theory*, IT-52:3778–3788, Aug. 2006.

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

[53] C. A. Rogers. *Packing and Covering*. Cambridge University Press, Cambridge, 1964.

[54] B. Schein and R. Gallager. The Gaussian parallel relay network. In *ISIT-2000, Sorrento, Italy*, page 22, 2000.

[55] S. Sridharan, A. Jafarian, S. Vishwanath, S. A. Jafar, and S. Shamai. A layered lattice coding scheme for a class of three user Gaussian interference channels. 2008.

[56] 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.

[57] H.S. Witsehausen, and A.D. Wyner. Interframe coder for video signals. US Patent number 4191970, 1980.

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

[59] 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.

[60] S.-Y. Chung W. Nam and Y. H. Lee. Capacity bounds for two-way relay channels. In *Proceedings of the 20th Biennial International Zurich Seminar on Commununication (IZS 2008), Zurich, Switzerland*, 2008.

[61] H. Yao and G. W. Wornell. Lattice-reduction-aided detectors for MIMO communication systems. In *Global Telecommunications Conference (GLOBECOM '02)*, volume 1, pages 424–428, Nov. 2002.

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

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

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

[65] 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.

[66] 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.

[67] 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.

[68] Z. Zhang, E.H. Yang, and V. Wei. The redundancy of source coding with a fidelity criterion - Part one: Known statistics. *IEEE Trans. Info. Theory*, IT-43:71–91, Jan. 1997.

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