Nested Linear/Lattice Codes for Structured Multiterminal Binning

Ram Zamir, *Senior Member, IEEE*, Shlomo Shamai (Shitz), *Fellow, IEEE*, and Uri Erez, *Associate Member, IEEE*

Invited Paper

Dedicated to the memory of Aaron Wyner, with deep respect and admiration.

***Abstract*—Network information theory promises high gains over simple point-to-point communication techniques, at the cost of higher complexity. However, lack of structured coding schemes limited the practical application of these concepts so far. One of the basic elements of a network code is the binning scheme. Wyner and other researchers proposed various forms of *coset codes* for efficient binning, yet these schemes were applicable only for lossless source (or noiseless channel) network coding. To extend the algebraic binning approach to lossy source (or noisy channel) network coding, recent work proposed the idea of *nested codes*, or more specifically, nested parity-check codes for the binary case and nested lattices in the continuous case. These ideas connect network information theory with the rich areas of linear codes and lattice codes, and have strong potential for practical applications. We review these recent developments and explore their tight relation to concepts such as combined shaping and precoding, coding for memories with defects, and digital watermarking. We also propose a few novel applications adhering to a unified approach.**

***Index Terms*—Binning, digital watermarking, error-correcting codes, Gelfand–Pinsker, memory with defects, multiresolution, multiterminal, nested lattice, side information, Slepian–Wolf, writing on dirty paper, Wyner–Ziv.**

I. Introduction

NETWORK information theory generalizes Shannon's original point-to-point communication model to systems with more than two terminals. This general framework allows to consider transmission of more than one source, and/or over more than one channel, possibly using auxiliary signals ("side information") to enhance performance. Existing theoretical results, although still partial, show strong potentials over conventional point-to-point communication techniques, at the cost of higher complexity. Classic problems in this theory are the multiple-access channel, the broadcast channel, multiterminal coding of correlated sources, the interference channel, and coding with side information. See [90], [3], [24], [21] for tutorials. Until now, however, most of these solutions have remained at the theoretical level, with the exception of, perhaps, the multiple-access channel for which theory and practice meet quite closely in cellular communication. Thus, communication systems ignore much of the useful information available about the topology and the statistical dependence between signals in the network.

One of the key elements in the solutions of information network problems is the idea of "binning" [21]. A binning scheme divides a set of codewords into subsets ("bins"), such that the codewords in each subset are as far apart as possible. As usual in the "direct" coding theorems in information theory, the proof constructs the bins at *random*, and therefore characterizes the scheme in probabilistic terms: the probability that some vector is close to (or "jointly typical" with) more than one codeword in a given bin is very small or high, depending on the application. This random construction, although convenient for analysis, is not favorable for practical applications.

The main goal of this work is to show that binning schemes may have structure. Our ideas originate from Wyner's linear coset code interpretation for the Slepian–Wolf solution [76], [90]. Wyner's construction may be thought of as an *algebraic* binning scheme for *noiseless* coding problems, i.e., a scheme that can be described in terms of a parity-check code and algebraic operations over a finite alphabet. His solution applies directly to *lossless* source coding where the decoder has access to an additive-noise side-information channel. In a dual fashion, this solution applies also to channel coding over an additive-noise channel with an input constraint, and where the encoder (but not the decoder) has *perfect* side information about the channel noise. See Section II-A.

Another example for a coset-code-based binning scheme is the Kuznetsov–Tsybakov code for a memory with defective cells [55]. Similarly to the additive noise problem previously discussed, the encoder has perfect knowledge about the defect location, which is completely unknown to the decoder. See [47] for a generalization of this model.

In common applications, however, source coding is often *lossy*, while channel coding is done with *im*perfect knowledge

Manuscript received October 15, 2001; revised March 4, 2002. This work was supported in part by the Israel Academy of Science. The material in this paper was presented in part at ISITA 96, Victoria, BC, Canada; ITW 98, Killarney, Ireland; ISITA 2000, Honolulu, HI; and ISIT 2001, Washington, DC.

R. Zamir and U. Erez are with the Department of Electrical Engineering–Systems, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel (e-mail: zamir@eng.tau.ac.il; uri@eng.tau.ac.il).

S. Shamai (Shitz) is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: sshlomo@ee.technion.ac.il).

Communicated by J. Ziv, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04802-2.

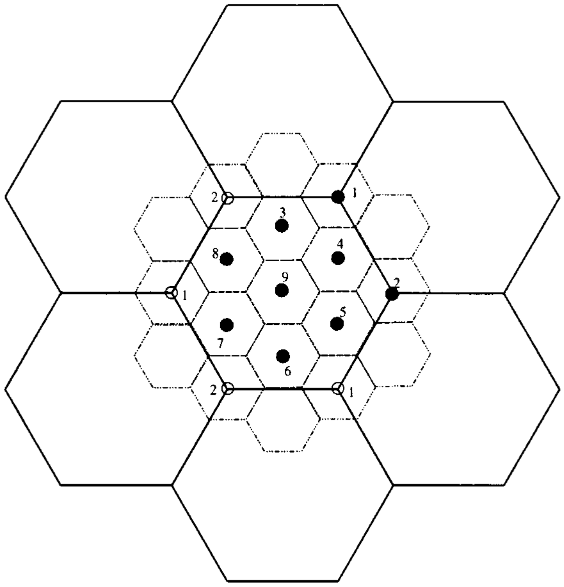

Fig. 1. Nested lattices: special case of self-similar lattices.

of the channel conditions or noise. In order to extend the idea of coset-code-based binning to "noisy" coding problems, we introduce the structure of *nested codes*, or more specifically, *nested linear codes* for the discrete case, and *nested lattices* for the continuous case. The idea is, roughly, to generate a diluted version of the original coset code; see Fig. 1. This structure allows one to construct algebraic binning schemes for more general coding applications, such as rate-distortion with side information at the decoder (the Wyner–Ziv problem) [91], and its dual problem of channel coding with side information at the encoder (the Shannon/Gelfand–Pinsker problems) [67], [41].

Specifically, nested codes apply to symmetric versions of the Wyner–Ziv problem, and to important special cases of the Gelfand–Pinsker problem such as "writing on dirty paper" (the Costa problem) [18], and "writing to a memory with known defects and unknown noise" (the Kuznetsov–Tsybakov/Heegard–El-Gamal problem) [55], [81], [47], [45]. In addition, nested codes can be used as algebraic building blocks for more general network configurations, such as multiterminal lossy source coding [3], coordinated encoding over mutually interfering channels (and specifically broadcast over Gaussian channels) [8], [98], [99], digital watermarking [9], and more. Nested lattices turn out also as a unifying model for some classical point-to-point coding techniques: constellation shaping for the additive white Gaussian noise (AWGN) channel, and combined shaping and precoding for the intersymbol interference (ISI) channel; see [29], [30], [27], [28] for background.

Nesting of codes is not a new idea in coding theory and digital communication. Conway and Sloane used nested lattice codes in [17] for constellation shaping. Forney extended and generalized their construction in [50], results which were subsequently applied to trellis shaping [51], trellis precoding [33], [12], etc. Related notions can be found in multilevel code constructions, proposed by Imai and Hirakawa [48], as well as in the work of Ungerboek and others for *set partitioning* in coded modulation [82]. In the lattice literature, Constructions B–E are all multilevel constructions [16], [52], [53].

In the context of network information theory, nested codes were proposed by Shamai, Verdú, and Zamir [71], [73], [72], [97] as an algebraic solution for the Wyner–Ziv problem. Their original motivation was systematic lossy transmission. Interestingly, the nested code structure is implicit already in Heegard's coding scheme for a memory with (a certain type of) defects [45], a problem which is a special case of channel coding with side information at the encoder. Willems proposed a scalar version of a nested code for channels with side information at the transmitter [88]. Barron, Chen, and Wornell [9], [1] showed the application of multidimensional nested codes to these channels as well as to digital watermarking. Independently of this work, Pradhan and Ramchandran [63] proposed similar structures for multiterminal source coding. Servetto [102] proposed explicit good nested lattice constructions for Wyner–Ziv encoding. Chou, Pradhan, and Ramchandran [11], Barron, Chen, and Wornell [1], and Su, Eggers, and Girod [77] pointed out the duality between the Wyner–Ziv problem and channel coding with side information at the encoder, and suggested using similar codes for both problems. A formal treatment of this duality under various side-information conditions is developed by Chiang and Cover [20].

This paper attempts to serve the dual roles of a focused tutorial and a unifying framework for algebraic coding schemes for symmetric/Gaussian multiterminal communication networks. We hope it gives a reliable coverage for this new and exciting area along with providing insights and demonstrating new applications. While demonstrating the effectiveness of the algebraic nested coding approach, we emphasize that for general (nonsymmetric/non-Gaussian) networks, this approach is not always suitable or it is inferior to *random* binning with *probabilistic* encoding–decoding. The paper is organized as follows. Section II considers noiseless side information problems associated with binary sources and channels, and describes Wyner's coset coding scheme. Section III introduces the basic definitions and properties of nested codes, for both the binary-linear case and the continuous-lattice case and discusses ways to construct such codes. Section IV uses nested codes to extend the discussion of Section II to noisy side information: the Wyner–Ziv, Costa, and Kuznetsov–Tsybakov–Heegard–El-Gamal problems. Section IV also discusses a hybrid approach of nested coding with probabilistic decoding. The rest of the paper describes various applications. Sections V and VI use the building blocks of Section IV for more general multiterminal communication problems. Section VII shows how these ideas reflect back on *point-to-point* communication problems, which include the standard additive and the dispersive Gaussian channels as well as multiple-input–multiple-output (MIMO) Gaussian channels.

II. Wyner's Noiseless Binning Scheme

A. Two Dual Side Information Problems

Figs. 2 and 3 show two problems of noiseless coding with side information, which involve binary sources and channels. As we shall see, if we make the correspondence $p \leftrightarrow \delta$, the problems and their solutions become dual [11], [1], [77], [10].

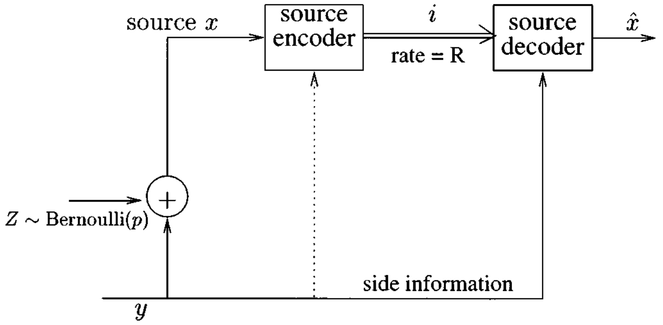

Fig. 2. Source coding with side information at decoder (SI = noisy version of source via binary-symmetric channel (BSC)).

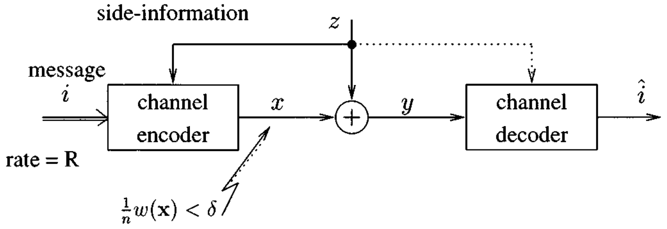

Fig. 3. BSC coding subject to an input constraint with side information at encoder (SI = channel noise).

The first problem, lossless source coding with side information at the decoder, is an important special case of the Slepian–Wolf setting [76]. A memoryless binary symmetric source $\{X_i\}$ is encoded and sent over a binary channel of rate R. The decoder needs to recover the source losslessly in the sense that its output $\{\hat{X}_i\}$ is equal to $\{X_i\}$ with probability larger than $1 - \varepsilon$ for some small positive number ε. In addition to the code at rate R, the decoder has access to a correlated binary source $\{Y_i\}$, generated by passing the source through a binary-symmetric channel (BSC) with crossover probability $0 \leq p \leq 1/2$. Were the side information available to the encoder as well, the encoder could use a "conditional" code of rate arbitrarily close to the conditional entropy

$$R = H(X|Y) = H(p) \tag{2.1}$$

for sufficiently large block length n, where

$$H(p) = -p\log(p) - (1-p)\log(1-p)$$

and where all logarithms are taken to base two. This result is a direct consequence of the conditional form of the asymptotic equipartition property (AEP) [21]: for each typical side-information sequence (known to both the encoder and the decoder) the source sequence belongs with high probability to a set of roughly $2^{nH(X|Y)}$ members, and thus can be described by a code at rate close to $H(X|Y)$.

The interesting result of Slepian and Wolf [76] shows that this rate can be approached even if the encoder does not have access to the side information. The idea underlying Slepian and Wolf's result is to randomly assign the source sequences to $2^{n[H(X|Y)+\varepsilon]}$ bins with a uniform probability, and reveal this partition to the encoder and the decoder. The encoder describes a source sequence by specifying the bin to which it belongs; the decoder looks for a source sequence in the specified bin that is *jointly typical* with the side-information sequence $\{Y_i\}$. The AEP guarantees that the true source sequence would pass this joint-typicality test. As for the other $\approx 2^{nH(X|Y)}$ source sequences which are jointly typical with the side information, the probability that the random binning scheme would assign any of them to the specified bin is very small; see [21].

Hence, as in other proofs by random coding, the proof shows that a good coding scheme *exists*. The proof even hints at a desired property of the binning scheme: it should not put together in one bin vectors which are "close" to (typical with) the same $\{Y_i\}$. In other words, each bin should play the role of a *good channel code*. However, the proof does not show how to *construct* a binning scheme with enough structure to allow efficient encoding and decoding. Can a good binning scheme have structure?

We shall soon see that indeed it can. To acquire some feeling for that in some hypothetic problem, suppose a party A wishes to specify an integer number to another party B, who knows a neighboring number, but A does not know which of its two neighboring numbers B has. An efficient solution, which requires only one bit of information—just as if A knew B, is the following. Tell B, supposing, say, that A is even, whether it divides by four or not (for a general integer, A tell B the result of $\lfloor A/2 \rfloor \bmod 2$). In terms of the Slepian–Wolf code above, this coding scheme partitions the even numbers into two bins, one of multiples of four and one of nonmultiples of four. In other words, the bins partition the source space into *lattice cosets*.

Before describing Wyner's algebraic binning scheme for the configuration of Fig. 2, let us consider the second problem, described in Fig. 3, of channel coding with perfect side information at the encoder. Here, we need to send information across a binary-symmetric channel, where the encoder knows *in advance* the channel noise sequence $\boldsymbol{z}$, i.e., the times at which the channel will invert the input bits. The decoder does not have this knowledge. To sharpen the ideas of this example, we shall assume that the channel crossover probability is half, i.e., $\boldsymbol{Z}$ is a Bernoulli-$1/2$ process. Suppose the encoder output $\boldsymbol{X}$ must satisfy the constraint (the equivalent, in essence, of the "power" constraint in the continuous case) that the average number of $\mathbf{1}$'s cannot exceed $n\delta$, where n is the block length and $0 \leq \delta \leq 1/2$. Now, if the side information were available to the decoder as well, it could cancel out the effect of the channel noise alltogether by XORing $\boldsymbol{y} \oplus \boldsymbol{z}$, and thus achieve capacity of

$$C = \max_{E\{\omega_H(\boldsymbol{X})\} \leq n\delta} \frac{1}{n} H(\boldsymbol{X}) = H(\delta) \tag{2.2}$$

where $\omega_H(\cdot)$ denotes the Hamming weight (number of $\mathbf{1}$'s). Due to the input constraint, however, the noise cannot be subtracted by the *encoder*; by XORing $\boldsymbol{x} \oplus \boldsymbol{z}$ the channel input vectors would have an average weight of

$$E\{\omega_H(\boldsymbol{x} \oplus \boldsymbol{Z})\} = n/2 \tag{2.3}$$

for any $\boldsymbol{x}$ sequence, thus violating the input constraint. On the other hand, ignoring the side information would nullify the capacity. Can the encoder make any use of knowing the noise Z?

Indeed, the result of Gelfand and Pinsker [41] implies that with a clever binning scheme we can achieve capacity of $H(\delta)$

even if the decoder does not have access to the side information $\boldsymbol{z}$, and without violating the input constraint. The idea is to randomly assign the 2^n possible binary n-vectors to $2^{n[H(\delta)-\varepsilon]}$ bins, and reveal this partition to the encoder and the decoder. The message to be sent specifies the bin. The encoder looks in that bin for a vector $\boldsymbol{u}$ whose Hamming distance from $\boldsymbol{z}$ is at most $n\delta$, and outputs the difference vector $\boldsymbol{x} = \boldsymbol{z} \oplus \boldsymbol{u}$. The decoder who receives $\boldsymbol{y} = \boldsymbol{x} \oplus \boldsymbol{z} = \boldsymbol{u}$ identifies the bin containing $\boldsymbol{u}$, and thus decodes the message unambiguously. Hence, we achieve a rate of $H(\delta) - \varepsilon$ under the desired input constraint provided that at least one vector in the bin is within a distance $n\delta$ from $\boldsymbol{z}$; indeed, by random selection of bins, for sufficiently large n it is very likely to find such a vector.

This solution for the configuration of Fig. 3 shows another angle of the desired property of a good binning scheme: each bin should contain a good collection of representative points which spread over the entire space. In other words, here each bin plays the role of a *good source code*. Again, however, random binning lacks structure, and therefore it is not practically efficient.

B. Parity-Check Codes

We now turn to show an algebraic construction for these two binning schemes. Following the intuition underlying the Slepian–Wolf solution, Wyner's basic idea in [90] was to generate the bins as the cosets of some "good" parity-check code.

To introduce Wyner's scheme, let an (n, k) binary parity-check code be specified by the $(n-k) \times n$ (binary) parity-check matrix H. The code $\mathcal{C} = \{\boldsymbol{c}\}$ contains all n-length binary vectors $\boldsymbol{c}$ whose syndrome $\boldsymbol{s} \stackrel{\Delta}{=} H\boldsymbol{c}$ is equal to zero, where here multiplication and addition are modulo 2. Assuming that all rows of H are linearly independent, there are 2^k codewords in $\mathcal{C}$, so the code rate is $(\log|\mathcal{C}|)/n = k/n$.

Given some general syndrome $\boldsymbol{s} \in \{0, 1\}^{n-k}$, the set of all n-length vectors $\boldsymbol{x}$ satisfying $H\boldsymbol{x} = \boldsymbol{s}$ is called a coset $\mathcal{C}_s$. The decoding function $f(\boldsymbol{s})$, where f: $\{0, 1\}^{n-k} \rightarrow \{0, 1\}^n$, is equal to the vector $\boldsymbol{v} \in \mathcal{C}_s$ with the minimum Hamming weight, where ties are broken arbitrarily. It follows from linearity that the coset is a shift of the code $\mathcal{C}$ by the vector $\boldsymbol{v}$, i.e.,

$$\mathcal{C}_s \stackrel{\Delta}{=} \{\boldsymbol{x}\colon H\boldsymbol{x} = \boldsymbol{s}\} = \{\boldsymbol{c} \oplus \boldsymbol{v}\colon \boldsymbol{c} \in \mathcal{C}\} \stackrel{\Delta}{=} \mathcal{C}^v \tag{2.4}$$

where the n-vector $\boldsymbol{v} = f(\boldsymbol{s})$ is called the *coset leader*.

Maximum-likelihood decoding of a parity-check code, over a BSC $y = x \oplus z$, amounts to quantizing $\boldsymbol{y}$ to the nearest vector in $\mathcal{C}$ with respect to the Hamming distance. This vector, $\hat{\boldsymbol{c}}$, can be computed by a procedure, called "syndrome decoding," which follows from the definition of the function f

$$\hat{\boldsymbol{c}} = \boldsymbol{y} \oplus \hat{\boldsymbol{z}}, \qquad \hat{\boldsymbol{z}} = f(H\boldsymbol{y}). \tag{2.5}$$

Hence, $f(H\boldsymbol{y})$ is the maximudm-likelihood estimate of the channel noise $\boldsymbol{z}$. Alternatively, we can interpret $f(H\boldsymbol{y})$ as the error vector in quantizing $\boldsymbol{y}$ by $\mathcal{C}$, or as reducing $\boldsymbol{y}$ modulo the code

$$f(H\boldsymbol{y}) = \boldsymbol{y} \bmod \mathcal{C}. \tag{2.6}$$

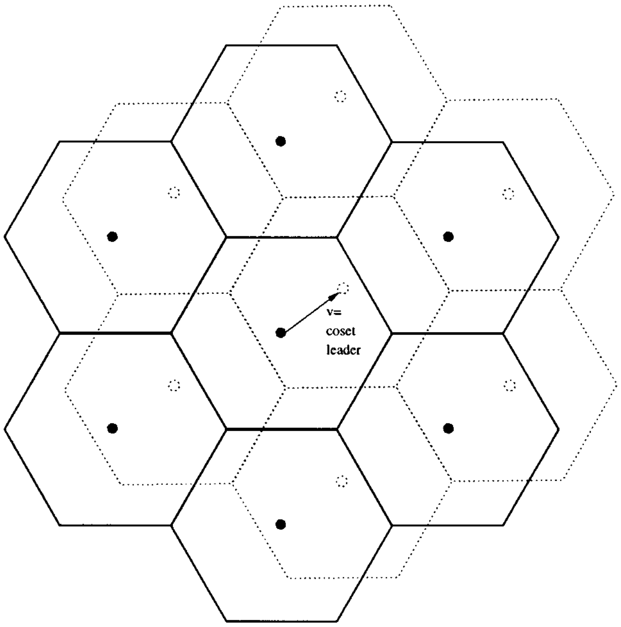

Fig. 4. Geometric interpretation of a parity-check code (solid) and one of its cosets (dashed) and their associated decision cells.

Fig. 4 illustrates the interrelations between a parity-check code and its cosets by interpreting the codewords as points of a two-dimensional hexagonal lattice. We may view the decoder (or quantizer) above as a partition of $\{0, 1\}^n$ to 2^k decision cells of size 2^{n-k} each, which are all shifted versions of the basic "Voronoi" set

$$\{\boldsymbol{z}\colon \boldsymbol{z} \oplus f(H\boldsymbol{z}) = 0\} \stackrel{\Delta}{=} \Omega_0. \tag{2.7}$$

Each of the 2^{n-k} members of Ω_0 is a coset leader (2.4) for a different coset.

An important asymptotic property of parity-check codes is that there exist "good" codes among them. Here "good" may have one of the following two definitions:

i) *Good channel codes over BSC(p):* For any $\varepsilon > 0$ and n large enough there exists an (n, k) code of rate $k/n > C - \varepsilon$, where $C = 1 - H(p)$ is the BSC(p) capacity, with a probability of decoding error smaller than ϵ

$$\Pr\{\hat{\boldsymbol{Z}} \neq \boldsymbol{Z}\} = \Pr\{f(H\boldsymbol{Z}) \neq \boldsymbol{Z}\} < \epsilon \tag{2.8}$$

where $\boldsymbol{Z}$ denotes the channel noise vector (a Bernoulli (p) vector), and $\hat{\boldsymbol{Z}}$ denotes its estimation (2.5). See [39]. We call such a code a "good BSC p-code."

ii) *Good source codes under Hamming distortion:* For any $0 \leq \delta \leq 1/2$, $\varepsilon > 0$, and sufficiently large n, there exists an (n, k) code of rate $k/n < R(\delta) + \varepsilon$, where $R(\delta) = 1 - H(\delta)$ is the rate-distortion function of a binary-symmetric source (BSS) $\boldsymbol{X}$, such that the expected quantization error Hamming weight satisfies

$$\frac{1}{n} E\left\{\omega_H\left(\boldsymbol{X} \oplus \hat{\boldsymbol{X}}\right)\right\} = \frac{1}{n} E\{\omega_H(\boldsymbol{U})\} < \delta + \varepsilon \tag{2.9}$$

where $\hat{\boldsymbol{X}}$ denotes the quantization of $\boldsymbol{X}$ by the code, and where $\boldsymbol{U} = \boldsymbol{X} \oplus \hat{\boldsymbol{X}} = f(H\boldsymbol{X})$ is the quantization error,

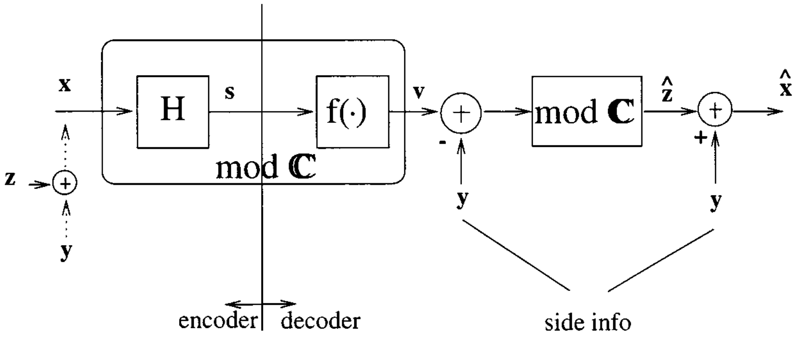

Fig. 5. Wyner's coset coding scheme for the binary-symmetric Slepian–Wolf problem in terms of modulo-code operations (both the $+$ and $-$ signs amount to the XOR operation).

which is uniformly distributed over Ω_0. We call such a code a "good BSS δ-code."

The geometric meaning of the asymptotic properties (2.8) and (2.9) is that the decision cells of a good $(k,\, n)$-parity-check code are approximately *Hamming balls* of radius np (or $n\delta$), where

$$k/n + H(p) \approx 1. \tag{2.10}$$

See [21]. Random parity-check arguments as in [39, Sec. 6.2] imply that the same code can be simultaneously good in both senses.

Another measure of goodness of a linear code, not necessarily asymptotic, is its *erasure* correction capability. For general q-ary alphabets (not necessarily binary) there exist $(n,\, k)$ codes, called minimum-distance separable (MDS) codes, which can correct $n-k$ erasures [6]. Asymptotically, good binary $(n,\, k)$ codes correct almost every pattern of $n-k$ erasures.

C. Coset Codes as Bins

Consider now the use of these algebraic structures for the two perfect side information problems of Section II-A. In the sequel, we will need to compute the error between a vector $\boldsymbol{y}$ and a coset $\mathcal{C}_s$

$$\boldsymbol{e} = \arg\min_{\boldsymbol{x}\in\mathcal{C}_s} \omega_H(\boldsymbol{x}\oplus\boldsymbol{y}). \tag{2.11}$$

Making the substitutions $\boldsymbol{x}' = \boldsymbol{x}\oplus\boldsymbol{y}$ and $\mathcal{C}_{s'} = \boldsymbol{y}\oplus\mathcal{C}_s$, we see that $\boldsymbol{e}$ minimizes the Hamming weight in $\mathcal{C}_{s'}$, thus by the definition of the decoding function and the $\bmod$-$\mathcal{C}$ operation in (2.6)

$$\begin{aligned}\boldsymbol{e} &= f(\boldsymbol{s}') && (2.12)\\ &= f(\boldsymbol{s}\oplus H\boldsymbol{y}) && (2.13)\\ &= \boldsymbol{y} \bmod \mathcal{C}_s && (2.14)\\ &= (\boldsymbol{v}\oplus\boldsymbol{y}) \bmod \mathcal{C} && (2.15)\end{aligned}$$

where $\boldsymbol{v} = f(\boldsymbol{s})$ is the coset leader associated with $\boldsymbol{s}$.

In the setting of lossless source coding with side information at the decoder (Fig. 2), we choose a good BSC p-code, and use as bins its $2^{n-k} \approx 2^{nH(p)}$ cosets. The encoding and decoding can be described by simple algebraic operations.

Encoding: transmit the syndrome $\boldsymbol{s} = H\boldsymbol{x}$; this requires $n-k \approx nH(p)$ bits.

Decoding: find the point in the coset $\mathcal{C}_s$ which is closest to the side information $\boldsymbol{y}$; by (2.12) this can be computed as

$$\hat{\boldsymbol{x}} = \boldsymbol{y}\oplus\hat{\boldsymbol{z}}, \qquad \text{where } \hat{\boldsymbol{z}} = f(\boldsymbol{s}\oplus H\boldsymbol{y}). \tag{2.16}$$

Note that the computation of (2.16) is unique, so unlike in *random* binning we never have ambiguous decoding. Hence, letting $\boldsymbol{z} = \boldsymbol{x}\oplus\boldsymbol{y}$ and noting from (2.16) that $\hat{\boldsymbol{z}} = f(H(\boldsymbol{x}\oplus\boldsymbol{y})) = f(H\boldsymbol{z})$, a decoding error event amounts to $\{\hat{\boldsymbol{x}}\neq\boldsymbol{x}\} \Longleftrightarrow \{\hat{\boldsymbol{z}}\neq\boldsymbol{z}\}$ so the probability of decoding error is

$$\Pr\{\hat{\boldsymbol{X}}\neq\boldsymbol{X}\} = \Pr\{f(H\boldsymbol{Z})\neq\boldsymbol{Z}\} \tag{2.17}$$

which by (2.8) is smaller than ε for a good BSC p-code. This is actually the probability that $\boldsymbol{Y}$ exceeds the cell Ω_0 shifted by $\boldsymbol{x}$, or that $\boldsymbol{Z} \bmod \mathcal{C} \neq \boldsymbol{Z}$.

Thus, we were able to encode $\boldsymbol{x}$ at rate close to $H(X|Y) = H(p)$, with a small probability of decoding error, using side information at the decoder, as desired.

Fig. 5 shows a useful way to describe the functioning of this coding scheme in terms of the "modulo-code" operation (2.6), using the identity (2.15). The modulo-code operation satisfies a distributive property [6]

$$((\boldsymbol{x} \bmod \mathcal{C}) + \boldsymbol{y}) \bmod \mathcal{C} = (\boldsymbol{x}+\boldsymbol{y}) \bmod \mathcal{C}, \qquad \forall \boldsymbol{x},\, \boldsymbol{y}. \tag{2.18}$$

Now, note that the successive operations H and $f(\cdot)$ at the beginning of the signal path are equivalent to a single $\bmod$-$\mathcal{C}$ operation. Hence, by the distributive property, due to the $\bmod$-$\mathcal{C}$ operation later in the signal path, we can eliminate these H and $f(\cdot)$ operations without affecting the output of the scheme. We then see immediately that $\hat{\boldsymbol{z}} = \boldsymbol{z} \bmod \mathcal{C}$.

We turn to the dual setting of channel coding with perfect side information at the encoder (Fig. 3). Here we choose a good BSS-δ-code, and, again, use its $2^{n-k} \approx 2^{nH(\delta)}$ cosets as bins. The encoding and decoding can be described in algebraic terms as follows.

Message selection: identify each syndrome $\boldsymbol{s}$ with a unique message; this amounts to $n-k \approx nH(\delta)$ information bits.

Encoding: transmit the error vector between the side information $\boldsymbol{z}$ and the message coset $\mathcal{C}_s$, i.e. (see (2.12))

$$\begin{aligned}\boldsymbol{x} &= \boldsymbol{z} \bmod \mathcal{C}_s && (2.19)\\ &= f(\boldsymbol{s}\oplus H\boldsymbol{z}) && (2.20)\\ &= (\boldsymbol{v}\oplus\boldsymbol{z}) \bmod \mathcal{C} && (2.21)\end{aligned}$$

where $\boldsymbol{v} = f(\boldsymbol{s})$.

Decoding: reconstruct the message as the syndrome $\hat{\boldsymbol{s}} = H\boldsymbol{y}$.

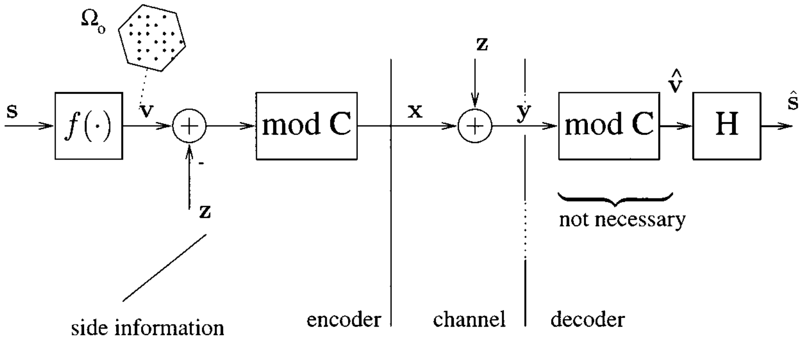

Fig. 6. Coset-based scheme for channel coding with perfect side information (Fig. 3).

It is easy to verify that the decoding is perfect, i.e.,

$$\hat{\boldsymbol{s}} = H \cdot (\boldsymbol{x} \oplus \boldsymbol{z}) = H \cdot f(\boldsymbol{s} \oplus H\boldsymbol{z}) \oplus H\boldsymbol{z} = \boldsymbol{s} \oplus H\boldsymbol{z} \oplus H\boldsymbol{z} = \boldsymbol{s} \tag{2.22}$$

due to the identity $H \cdot f(\boldsymbol{s}') \equiv \boldsymbol{s}' \, \forall \boldsymbol{s}' \in \{0, 1\}^{n-k}$. Moreover, for any $\boldsymbol{s}$, the average transmission Hamming weight satisfies

$$\frac{1}{n} E\omega_H(f(\boldsymbol{s} \oplus H\boldsymbol{Z})) < \delta + \varepsilon \tag{2.23}$$

by the BSS δ-goodness of the code and the symmetry of $\boldsymbol{Z}$.[1] Thus, we were able to transmit at rate $\approx H(\delta)$ with input constraint δ, using side information at the encoder, as desired.

Fig. 6 shows an equivalent formulation of this scheme in terms of modulo-$\mathcal{C}$ operations. For illustration purposes, we have inserted a second mod-$\mathcal{C}$ operation that does not affect the output. As in Fig. 5, the functioning of the scheme becomes transparent by applying the distributive property (2.18) of the mod-$\mathcal{C}$ operation, and eliminating the first mod-$\mathcal{C}$ operation. It immediately follows that the noise cancels out so that $\hat{\boldsymbol{v}} = \boldsymbol{v} \bmod \mathcal{C} = \boldsymbol{v}$, and clearly $\hat{\boldsymbol{v}} = \boldsymbol{v}$ implies $\hat{\boldsymbol{s}} = \boldsymbol{s}$.

D. Other Variants

1) Nonsymmetric Channels and Sources: We can generalize the two side information problems discussed throughout this section in various ways. One way is to consider more general distributions for the signals in the system. It is clear from the equivalent formulation of Wyner's scheme in Fig. 5, that the scheme is insensitive to the structure of the $\boldsymbol{y}$ vector, as long as $\boldsymbol{X}$ is obtained by passing the side information $\boldsymbol{y}$ through a BSC. Likewise, it is easily seen from Fig. 6 that in the second problem the side-information signal $\boldsymbol{z}$ may be arbitrary; only, to ensure that the δ-input constraint is satisfied, we need to smooth out the effect of adding $\boldsymbol{z}$ using a technique called "dithering" before applying the mod-$\mathcal{C}$ operation at the encoder; see Section IV. It follows from this discussion, that the same schemes can achieve the optimum rates of $H(p)$ in the former case and $H(\delta)$ in the latter case for *arbitrarily varying* side-information signals.

Note, however, that if the channel connecting X and Y in the first problem is *non*symmetric, or if the input constraint in the second problem is more complex (e.g., depends on Z or has memory), then the algebraic binning schemes above are no longer optimal. This is similar to the difficulty of applying parity-check codes to general, nonsymmetric channels, or to *non*difference distortion measures for source coding.

2) Digital Watermarking/Information Embedding: The algebraic construction for channel coding with perfect side information is based on the equivalent formulation of digital watermarking by Barron, Chen, and Wornell [1] and by Chou, Pradhan, and Ramchandran [11]. In these formulations, the side-information signal $\boldsymbol{z}$ is considered as a "host" signal, which carries information under the constraint that the Hamming *distortion* due to the watermark code should not exceed δ. An extension of this setting to watermarking in the presence of noise is equivalent to the nonperfect side information case (the Costa problem) which we discuss in Section IV. See [14], [10], [1], [11] for more settings and literature about the digital watermarking problem and its equivalence to channel coding with side information.

3) Writing to Computer Memory With Defects: Another well-known example of coset-code-based binning is that of computer memory with defects [55], [47]. Here, k out of n binary digits are stuck at arbitrary positions, so the encoder can write new information only at the remaining $n-k$ binary digits. The location of the defective cells is arbitrary, and is detected by the encoder prior to writing. Various authors (mostly in the Russian literature) developed schemes and performance bounds for this channel model, and showed that it is possible to achieve the capacity of $n-k$ bits, even if the location of the k defective cells is not known to the decoder. See [47], [106], and the references therein.

To prove this fact asymptotically by a random binning argument, assume that the binary n-vectors are randomly assigned to $2^{n-k-\epsilon n}$ bins, where $\epsilon > 0$. This assignment is fixed prior to encoding. A message containing $n-k-\epsilon n$ bits selects the bin. The encoder looks for a vector in the selected bin which agrees with the values of the k defective cells, and writes this vector to the memory. Since each vector identifies a unique bin, the decoder can decode the message correctly, provided that the encoder indeed finds a "defect-matching" vector in the selected bin. Otherwise, an encoding error is declared. A standard calculation shows that the probability of an error event is given by

$$\left[1 - 2^{-[n-k-\epsilon n]}\right]^{2^{n-k}}$$

[1] Dithering can be used to guarantee (2.23) for a nonsymmetric $\boldsymbol{Z}$; see the discussion in the sequel.

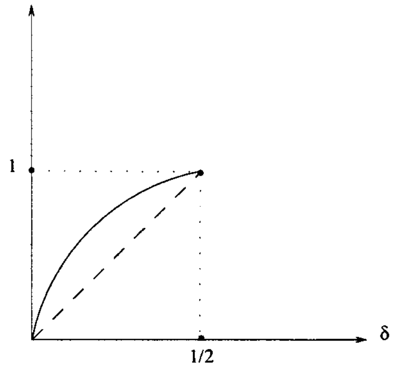

Fig. 7. Capacity with causal (dashed), and noncausal (solid) SI at the encoder in the setting of Fig. 3.

which goes to zero as n goes to infinity for any $\epsilon > 0$. (There are 2^{n-k} valid n-vectors, and each of them is in the selected bin with probability $1/2^{n-k-\epsilon n}$.)

An algebraic coding scheme which achieves this capacity uses the cosets of an *erasure correction code* as bins. Specifically, assume a q-ary linear code of length n which can correct $n-k$ erasures (each erasure is a full q-ary symbol). If the code is MDS (e.g., a Reed–Solomon code) [6], then it contains q^k codewords. This implies that each fixed pattern of k symbols takes all the q^k possible values as we scan the q^k codewords; furthermore, the code has q^{n-k} distinct cosets (including the original code), each of which satisfies this property. It follows that if we use these cosets as bins, the encoder can find a defect-matching vector in each bin, for any pattern of k defective cells.

A noisy generalization of this problem will be discussed in Section IV-D.

4) Causal Side Information: Shannon proposed the perfect binary side information problem, without the input constraint, as a motivation for his treatment of the *causal* side information case [67]. Unlike the setting in Fig. 3, in Shannon's formulation the channel input depends only on the past and present samples of the channel noise, i.e.,

$$x_i = \varphi_i(w, z_1, \ldots, z_i), \qquad i = 1 \cdots n$$

where w denotes the message to be sent. It follows from the analysis of Erez *et al.* [30], [28], [27] that the capacity with *causal* side information and input constraint δ as above, is given by $C = 2\delta, 0 \le \delta \le \frac{1}{2}$. See the dashed line in Fig. 7. This capacity, which is of course lower than that achieved by the *non*causal binning scheme solution, is realized by appropriate time sharing of two strategies: $x_i = u_i \oplus z_i$ ("perfect precancellation") a fraction 2δ of the transmission time, and $x_i \equiv 0$ ("idle") a fraction $(1-2\delta)$ of the transmission time, where $u_1, u_2, \ldots, u_{n2\delta}$ are the information bits, i.e., $w = (u_1, \ldots, u_{n2\delta})$.

III. Nested Codes: Preliminaries

The binning schemes discussed so far are not suitable for "noisy" coding situations, i.e., source coding with *distortion*, or transmission in the presence of an *unknown* (random) noise component. In the noiseless case, the cosets (= bins) filled the binary space completely. To allow further compression in source coding, or noise immunity in channel coding, we need to dilute the coset density in space. Nested parity-check codes generate such a diluted system of cosets in an efficient way.

The continuous analog of a parity-check code is the lattice code. Being a construction in Euclidean space, the lattice has continuously many cosets. The notion of a nested lattice code allows to define a *finite* sample of lattice cosets efficiently. This will provide the basis for algebraic binning schemes for continuous signals.

This section establishes the basic definitions of these concepts. It is an extended and more complete version of the discussion by Zamir and Shamai [97]. We start with the binary case and nested parity-check codes, and then continue to the continuous alphabet case and nested lattice codes.

A nested code is a pair of linear or lattice codes $(\mathcal{C}_1, \mathcal{C}_2)$ satisfying

$$\mathcal{C}_2 \subset \mathcal{C}_1 \tag{3.1}$$

i.e., each codeword of $\mathcal{C}_2$ is also a codeword of $\mathcal{C}_1$. We call $\mathcal{C}_1$ the "fine code" and $\mathcal{C}_2$ the "coarse code."

A. Nested Parity-Check Codes

If a pair $\{(n, k_1), (n, k_2)\}$ of parity-check codes, $k_1 > k_2$, satisfies condition (3.1), then the corresponding parity-check matrices H_1 and H_2 are interrelated as

$$H_2 = \begin{bmatrix} H_1 \\ \cdots \\ \Delta H \end{bmatrix} \tag{3.2}$$

where H_1 is an $(n-k_1) \times n$ matrix, H_2 is an $(n-k_2) \times n$ matrix, and ΔH is a $(k_1 - k_2) \times n$ matrix. This implies that the syndromes $\boldsymbol{s}_1 = H_1\boldsymbol{x}$ and $\boldsymbol{s}_2 = H_2\boldsymbol{x}$ associated with some n-vector $\boldsymbol{x}$ are related as $\boldsymbol{s}_2^t = [\boldsymbol{s}_1^t, \boldsymbol{\Delta s}^t]$, where the length of $\boldsymbol{\Delta s}$ is $k_1 - k_2$ bits. In particular, if $\boldsymbol{x} \in \mathcal{C}_1$, then $\boldsymbol{s}_2^t = [0, \ldots, 0, \boldsymbol{\Delta s}^t]$. We may, therefore, partition $\mathcal{C}_1$ into $2^{k_1-k_2}$ cosets of $\mathcal{C}_2$ by setting $\boldsymbol{s}_1 \equiv \mathbf{0}$, and varying $\boldsymbol{\Delta s}$, i.e.,

$$\mathcal{C}_1 = \bigcup_{\boldsymbol{\Delta s} \in \{0,1\}^{k_1-k_2}} \mathcal{C}_{2, \boldsymbol{s}_2}, \qquad \text{where } \boldsymbol{s}_2^t = [\mathbf{0}, \boldsymbol{\Delta s}^t]. \tag{3.3}$$

Of fundamental importance is the question: can we require both components of a nested code, the fine code and the coarse code, to be good in the sense of (2.8) and (2.9)? More interestingly, it turns out that in the network problems discussed below, one of the component codes should be a good channel code, while the other component code should be a good source code; see the discussion in Section III-C. If a nested code is indeed "good," where the fine code is a good q_1-code and the coarse code is a good q_2-code, $q_2 > q_1$, then by (2.10) the number of cosets in (3.3) is about

$$2^{k_1-k_2} \approx 2^{n[H(q_2)-H(q_1)]} \tag{3.4}$$

where $\approx$ means approximation in an exponential sense (i.e., the difference between the normalized logarithms is small).

B. Lattices and Nested Lattice Codes

We turn to Euclidean space and to nested lattices. Let us first introduce the basic properties of a lattice code. 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.,

$$\Lambda = \{\boldsymbol{\ell} = G \cdot \boldsymbol{i} \colon \boldsymbol{i} \in \mathbb{Z}^n\} \tag{3.5}$$

where $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the $n \times n$ *generator matrix* G is given by $G = [\boldsymbol{g}_1 \,|\, \boldsymbol{g}_2 | \cdots | \, \boldsymbol{g}_n]$. Note that the zero vector is always a lattice point, and that G is not unique for a given Λ. See [16].

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

$$Q(\boldsymbol{x}) = \boldsymbol{\ell} \in \Lambda \quad \text{if } \|\boldsymbol{x} - \boldsymbol{\ell}\| \le \|\boldsymbol{x} - \boldsymbol{\ell}'\|, \qquad \forall \boldsymbol{\ell}' \in \Lambda \tag{3.6}$$

where $\|\cdot\|$ denotes Euclidean norm. In analogy with the basic decision cell Ω_0 in the binary case, the basic Voronoi cell of Λ is the set of points in $\mathbb{R}^n$ closest to the zero codeword, i.e.,

$$\mathcal{V}_0 = \{\boldsymbol{x} \colon Q(\boldsymbol{x}) = \boldsymbol{0}\} \tag{3.7}$$

where ties are broken arbitrarily. The Voronoi cell associated with each $\boldsymbol{\ell} \in \Lambda$ is a shift of $\mathcal{V}_0$ by $\boldsymbol{\ell}$. In analogy with (2.6) the $\bmod$-Λ operation is defined as

$$\boldsymbol{x} \mod \Lambda = \boldsymbol{x} - Q(\boldsymbol{x}) \tag{3.8}$$

which is also the quantization error of $\boldsymbol{x}$ with respect to Λ. The second moment of Λ is defined as the second moment per dimension of a uniform distribution over $\mathcal{V}_0$

$$\sigma^2 = \frac{1}{V} \cdot \frac{1}{n} \int_{\mathcal{V}_0} \|\boldsymbol{x}\|^2 \, d\boldsymbol{x} \tag{3.9}$$

where $V = \mathrm{Vol}\,(\mathcal{V}_0) = \int_{\mathcal{V}_0} d\boldsymbol{x}$ is the volume of $\mathcal{V}_0$. A figure of merit of a lattice code with respect to the mean squared error distortion measure is the normalized second moment

$$G(\Lambda) = \frac{1}{V^{1+2/n}} \cdot \frac{1}{n} \cdot \int_{\mathcal{V}_0} \|\boldsymbol{x}\|^2 \, d\boldsymbol{x} = \sigma^2 / V^{2/n}. \tag{3.10}$$

The minimum possible value of $G(\Lambda)$ over all lattices in $\mathbb{R}^n$ is denoted G_n. The isoperimetric inequality implies that $G_n \ge 1/2\pi e \, \forall \, n$. When used as a channel code over an unconstrained AWGN channel, [62], [30], the decoding error probability is the probability that a white Gaussian noise vector $\boldsymbol{Z}$ exceeds the basic Voronoi cell

$$P_e = \Pr\{\boldsymbol{Z} \notin \mathcal{V}_0\}. \tag{3.11}$$

The use of high-dimensional lattice codes is justified by the existence of asymptotically "good" lattice codes. As for parity-check codes in the binary case (Section II-B), we consider two definitions of goodness.

i) *Good channel codes over AWGN channel:* For any $\varepsilon > 0$ and sufficiently large n, there exists an n-dimensional lattice Λ whose cell volume $V < 2^{n[h(Z)+\varepsilon]}$, where $h(Z) = \frac{1}{2}\log(2\pi e \sigma_z^2)$ and σ_z^2 are the differential entropy and the variance of the AWGN Z, respectively, such that

$$P_e = \Pr\{\boldsymbol{Z} \notin \mathcal{V}_0\} < \varepsilon. \tag{3.12}$$

Such codes approach the capacity per unit volume of the AWGN channel, and are called "good AWGN channel σ_z^2-codes;" see [62], [16].

ii) *Good source codes under mean squared distortion measure:* For any $\varepsilon > 0$ and sufficiently large n, there exists an n-dimensional lattice Λ with

$$\log(2\pi e G_n) < \varepsilon \tag{3.13}$$

i.e., the normalized second moment of good lattice codes approaches the bound $1/2\pi e$ as n goes to infinity; see [95]. Such codes, scaled to second moment D, approach the quadratic rate-distortion function $R(D)$ at high-resolution quantization conditions [96] and are called "good source D-codes."

In analogy with the binary case, the meaning of i) and ii) is that the basic Voronoi cells of good lattice codes approximate Euclidean *balls* of radius $\sqrt{n\sigma_z^2}$ (or $\sqrt{nD}$); see [16], [95], [62]. This implies that the volume of the Voronoi cells of good δ-codes satisfies asymptotically

$$\frac{1}{n}\log V \approx \frac{1}{2}\log(2\pi e \delta) \tag{3.14}$$

where δ corresponds to σ_z^2 (or D).

It is interesting to note that a lattice which is good in one sense need not necessarily be good in the other. This is analogous to the well-known fact that lattice sphere packing is not equivalent to lattice sphere covering; see [16] and [100].

A pair of n-dimensional lattices (Λ_1, Λ_2) is nested in the sense of (3.1), i.e., $\Lambda_2 \subset \Lambda_1$, if there exists corresponding generator matrices G_1 and G_2, such that

$$G_2 = G_1 \cdot \boldsymbol{J}$$

where $\boldsymbol{J}$ 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

$$V_2 = \det\{\boldsymbol{J}\} \cdot V_1 \tag{3.15}$$

where $V_2 = \mathrm{Vol}\,(\mathcal{V}_{0,2})$ and $V_1 = \mathrm{Vol}\,(\mathcal{V}_{0,1})$. We call

$$\sqrt[n]{\det\{\boldsymbol{J}\}} = \sqrt[n]{V_2/V_1}$$

the *nesting ratio*.

Fig. 1 shows nested hexagonal lattices with $\boldsymbol{J} = 3 \cdot I$, 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 [15].

The points of the set

$$\{\Lambda_1 \mod \Lambda_2\} \triangleq \{\Lambda_1 \cap \mathcal{V}_{0,2}\} \tag{3.16}$$

are called the *coset leaders* of Λ_2 relative to Λ_1; for each $\boldsymbol{v} \in \{\Lambda_1 \mod \Lambda_2\}$ the shifted lattice $\Lambda_{2,v} = \boldsymbol{v} + \Lambda_2$ is called a *coset* of Λ_2 relative to Λ_1. Mapping of border points in (3.16) (i.e., points of Λ_1 that fall on the envelope of the Voronoi region $\mathcal{V}_{0,2}$) to the coset leader set is done in a systematic fashion, so that the cosets $\Lambda_{2,v}$, $\boldsymbol{v} \in \{\Lambda_1 \mod \Lambda_2\}$ are disjoint. It follows that there are $V_2/V_1 = \det\{J\}$ different cosets, whose union gives the fine lattice

$$\dot{\bigcup_{\boldsymbol{v}\in\{\Lambda_1 \bmod \Lambda_2\}}} \Lambda_{2,\,v} = \Lambda_1. \tag{3.17}$$

Note that for any $\ell_1 \in \Lambda_1$, reducing $\ell_1 \bmod \Lambda_2$ (see (3.8)) gives the leader of the (unique) coset which contains ℓ_1. Enumeration of the cosets can be obtained using a parity-check-like matrix [16].

As mentioned in the binary case, of fundamental importance is the question of 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 is good for source coding under mean squared distortion. See the discussion in Section III-C. If a nested lattice pair is indeed good in this sense, where the fine lattice is a good δ_1-code and the coarse lattice is a good δ_2-code, $\delta_2 > \delta_1$, then by (3.14) the number of cosets of Λ_2 relative to Λ_1 in (3.17) is about

$$|\{\Lambda_1 \bmod \Lambda_2\}| = V_2/V_1 \approx (\delta_2/\delta_1)^{n/2} \tag{3.18}$$

where $\approx$ means approximation in an exponential sense.[2]

Another special issue that arises in the application of nested codes is the "self-noise" phenomenon. In simple words, it is the immunity of the channel-code component of the nested code to noise induced by the quantization error of the source-code component. This issue will be discussed in detail in Section IV.

C. Construction of Good Nested Codes

For the binary case, existence is straightforward by the properties of random ensembles of parity-check codes [39, Sec. 6.2]. For a more explicit construction one may proceed as follows [58]. Let G_2 be the generating matrix of a $(k,\, n)$ code $\mathcal{C}_2$ that has roughly a binomial distance spectrum. This property guarantees that $\mathcal{C}_2$ is a good parity-check code. One can now add cosets to $\mathcal{C}_2$ (or equivalently, rows to G) and still retain a binomial spectrum for the new code, denoted by $\mathcal{C}_1$. Furthermore, from the construction it is evident that $\mathcal{C}_2 \subset \mathcal{C}_1$. See also Heegard's construction of partitioned Bose–Chaudhuri–Hocquenghem (BCH) codes [45].

We present a detailed construction of good nested lattice ensembles (Construction-U) in a future work [30], [31]. We shall point out here the basic elements. Our construction is based on Loeliger's construction of lattice ensembles [59], and is similar to common approaches aiming at incorporating shaping gain into coded modulation [33], [86], in that the "effective dimensionality" of the coarse and fine lattice may greatly differ. That is, at large nesting ratios it might suffice to use a relatively low-dimensional source-coding (shaping) lattice to make $\log 2\pi e G(\Lambda)$ small enough as required by (3.13). Denoting such a k-dimensional lattice by Λ_2', the construction forms the $n = m \times k$-dimensional coarse lattice by a Cartesian product of this basic lattice, i.e.,

$$\Lambda_2 = \Lambda_2' \times \Lambda_2' \times \cdots \times \Lambda_2'. \tag{3.19}$$

The fine lattice $\Lambda_1 \supset \Lambda_2$ is typically much more complex in order to achieve large coding gains, i.e., make the decoding error probability P_e small as required by (3.12). Therefore, its "effective dimension" is n.[3]

Loeliger's construction is based on drawing a random n-dimensional code over $\mathcal{Z}_p$ where p is a prime number, and applying construction A [16]. This forms a good fine lattice in n-dimensional Euclidean space nested with a coarse *cubic* lattice with nesting ratio $\sqrt{[n]}\,p$. While this nesting in a coarse cubic lattice is just an artifact of any type A construction, we can utilize it to obtain a fine code nested in a good *coarse* lattice as well. Specifically, denoting the $k \times k$ generator matrix of Λ_2' as G, we transform the $n = k \times m$-dimensional Euclidean space by applying G to each of the m consecutive k-blocks. This transformation preserves the random code properties required in Loeliger's construction, which for the appropriate choice of G imply the goodness properties i) and ii) in Section III-B. As discussed in Section IV-C, this factorizable form also has some practical merits, but it requires modifications for small nesting ratios.

An explicit (and practical) construction of good nested codes in real space was introduced by Forney and Eyuboglu [51], [33]. Here, a trellis code plays the role of a finite complexity infinite-dimensional lattice. The preceding existence argument for good nested lattices can be extended to such trellis-based nested codes. In fact, in the applications discussed later it may be practically advantageous to replace the nested lattice codes with nested trellis codes.

IV. Noisy Side Information Problems

Relative cosets of good nested codes generate efficient binning schemes for "noisy" network coding problems. To demonstrate that, we first consider the simpler settings of coding with side information. These settings are in a sense noisy extensions of the two basic settings of Section II, Figs. 2 and 3, and are based on [72], [97], [1]. In the sequel, we switch back and forth between the binary case and the continuous case, and for convenience we use the same letters $x,\, y,\, z$ to denote source/channel variables in both cases.

A. The Wyner–Ziv Problem

Consider the lossy extension of the configuration in Fig. 2 of source coding with side information. As in the lossless case, the encoding and decoding functions take the form

$$i = f(\boldsymbol{x}) \quad \text{and} \quad \hat{\boldsymbol{x}} = g(i,\, \boldsymbol{y}) \tag{4.1}$$

respectively. However, in the lossy case we allow some distortion between the source and the reconstruction

$$\frac{1}{n} Ed(\boldsymbol{X},\, \hat{\boldsymbol{X}}) \le D \tag{4.2}$$

for some distortion measure d. Wyner and Ziv [91] showed that if X and Y are doubly binary symmetric, where $Y = X \oplus Z$ with Z Bernoulli-p, and d is the Hamming distance, then the minimum coding rate is given by

$$R_{\mathrm{WZ}}(D) = \ell.c.e\left\{H(p * D) - H(D),\ (p,\, 0)\right\}, \qquad 0 \le D \le p \tag{4.3}$$

[2] Note that for the good channel code component, the "δ" indicates the AWGN power, which is in general smaller than, or equal to the second moment of the lattice. For the good source code component, the "δ" indicates the mean square distortion, which coincides with the second moment of the lattice.

[3] The dual case of complex-coarse/simple-fine nested lattices can be achieved using concatenated codes, and it will be discussed elsewhere.

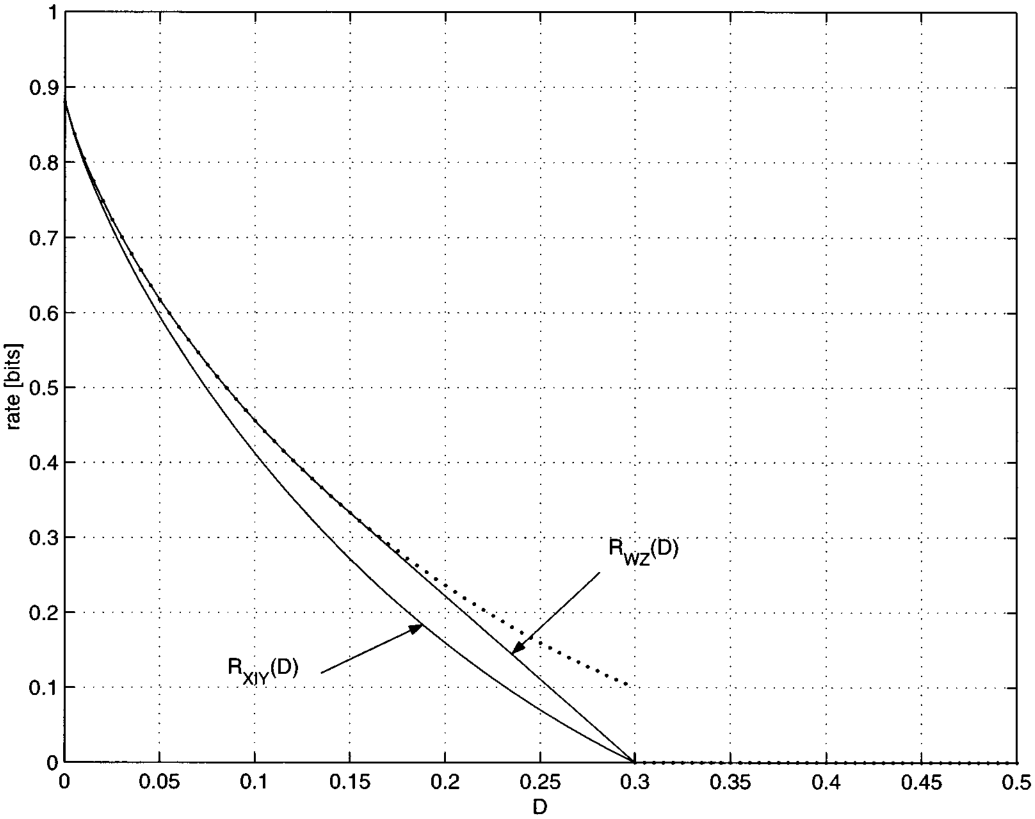

Fig. 8. $R_{\rm WZ}(D)$ for a doubly symmetric binary source.

where $p * D = p(1-D) + (1-p)D$ is the binary convolution of p and D, i.e., $R_{\rm WZ}(D)$ is the lower convex envelope of the function $H(p*D) - H(D)$ and the point $(D=p,\, R=0)$; see Fig. 8.

In the continuous case, Wyner [89] showed that if X and Y are jointly Gaussian, and the distortion measure the a squared error, then

$$R_{\rm WZ}(D) = \frac{1}{2}\log\left(\frac{\sigma^2_{x|y}}{D}\right), \qquad 0 \le D \le \sigma^2_{x|y} \tag{4.4}$$

where $\sigma^2_{x|y}$ is the conditional variance of X given Y.

Interestingly, the Wyner–Ziv rate-distortion function (4.3) in the binary case is strictly greater than the conditional rate-distortion function $R_{x|y}(D) = H(p) - H(D)$, which corresponds to the case where the side information is available to both the encoder and the decoder. On the other hand, in the quadratic-Gaussian case the two functions coincide, i.e., $R_{\rm WZ}(D) = R_{x|y}(D)$.

The standard proof of the achievability of the Wyner–Ziv function is by random binning; see, e.g., [21]. We now show how to achieve these rate-distortion functions using relative cosets of nested codes, following the constructions in [72], [97]. Our constructions generalize (4.3) and (4.4) in the sense that the side information $\boldsymbol{y}$ may be an arbitrary signal (not necessarily Bernoulli/Gaussian).

In the binary-Hamming case, we use a pair of nested parity-check codes with check matrices H_1 and H_2, where $H_2^t = [\Delta H^t;\, H_1^t]$ and $(\cdot)^t$ denotes transpose, as defined in (3.2). We require the fine code $\mathcal{C}_1$ to be a *good source D-code*, and the coarse code $\mathcal{C}_2$ to be a *good channel $p * D$-code*.

Encoding: quantize $\boldsymbol{x}$ to the nearest point in $\mathcal{C}_1$, resulting in $\boldsymbol{x}_q = \boldsymbol{x} \oplus f(H\boldsymbol{x}) \in \mathcal{C}_1$; then transmit $\Delta\boldsymbol{s} = \Delta H \cdot \boldsymbol{x}_q$, which requires $k_1 - k_2 \approx n \cdot [H(p*D) - H(D)]$ bits (see (3.4)).

Decoding: compute $\boldsymbol{s}_2 = H_2\boldsymbol{x}_q$ by zero padding, i.e., $\boldsymbol{s}_2 = (\boldsymbol{0},\, \boldsymbol{\Delta s})$; then reconstruct $\boldsymbol{x}$ by the point in the coset $\mathcal{C}_{2,\boldsymbol{s}_2}$ which is closest to $\boldsymbol{y}$, an operation that can be written as (see (2.11) and (2.12))

$$\hat{\boldsymbol{x}} = \boldsymbol{y} \oplus \hat{\boldsymbol{w}}, \quad \text{where } \hat{\boldsymbol{w}} = f_2(\boldsymbol{s}_2 \oplus H_2\boldsymbol{y}). \tag{4.5}$$

Time sharing this procedure with the "idle point" $(D = p,\, R = 0)$ gives the $R_{\rm WZ}$ function (4.3). It is left to be shown that the reconstruction $\hat{\boldsymbol{x}}$ is equal with high probability to $\boldsymbol{x}_q$, and, therefore, by the definition of the fine code, $\frac{1}{n}\, Ed_H(\boldsymbol{X}, \hat{\boldsymbol{X}})$ satisfies the distortion constraint D. To that end, consider Fig. 9, which shows an equivalent schematic formulation of this coding–decoding procedure in terms of $\bmod$-$\mathcal{C}_2$ operations based on the identity (2.15). Note that the concatenation of ΔH, zero padding, and $f_2(\cdot)$ in the signal path can be replaced by a single $\bmod$-$\mathcal{C}_2$ operation, whose output is $\boldsymbol{v}_2 = \boldsymbol{x}_q \bmod \mathcal{C}_2$.[4] Since we have two successive $\bmod$-$\mathcal{C}_2$ operations at the signal path, we use the distributive property (2.18) to eliminate the first, and arrive at the equivalent channel shown in Fig. 10, with $\boldsymbol{e}_q = \boldsymbol{x} \oplus \boldsymbol{x}_q$ denoting the quantization error. It follows that

$$\hat{\boldsymbol{x}} = ((\boldsymbol{e}_q \oplus \boldsymbol{z}) \bmod \mathcal{C}_2) \oplus \boldsymbol{y} \tag{4.6}$$

$$\overset{\text{c.d.}}{=} (\boldsymbol{e}_q \oplus \boldsymbol{z}) \oplus \boldsymbol{y} \tag{4.7}$$

$$= \boldsymbol{x}_q \tag{4.8}$$

[4] Using this formulation, the vector $\hat{\boldsymbol{w}}$ in (4.5) is given by $\hat{\boldsymbol{w}} = (\boldsymbol{v}_2 \oplus \boldsymbol{y}) \bmod \mathcal{C}_2$

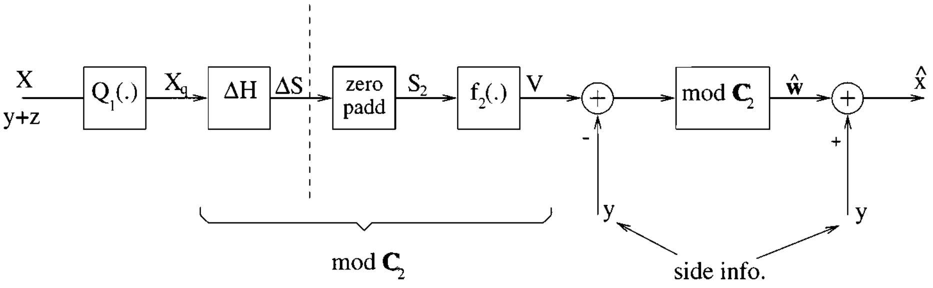

Fig. 9. Wyner–Ziv encoding of a doubly symmetric binary source using nested linear codes.

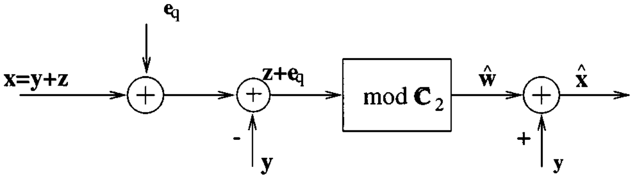

Fig. 10. Equivalent channel for the coding scheme of Fig. 9.

where $\overset{\text{c.d.}}{=}$ denotes *equality conditional on correct decoding*, and in the last line we used $\boldsymbol{y} \oplus \boldsymbol{z} = \boldsymbol{x}$. Thus, correct decoding amounts to $(\boldsymbol{e}_q \oplus \boldsymbol{z}) \bmod \mathcal{C}_2 = \boldsymbol{e}_q \oplus \boldsymbol{z}$, which implies that $\hat{\boldsymbol{x}} = \boldsymbol{x}_q$, as desired. The decoding error probability is equal to

$$P_e = \Pr\{(\boldsymbol{E}_q \oplus \boldsymbol{Z}) \bmod \mathcal{C}_2 \neq \boldsymbol{E}_q \oplus \boldsymbol{Z}\}. \tag{4.9}$$

Note that $\boldsymbol{E}_q$ and $\boldsymbol{Z}$ are statistically independent, $\boldsymbol{Z}$ is Bernoulli-p, and

$$\frac{1}{n}\, E w_H(\boldsymbol{E}_q) = \frac{1}{n}\, E d_H(\boldsymbol{X}, \boldsymbol{X}_q) = D$$

thus,

$$\frac{1}{n}\, E w_H(\boldsymbol{E}_q \oplus \boldsymbol{Z}) = D * p. \tag{4.10}$$

Hence, if the quantization error $\boldsymbol{E}_q$ were a Bernoulli process, then $\boldsymbol{E}_q \oplus \boldsymbol{Z}$ were a Bernoulli $-(D * p)$ process, and the $(D * p)$-goodness of the coarse code $\mathcal{C}_2$ would have implied that $P_e < \varepsilon$. However, $\boldsymbol{E}_q$ is not a Bernoulli process. In fact, $\boldsymbol{E}_q$ is distributed uniformly over $\Omega_{0,1}$, the basic Voronoi cell of $\mathcal{C}_1$ (see (2.7)).[5]

We see that the quantization error generated by the fine code plays the role of a noise component for the coarse code. We call this phenomena "self noise," and as we shall see, it appears in almost all applications of nested codes to binning schemes. Can the coarse code, the good channel code component of the nested code, protect against errors induced by the self-noise? We will address this question momentarily, in the context of nested lattice codes.

To achieve the Wyner–Ziv function in the quadratic Gaussian case, we assume that X and Y are related as

$$X = Y + Z \tag{4.11}$$

where Z is an independent zero mean Gaussian with variance σ_z^2, i.e., $\sigma_{x|y}^2 = \text{Var}(X|Y) = \sigma_z^2$.[6] The random variable Y may be arbitrary (not necessarily Gaussian). Our nested code construction discussed next is an improved version of the basic construction of [97] (which was optimal only for $\sigma_x^2 \gg \sigma_z^2$), and of [1] (which extended [97] to any ratio of σ_x^2 to σ_z^2, but did not take into account the exact effects of the self-noise).

Use a nested lattice pair (Λ_1, Λ_2) whose generator matrices are related by $G_2 = G_1 \cdot J$, as discussed in Section III-B. Require the fine lattice Λ_1 to be a good source D-code, and the coarse lattice Λ_2 to be a good channel σ_z^2-code. Let the (pseudo) random vector $\boldsymbol{U}$, the "dither," be uniformly distributed over $\mathcal{V}_{0,1}$, the basic Voronoi cell of the fine lattice. We shall assume that the encoder and the decoder share *common randomness*, so that $\boldsymbol{U}$ is available to both of them. Let $\alpha = \sqrt{1 - D/\sigma_z^2}$ denote the optimum estimation coefficient to be used in the following.

Encoding: quantize $\alpha\boldsymbol{x} + \boldsymbol{u}$ to the nearest point in Λ_1, resulting in $\boldsymbol{x}_q = Q_1(\alpha\boldsymbol{x} + \boldsymbol{u})$, then transmit an index which identifies $\boldsymbol{v}_2 = \boldsymbol{x}_q \bmod \Lambda_2$, the leader of the unique relative coset containing $\boldsymbol{x}_q$; by (3.18), this index requires $\log(V_2/V_1) \approx \frac{n}{2}\log(\sigma_z^2/D)$ bits.

Decoding: decode the coset leader $\boldsymbol{v}_2$, and reconstruct $\boldsymbol{x}$ as

$$\hat{\boldsymbol{x}} = \boldsymbol{y} + \alpha\hat{\boldsymbol{w}}, \quad \text{where } \hat{\boldsymbol{w}} = [\boldsymbol{v}_2 - \boldsymbol{u} - \alpha\boldsymbol{y}] \bmod \Lambda_2. \tag{4.12}$$

This procedure is unique up to scaling. For example, we can equivalently inflate (Λ_1, Λ_2) by a factor $1/\alpha$, quantize $\boldsymbol{x}$ directly (instead of $\alpha\boldsymbol{x}$), and multiply the output of the second $\text{mod-}\Lambda_2$ operation by α^2 (instead of α).

Note that the coding rate coincides with (4.4) as desired. To complete the analysis of the scheme, we show that the expected mean squared reconstruction error $\frac{1}{n} E\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|^2 \approx D$.

To that end, consider Fig. 11, which shows a schematic formulation of this coding–decoding procedure. Note that in the figure we suppressed the intermediate mapping of $\boldsymbol{v}_2$ into the transmitted index. As in the binary case (2.18), the $\text{mod-}\Lambda_2$ operation satisfies a distributive property

$$((\boldsymbol{x} \bmod \Lambda) + \boldsymbol{y}) \bmod \Lambda = (\boldsymbol{x} + \boldsymbol{y}) \bmod \Lambda, \qquad \forall \boldsymbol{x}, \boldsymbol{y} \tag{4.13}$$

(which easily follows by $Q(x + Q(x')) = Q(x) + Q(x')$). This property implies that we can eliminate the first $\text{mod-}\Lambda_2$ operation in the signal path, and arrive at the equivalent channel

[5] Even for a nonuniform source the encoding scheme can force $\boldsymbol{E}_q$ to be uniform over $\Omega_{0,1}$ using subtractive dithering based on common randomness; see the Gaussian case later.

[6] Note that any jointly Gaussian pair (X, Y) can be described in the form (4.11), replacing Y with aY.

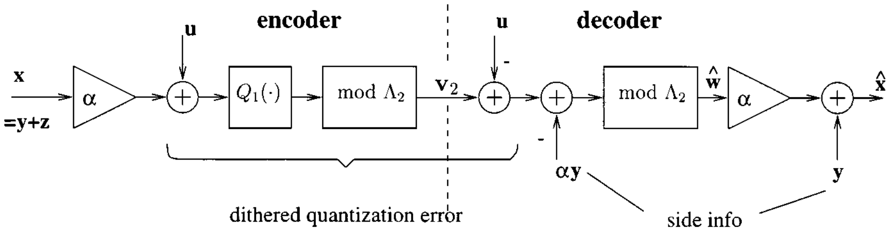

Fig. 11. Wyner–Ziv encoding of a jointly Gaussian source using nested lattice codes.

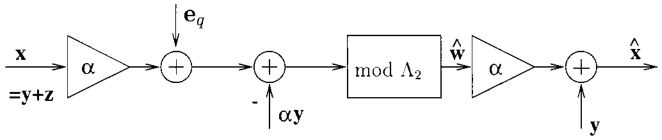

Fig. 12. Equivalent channel for the scheme of Fig. 11.

of Fig. 12, where $\boldsymbol{e}_q$ denotes the subtractive dither quantization error [95]

$$\boldsymbol{e}_q = Q_1(\alpha \boldsymbol{x} + \boldsymbol{u}) - (\alpha \boldsymbol{x} + \boldsymbol{u}).$$

Observing that the input to the mod-Λ_2 operation in Fig. 12 is $(\alpha \boldsymbol{z} + \boldsymbol{e}_q)$, we write the final reconstruction as

$$\hat{\boldsymbol{x}} = \alpha \left((\alpha \boldsymbol{z} + \boldsymbol{e}_q) \mod \Lambda_2 \right) + \boldsymbol{y} \tag{4.14}$$

$$\overset{\text{c.d.}}{=} \alpha \left(\alpha \boldsymbol{z} + \boldsymbol{e}_q \right) + \boldsymbol{y} \tag{4.15}$$

$$= \boldsymbol{x} + \left(\alpha \boldsymbol{e}_q - (1-\alpha^2) \boldsymbol{z} \right) \tag{4.16}$$

where, as earlier, $\overset{\text{c.d.}}{=}$ denotes *equality conditional on correct decoding*, and in the last line we used $\boldsymbol{y} + \boldsymbol{z} = \boldsymbol{x}$. We conclude that conditional on correct decoding, the equivalent error vector is

$$\hat{\boldsymbol{x}} - \boldsymbol{x} = \alpha \boldsymbol{e}_q - (1-\alpha^2) \boldsymbol{z} \tag{4.17}$$

while the decoding error probability is given by

$$P_{e,n} = \Pr \left\{ (\alpha \boldsymbol{Z} + \boldsymbol{E}_q) \mod \Lambda_2 \neq \alpha \boldsymbol{Z} + \boldsymbol{E}_q \right\}. \tag{4.18}$$

As we shall show later, for a sequence of good nested codes the probability of decoding error vanishes asymptotically, i.e.,

$$P_{e,n} \rightarrow 0, \qquad \text{as } n \rightarrow \infty. \tag{4.19}$$

Hence, the reconstruction error $\hat{\boldsymbol{X}} - \boldsymbol{X}$ converges in probability to the right-hand side of (4.17). Now, the second moment per dimension of the right-hand side of (4.17) is given by

$$\frac{1}{n} E \| \alpha \boldsymbol{E}_q - (1-\alpha^2) \boldsymbol{Z} \|^2$$

$$= \frac{1}{n} E \| \alpha (-\boldsymbol{U}) \|^2 + \frac{1}{n} E \| (1-\alpha^2) \boldsymbol{Z} \|^2 \tag{4.20}$$

$$= \alpha^2 D + (1-\alpha^2)^2 \sigma_z^2 \tag{4.21}$$

$$= D \tag{4.22}$$

for any $\boldsymbol{y}$, where in (4.20) we used a property of subtractive dithered quantization, [96], [95]; namely, that $\boldsymbol{E}_q$ is independent of $\boldsymbol{X}$ (and therefore of $\boldsymbol{Z}$), and is equal in distribution to $-\boldsymbol{U}$; and in (4.22) we substituted $\alpha^2 = 1 - D/\sigma_z^2$. On the other hand, in view of (4.14) and since the mod-Λ_2 operation only reduces the magnitude, $\hat{\boldsymbol{X}} - \boldsymbol{X}$ has a finite second moment as well. Thus, both sides of (4.17) have a finite second moment per dimension, implying that their convergence in probability (implied by (4.19)) implies convergence of their second moments, and we conclude that the reconstruction error is indeed arbitrarily close to D, provided that (4.19) holds, i.e., that the decoding error indeed vanishes.

Good Nested Codes and the Self-Noise Phenomenon: Proof of (4.19)

To show (4.19), consider the definition of the error event in (4.18). Note that the argument of the mod-Λ_2 operation satisfies

$$\frac{1}{n} E \| \alpha \boldsymbol{Z} + \boldsymbol{E}_q \|^2 = \frac{1}{n} E \| \alpha \boldsymbol{Z} \|^2 + \frac{1}{n} E \| -\boldsymbol{U} \|^2 = \alpha^2 \sigma_z^2 + D = \sigma_z^2$$

where we used the properties of subtractive dithered quantization as in (4.20); see [96], [95]. Thus, if $\alpha \boldsymbol{Z} + \boldsymbol{E}_q$ were AWGN, then the σ_z^2-channel-goodness of the coarse code would imply that $P_{e,n} \rightarrow 0$ as $n \rightarrow \infty$ and (4.19) is proved. But the quantization error $\boldsymbol{E}_q$ is not AWGN and, therefore, $\alpha \boldsymbol{Z} + \boldsymbol{E}_q$ is not either. Thus, we again encounter the "self-noise" phenomenon, where part of the noise seen by the channel code component of the nested lattice pair is induced by the quantization error of the source code component.

The "self-noise" phenomenon was observed in [72], [97], where it was conjectured that asymptotically its effect is similar to a Bernoulli process in the binary case, and to AWGN in the continuous case. This is, indeed, plausible by the source-coding goodness of the fine code. Other works which dealt with nested-like constructions adopted this argument to justify their derivations [1] or tended to disregard this phenomenon. However, there was no rigorous treatment until recently. Now, if the fine and coarse code components were independent, then the effect of the self-noise could have been made identical to a Bernoulli/AWGN process by appropriate randomization of the coarse code, e.g., interleaving in the binary case. However, we cannot randomize one code component while keeping the other component fixed, because the nesting relation connects the two components. In a recent work [30], Erez and Zamir confirm the conjecture made in [72], [97] by putting an additional condition on the nested code. This condition extends the meaning of a "good channel code," as defined for the lattice case in item i) of Section III-B:

i)* *Exponentially good channel codes over AWGN channel:* For any n and $\varepsilon > 0$, there exists an n-dimensional lattice with cell volume $< 2^{n[h(Z)+\varepsilon]}$, where $h(Z) = \frac{1}{2} \log 2\pi e \sigma_z^2$ and σ_z^2 are the entropy and the variance of the AWGN $\boldsymbol{Z}$, respectively, such that

$$P_e = \Pr\{ \boldsymbol{Z} \notin \mathcal{V}_0 \} < e^{-n \cdot E(\varepsilon)} \tag{4.23}$$

where $E(\varepsilon) > 0$.

As discussed in Section III-B, for good codes, the Voronoi region $\mathcal{V}_0$ tends to a Euclidean ball, implying that the quantization error $\boldsymbol{E}_q$ is roughly uniform over a Euclidean ball of radius $\sqrt{nD}$. The analysis of [30] shows that the effect of such a noise on the decoding error probability is *sub*exponential in n relative to an AWGN with the same power. Thus, if the coarse lattice Λ_2 is exponentially good, the effect of the self-noise on the decoding error probability (4.18) is asymptotically equivalent to AWGN. A similar analysis shows this fact with respect to the binary case, and it can be found in [32].

Note that the existence of a sequence of lattice channel codes with exponential decay of the decoding error probability (as in (4.23)) was shown in [62]. Yet, the existence of a good nested *pair* of lattices, as required by the Wyner–Ziv encoding schemes above, is a more delicate question, which we address in another work [31].

In the next section, we shall see that the "self-noise" phenomenon occurs in a "reversed" manner also in the dual problem of channel coding with side information. We note also that Eggers, Su, and Girod observed this phenomenon in their analysis of scalar digital watermarking schemes ("scalar Costa") at high host-to-watermark ratio [25].

B. The Costa Problem: "Writing on Dirty Paper"

The second setting we consider is a special case of channel coding with side information at the transmitter, known as "writing on dirty paper."

Consider the channel

$$Y = X + S + N \tag{4.24}$$

where X and Y are the channel input and output, respectively, N is an unknown additive noise, and S is an interference signal known to the transmitter but not to the receiver.[7] As in the configuration of Fig. 3 (and unlike Shannon's causal formulation [67], [28]), the encoder knows the *entire* interference vector $\boldsymbol{S} = (S_1, \cdots, S_n)$ prior to transmission. Hence, the encoding and decoding functions have the form

$$\boldsymbol{x} = f(w, \boldsymbol{s}) \quad \text{and} \quad \hat{w} = g(\boldsymbol{y}) \tag{4.25}$$

respectively, where w denotes the message. This setting extends the configuration of channel coding with perfect side information discussed in Section II (Fig. 3), in the sense that here there is an additional noise component (N) which is unknown to both the transmitter and the receiver.

Costa [18] adhering to the Gelfand–Pinsker setting [41], showed that if S and N are statistically independent Gaussian variables, and the channel input must satisfy an average power constraint $\frac{1}{n}E\|\boldsymbol{X}\|^2 \leq P$, then the capacity with side information at the transmitter is given by

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_N^2}\right) \tag{4.26}$$

where σ_N^2 is the variance of N. Thus, the effect of the interference S is canceled out completely, as if it were zero or it were available also at the receiver. The proof is based on the random binning solution for general channels with side information [41].

In the binary mod-2 additivie-noise channel case, Barron *et al.* [1] showed that if the known interference S is a binary-symmetric source, the unknown noise N is an independent Bernoulli-p source, and the channel input satisfies an input Hamming constraint $\frac{1}{n}Ew_H(\boldsymbol{X}) \leq \delta$, then the capacity with side information at the transmitter is given by

$$C(\delta) = \text{u.c.e.}\{H(\delta) - H(p),\, (0,\, 0)\}, \qquad 0 \leq \delta \leq 0.5 \tag{4.27}$$

where u.c.e.$\{\cdot\}$ denotes upper convex envelope as a function of δ. Thus, in this case we loose in capacity for not knowing S at the receiver, because with an informed receiver we could achieve capacity of $C = H(\delta * p) - H(p)$, which is larger for any $0 < \delta < 1/2$.

We now demonstrate how to achieve these side-information capacities using algebraic binning schemes, based on the relative cosets of a good nested linear/lattice code pair. In fact, our scheme generalizes (4.26) and (4.27) in the sense that the interference $\boldsymbol{S}$ may be an arbitrary signal (not necessarily Gaussian/Bernoulli). The following configurations are based on [1], yet with a more exact analysis of the effects of the self-noise.

In both cases, we tune the fine code to the (effective) unknown noise level, and the coarse code to the input constraint. Specifically, in the continuous case, we choose for the fine lattice Λ_1 an exponentially good channel $\frac{\sigma_N^2 P}{\sigma_N^2 + P}$-code (as defined in (4.23)). The coarse lattice Λ_2 should be a good source P-code. In the binary case, we choose for the fine code $\mathcal{C}_1$ an exponentially good channel p-code, and for the coarse code $\mathcal{C}_2$ a good source δ-code. Let the (pseudo) random vector $\boldsymbol{U}$ be uniform over the basic Voronoi cell of the coarse code, i.e., $\mathcal{V}_{0,\,2}$ (lattice case) or $\Omega_{0,\,2}$ (binary case). As in our scheme for the Wyner–Ziv problem, $\boldsymbol{U}$ is a dither signal, known to both the transmitter and the receiver via common randomness. For the continuous case define also the estimation coefficient $\alpha = P/(P + \sigma_N^2)$, and use the following encoding procedure.

Message selection: identify each coset $\Lambda_{2,\,v}$, $\boldsymbol{v} \in \{\Lambda_1 \cap \mathcal{V}_{0,\,2}\}$, with a unique message; by (3.18) this amounts to $\log(V_2/V_1) \approx \frac{n}{2}\log(1 + P/\sigma_N^2)$ bits per n-block.

Encoding: transmit the error vector between $\alpha\boldsymbol{s} + \boldsymbol{u}$ and the selected coset $\Lambda_{2,\,v}$, i.e.,

$$\boldsymbol{x} = [\boldsymbol{v} - \alpha\boldsymbol{s} - \boldsymbol{u}] \mod \Lambda_2 \tag{4.28}$$

where $\boldsymbol{s}$ is the interference vector and $\boldsymbol{u}$ is the dither. By the properties of dithered quantization

$$\frac{1}{n}E\{\|\boldsymbol{X}\|^2 \,|\, \boldsymbol{V} = \boldsymbol{v},\, \boldsymbol{S} = \boldsymbol{s}\} = P$$

independently of the values of $\boldsymbol{v}$ and $\boldsymbol{s}$, where the expectation is over the dither $\boldsymbol{U}$.

Decoding: reconstruct the message as the unique coset containing $Q_1(\alpha\boldsymbol{y} + \boldsymbol{u})$; the leader of this coset can be computed as

$$\hat{\boldsymbol{v}} = Q_1(\alpha\boldsymbol{y} + \boldsymbol{u}) \mod \Lambda_2. \tag{4.29}$$

[7] Note that any channel $Y = X + Z$ with side information S, where Z and S are jointly Gaussian, can be reduced to the form (4.24).

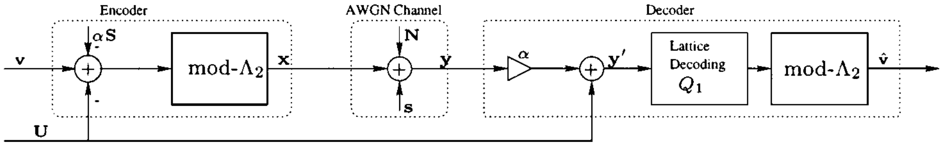

Fig. 13. Nested encoding and decoding scheme for the Costa problem.

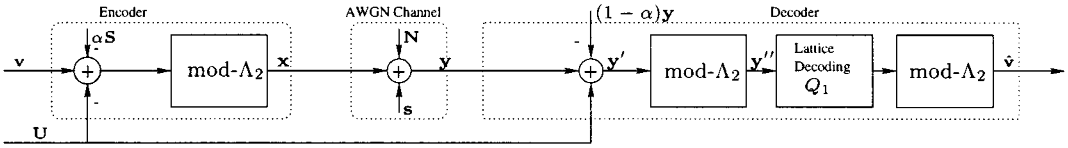

Fig. 14. Equivalent representation of encoding–decoding for the Costa problem.

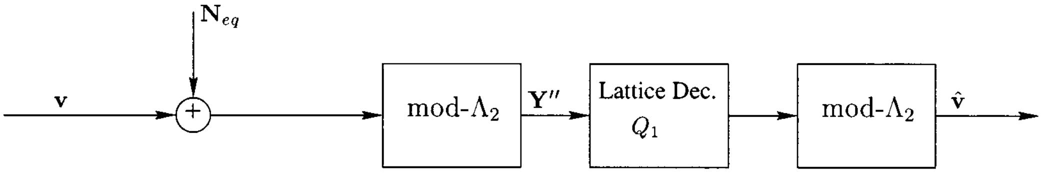

Fig. 15. Equivalent channel for the Costa problem.

Fig. 13 depicts this procedure. In the binary case, the procedure is the same, setting $\alpha = 1$, replacing Λ_2 by $\mathcal{C}_2$, and regarding the "lattice decoding" $Q_1(\cdot)$ as minimum Hamming distance decoding with respect to the fine code $\mathcal{C}_1$. This results in

$$\log(|\Omega_{02}|/|\Omega_{01}|) \approx n[H(\delta) - H(p)]$$

bits of information, and an average codeword Hamming weight of δ. Time-sharing this procedure with the "idle point" ($\delta = 0, R = 0$) gives the $C(\delta)$ function (4.27).

We want to show that $\hat{\boldsymbol{V}} = \boldsymbol{V}$ with high probability, so that the message is decoded correctly. In the binary case, the distributive property of the mod-$\mathcal{C}_2$ operation allows to eliminate the first mod-$\mathcal{C}_2$ operation, and obtain that the decoded coset leader is given by

$$\hat{\boldsymbol{v}} = Q_1(\boldsymbol{v} + \boldsymbol{N}) \bmod \mathcal{C}_2 \tag{4.30}$$

$$\overset{\text{c.d.}}{=} \boldsymbol{v} \tag{4.31}$$

where the second equality holds conditional on correct decoding. By the p-goodness of the fine code and since N is Bernoulli-p, the second line indeed holds with high probability.

The proof that the decoding error probability

$$P_e = \Pr(\hat{\boldsymbol{V}} \neq \boldsymbol{V}) \tag{4.32}$$

is small also in the continuous case is slightly more involved due to the presence of the estimation coefficient $\alpha < 1$. Consider first Fig. 14, where we replaced the gain α in the signal path by a shortcut and compensated for that by subtracting $(1-\alpha)\boldsymbol{y}$ from $\boldsymbol{y}'$; we also inserted another mod-Λ_2 operation which by the distributive property does not affect the final result, and denoted its output $\boldsymbol{y}''$, i.e., $\boldsymbol{y}'' = [\alpha\boldsymbol{y} + \boldsymbol{u}] \bmod \Lambda_2$. We then used the distributive property to eliminate the first mod-Λ_2 operation, and arrived at the equivalent channel of Fig. 15. It follows that

$$\begin{aligned}\boldsymbol{y}'' =& [(\boldsymbol{v} - \alpha\boldsymbol{s} - \boldsymbol{u}) + (\boldsymbol{s} + \boldsymbol{n}) \\ & - (1-\alpha)(\boldsymbol{x} + \boldsymbol{s} + \boldsymbol{n}) + \boldsymbol{u}] \bmod \Lambda_2 \end{aligned} \tag{4.33}$$

$$= [\boldsymbol{v} + \alpha\boldsymbol{n} - (1-\alpha)\boldsymbol{x}] \bmod \Lambda_2. \tag{4.34}$$

Thus, the equivalent noise component is $\boldsymbol{n}_{\text{eq}} = \alpha\boldsymbol{n} - (1-\alpha)\boldsymbol{x}$. Quite surprisingly, an interesting lemma in [28], [27] shows that this noise is independent of the input so the equivalent channel is modulo-additive.

Lemma 1 "Inflated Lattice" [28], [27]: The channel from $\boldsymbol{v}$ to $\boldsymbol{Y}''$, for $\boldsymbol{U}$ uniformly distributed over $\mathcal{V}_{0,2}$, is equivalent in distribution to the channel

$$\boldsymbol{Y}'' = [\boldsymbol{v} + \boldsymbol{N}_{\text{eq}}] \bmod \Lambda_2, \qquad \text{where } \boldsymbol{N}_{\text{eq}} = [(1-\alpha)\boldsymbol{U} + \alpha\boldsymbol{N}] \tag{4.35}$$

and where $\boldsymbol{N}_{\text{eq}}$ is independent of $\boldsymbol{v}$.

Substituting $\alpha = P/(P + \sigma_N^2)$, the second moment per dimension of $\boldsymbol{N}_{\text{eq}}$ is equal to

$$(1-\alpha)^2 P + \alpha^2\sigma_N^2 = P\sigma_N^2/(P + \sigma_N^2).$$

Thus, if $\boldsymbol{N}_{\text{eq}}$ were AWGN, then the $P\sigma_N^2/(P + \sigma_N^2)$-channel-goodness of the fine lattice would imply that $Q_1(\boldsymbol{Y}'') \bmod \Lambda_2$ is equal to $\boldsymbol{v}$ with high probability, and, therefore, P_e is small as desired. However, the "self-noise" component $(1-\alpha)\boldsymbol{U}$ is not Gaussian, but rather uniform over $\mathcal{V}_{02}$. Yet, as in Section IV-A, if the fine lattice is exponentially good, then the probability of decoding error goes to zero in spite of the slight deviation of $\boldsymbol{N}_{\text{eq}}$ from AWGN. See [30] for a detailed proof. Thus, with good nested codes, the proposed scheme approaches capacity, as desired.

As discussed in Section VII, if we choose the coarse lattice to be cubic, the preceding scheme becomes a dithered version of Tomlinson precoding [80], where S plays the role of "intersymbol interference."

C. A Note on Coding Complexity

In practice, the "goodness" requirement from the two components of the nested code is not equally stringent. In the quadratic Wyner–Ziv problem of Section IV-A (source coding with side information) the channel coding goodness of the coarse code determines the decoding error probability (4.19). Hence, the coarse lattice must be sufficiently complex to make this probability small. On the other hand, the source coding goodness of the fine code has only a slight effect on the rate, and therefore in practice the fine lattice can be simple. For example, if the fine code is simply a cubic lattice (corresponding to scalar quantization), then $G_n = G_1 = 1/12$, instead of the optimum value $G_\infty = 1/2\pi e$; see (3.10). This implies cell volume $V_1 = \frac{n}{2}\log(12D)$, and rate redundancy of $\frac{1}{2}\log(2\pi e/12) \approx 0.254$ bit per sample above the Wyner–Ziv-rate distortion function. For a general lattice Λ, the rate redundancy becomes $\frac{1}{2}\log(2\pi e G(\Lambda))$.

In the dual Costa problem of Section IV-B (channel coding with side information) this behavior is reversed. The channel coding goodness of the *fine* code determines the decoding error probability P_e in (4.32); in the terminology of digital communication [52], the "coding gain" of the fine lattice should be high enough to achieve a desired P_e without too much excess power. The source coding goodness of the *coarse* code only slightly affects the rate; the coarse lattice cell volume is $\frac{n}{2}\log(P/G(\Lambda))$, so the rate gain relative to a cubic lattice is given by (see (3.10))

$$\frac{1}{2}\log(1/12G(\Lambda))$$

a term called "space-filling gain" in the quantization literature [43] and "shaping gain" in the digital communication literature [52], [100]. Hence, the asymptotic gain of a good coarse lattice with respect to a simple cubic coarse lattice is only $\frac{1}{2}\log(2\pi e/12) \approx 0.254$ bit. See discussion regarding constellation shaping in [51], [33].

The perfect source/channel decoupling of the nested code discussed above ceases to hold at low coding rates, i.e., at small nesting ratios. In this regime the "self-noise" becomes a significant portion of the equivalent channel noise. This implies that the decoding error probability is determined not only by the properties of the channel coding component of the nested code, but also by the source coding component.

D. Writing to a Memory With Noise and Defects

The third setting we consider is another special case of channel coding with side information at the encoder, proposed by Tsybakov [81] and Heegard and El-Gamal [47]. A binary memory is modeled as a channel with three states

$$S \in \{\text{stuck at zero, stuck at one, BSC}(\epsilon)\}$$

where the state process is memoryless, with probability $p/2$ of each of the "stuck-at" states, and probability $1-p$ of the "BSC" state. As shown in [47], if the state sequence is known in advance to the encoder, or to both the encoder and the decoder, then the capacity of this channel is given by

$$C = (1-p)(1-H(\epsilon)).$$

Note that the case $\epsilon = 0$ corresponds to the Kuznetsov–Tsybakov problem, [55], which can be solved using a good *erasure* correction coset code (correcting $\approx n(1-p)$ erasures per n-block), as discussed in Section II-D.

A nested coding approach for this problem was proposed by Tsybakov [81] and Heegard [45], where the codes are referred to as "partitioned" codes. Heegard showed that nested parity-check codes may achieve the capacity of the channel above as well as suggested specific code constructions for this problem based on BCH codes. We now give a heuristic description of a nested parity-check code solution in the terminology of this paper. The fine code component is a good BSC ϵ-code of dimension n, as defined in Section II-B. The coarse code component is good in a different sense than discussed so far. It should have the property that the projection of the code on almost every subset of np components is a good BSS ϵ-code of dimension np. Note that for $\epsilon = 0$, this property amounts to good *erasure* correction at coding rate slightly higher than p. Observe also that the requirement becomes less restrictive as ϵ increases; random coding arguments imply that there exist parity-check codes satisfying this property at coding rates slightly higher than $p(1-H(\epsilon))$. It follows that such codes have slightly less than $2^{n[1-p(1-H(\epsilon))]}$ cosets, each of which satisfies the desired property above.

Message selection: identify each relative coset with a unique message; the number of bits per n-block are thus

$$\log\left(\frac{|\Omega_{0,2}|}{|\Omega_{0,1}|}\right) \approx \log\left(\frac{2^{n[1-p(1-H(\epsilon))]}}{2^{nH(\epsilon)}}\right) = n(1-p)(1-H(\epsilon)).$$

Encoding: look for a vector in the message coset which is ϵ-compatible with the known defect, that is, it agrees with the np stuck-at values a fraction $1-\epsilon$ of these cells. (Such a vector exists with high probability by the BSS ϵ-goodness of every np projection of the coarse code.) Store this vector in the memory, or declare an error if it does not exist.

Decoding: do conventional BSC decoding of the fine code from the stored vector, and identify the coset ($=$ message) to which the decoded vector belongs. The decoding error is small by the BSC ϵ-goodness of the fine code, and because the stored vector has a total of $n\epsilon$ errors (intentional $np\epsilon$ at the stuck-at positions, and random $n(1-p)\epsilon$ errors at the other positions).

E. Nested Codes with Probabilistic Decoding

So far we assumed that the unknown noise components (Z in the Wyner–Ziv problem, and N in the Costa problem) correspond to memoryless binary-symmetric/Gaussian-noise channels. For such channels, maximum likelihood (ML) is equivalent to minimum Hamming/Euclidean distance, so it lends itself to *algebraic* decoding of the nested code. We may extend the schemes above to more general additive noise channels if we keep the nested coding structure, but allow *probabilistic* decoding. This amounts to replacing the Voronoi decision cells of the channel code component (the coarse code in the Wyner–Ziv problem or the fine code in the Costa problem) by ML decoding decision cells.

Information-wise, a good nested code is equivalent to a uniformly distributed random code over $\mathcal{V}_2$; see Section III-C. Such a code is not necessarily optimal for a general (non-Gaussian/correlated) noise and a general coding rate. Nevertheless, in the limit of high coding rate (i.e., large nesting ratio), a good nested code with probabilistic decoding becomes optimal for any additive noise. Specifically, it can be shown that asymptotically as $D \to 0$, the rate-distortion performance of the ML-decoded Wyner–Ziv nested coding scheme is

$$\approx h(X|Y) - \frac{1}{2}\log(2\pi e D)$$

provided that $h(X|Y) = h(Z)$ is finite. Also, the asymptotic rate–power performance of the ML-decoded Costa nested coding scheme becomes, as $P \to \infty$

$$\approx \frac{1}{2}\log(2\pi e P) - h(N)$$

provided that $h(N)$ is finite. Note that these expressions coincide asymptotically with the rate-distortion and the capacity functions, respectively [27], [105].

V. Multiterminal Coding of Correlated Sources

The nested coding schemes above, in the presence of side information, provide the basic blocks for more general network coding schemes. In the context of source coding, the main application we address is multiterminal (or distributed) source coding [3], [94]. This configuration generalizes the problem of source coding with side information at the decoder, discussed in Sections II-A and IV-A. Before considering this application, we briefly consider two related problems: multiresolution source coding, variants of which are known also as *successive refinement* or *multistage source coding* [26]; and multiple side-information terminals.

A. Two Related Problems

Multiresolution Source Coding: A multiresolution source code consists of two codewords, the first contains a coarse description of the source, while the second contains a refinement of this description. Nested codes provide a straightforward mechanism for multiresolution source coding, where the coarse and fine components, $\mathcal{C}_2$ and $\mathcal{C}_1$, generate the coarse and fine descriptions, respectively. By the structure of the nested code, the relative coset information, $\mathcal{C}_1 \bmod \mathcal{C}_2$, provides a natural construction for the refinement codeword. However, unlike in side information problems, both components of the nested multiresolution code should be "good *source* codes." More importantly, this code is *not* induced by a binning mechanism. Hence, this problem is, in fact, conceptually different than the problems discussed so far, and it will not be developed here further.

Multiple Side-Information Terminals: Consider the problem of Fig. 2, but assume that the decoder has access to two side-information variables instead of one, $Y^{(1)}$ and $Y^{(2)}$, which are correlated with the source X. From a *random* binning viewpoint, this configuration is basically the same as before, since we can view the two side information variables as a single variable with two components. However, an algebraic formulation for the solution is not straightforward. Specifically, following the solution in the single side-information terminal case, we would like the decoder to reconstruct the source $\boldsymbol{x}$ via an algebraic operation on some coset information and the side information, i.e., using

$$\boldsymbol{x} \in \mathcal{C}_s, \quad \boldsymbol{y}^{(1)} = \boldsymbol{x} + \boldsymbol{z}^{(1)}, \quad \text{and} \quad \boldsymbol{y}^{(2)} = \boldsymbol{x} + \boldsymbol{z}^{(2)} \tag{5.1}$$

where $\mathcal{C}_s$ is a suitable coset code. In the binary case, maximum-likelihood decoding of $\boldsymbol{x}$ from (5.1) amounts to *weighted* minimum Hamming distance decoding (because $p(x|y^{(1)}, y^{(2)})$ has four values, one per each of the four possible pairs $(y^{(1)}, y^{(2)})$). Thus, we must use "soft" decisions, rather than basic algebraic operations as desired. Nevertheless, in the joint Gaussian case, the pair $(\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)})$ in (5.1) can be reduced to a single *sufficient statistic*, $\alpha\boldsymbol{y}^{(1)} + \beta\boldsymbol{y}^{(2)}$, where α and β are suitable minimum mean square error (MMSE) estimation coefficients. The conditional distribution of $\boldsymbol{X}$ given this statistic is Gaussian, with variance equals to the conditional variance of X given $X + Z^{(1)}, X + Z^{(2)}$. Thus, the problem reduces to conventional lattice decoding in the presence of AWGN, as in the single side-information terminal case.

Combining the two cases (multiresolution and multiple side-information terminals), we can devise nested coding schemes for more general configurations, e.g., rate distortion when side information may be present/absent [46], [54].

B. The Theoretic Multiterminal Rate Region

The general form of the Slepian–Wolf problem allows encoding of both correlated sources X and Y, at rates R_1 and R_2, respectively [76]. The resulting configuration consists of two separate encoders and a common decoder, as shown in Fig. 16.

Lossless reconstruction of X and Y is possible if and only if

$$\begin{gathered} R_1 \geq H(X|Y), \quad R_2 \geq H(Y|X) \\ R_1 + R_2 \geq H(X, Y). \end{gathered} \tag{5.2}$$

A lossy version of the Slepian–Wolf problem was considered by several researchers, but a tight solution was found only in special cases [91], [3], [4], [61], [94]. In the quadratic-Gaussian case, the largest known explicit single-letter characterization of the set of achievable rates is given by a "long Markov chain solution" [3]

$$\begin{aligned} R_1 &\geq I(X; X + N_1|Y + N_2) \\ R_2 &\geq I(Y; Y + N_2|X + N_1) \\ R_1 + R_2 &\geq I(X, Y; X + N_1, Y + N_2) \end{aligned} \tag{5.3}$$

where (N_1, N_2) are any independent Gaussian variables, such that

$$\mathrm{Var}\,(X|X + N_1, Y + N_2) \leq D_1$$

and

$$\mathrm{Var}\,(Y|X + N_1, Y + N_2) \leq D_2$$

and where $\mathrm{Var}\,(A|B)$ denotes the conditional variance of A given B.

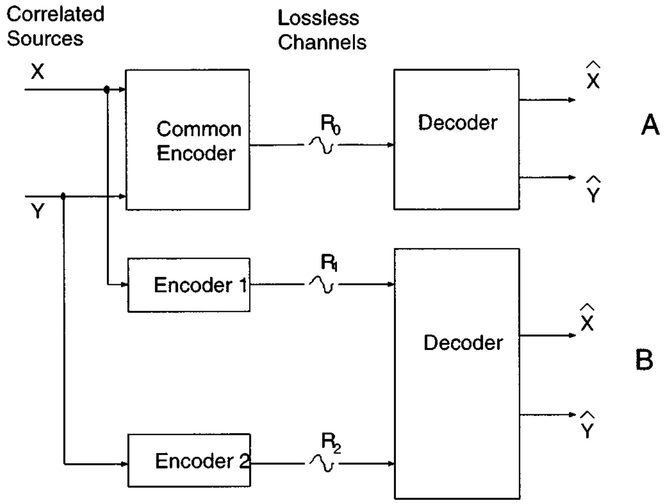

Fig. 16. General multiterminal source coding configuration.

As usual, the information-theoretic solutions above are based on random binning coding schemes and jointly typical decoding [22], [21]. Our goal is to realize these rate regions using algebraic coding–decoding schemes, and specifically, using nested codes as building blocks. Similar formulations were derived by Pradhan and Ramchandran [63].

C. Multiterminal Nested Codes

Lossless Case: We start with the lossless version, known as the Slepian–Wolf problem. As in Section II, we assume that X and Y form a doubly symmetric memoryless source, and are related as $X = Y + Z$, so $H(X) = H(Y) = 1$, and $H(X|Y) = H(Y|X) = H(Z) = H(p)$. Clearly, the case where $R_1 \geq H(X)$ and $R_2 \geq H(Y)$ is redundant, and trivially solved by single terminal codes. Hence, without loss of generality we can assume that $R_1 < H(X)$. If $R_2 \geq H(Y) = 1$, then we can compress Y losslessly using a single terminal code, and the problem reduces to that of encoding (X) with side information at the decoder (Y), as discussed in Section II-A. The interesting case is thus $H(X, Y) - R_1 \leq R_2 < H(Y)$. One way to solve this case is by time-sharing the "corner points" $(H(X|Y), H(Y))$ and $(H(X), H(Y|X))$ in the (R_1, R_2)-plain, each corresponds to a simple side information problem as discussed earlier.

An alternative way to solve the latter case is to use a *source-splitting* approach [66]: quantize the first source into $\boldsymbol{X}'$ such that $nR_2 = H(\boldsymbol{Y}|\boldsymbol{X}')$, and transmit $\boldsymbol{X}'$ across the first channel at rate R_1'; then, encode Y given side information $\boldsymbol{X}'$ at the decoder, and transmit the corresponding (coset) codeword across the second channel at rate R_2; finally, encode $\boldsymbol{X}$ losslessly given the side-information pair $(\boldsymbol{X}', \boldsymbol{Y})$, and transmit the codeword across the first channel at rate R_1'', as an addendum to the codeword of the first stage. If we generate $\boldsymbol{X}'$ in the first stage using a good source δ-code, the error $\boldsymbol{Z}' = \boldsymbol{X}' - \boldsymbol{X}$ is roughly a Bernoulli-δ process. Hence, we realize the first stage at rate $R_1' \approx 1 - H(\delta)$. Furthermore, since $X' = Y + Z + Z'$, we can realize the second stage using the coset information of a good channel $p * \delta$-code, at rate $R_2 \approx H(p * \delta)$, as discussed in Section II. (Note that similarly to our treatment of the "self-noise" problem, we can avoid the effect of the deviation of $\boldsymbol{Z}'$ from a Bernoulli process by a suitable choice of these codes.) The third stage requires, *in principle*, rate of

$$R_1'' = \frac{1}{n} H(\boldsymbol{X}|\boldsymbol{X}', \boldsymbol{Y}) \tag{5.4}$$

$$\approx H(X|X + Z', X + Z) \tag{5.5}$$

$$= H(p) + H(\delta) - H(p * \delta) \tag{5.6}$$

where (5.5) uses the approximate Bernoulli form of $\boldsymbol{Z}'$, while (5.6) follows using the chain rule and using the fact that X is a binary-symmetric source. It follows that the total rate $R_1 + R_2 = R_1' + R_2 + R_1''$ is approximately

$$(1 - H(\delta)) + (H(p * \delta)) + (H(p) + H(\delta) - H(p * \delta)) \\ = 1 + H(p) = H(X, Y)$$

and by varying δ in the range $0 \leq \delta \leq 1/2$ we obtain the entire boundary of the rate region (5.2). Unfortunately, however, achieving (5.4) is problematic in practice, as we need to encode a binary source (X) given *two* noisy binary versions (X', Y); as discussed in Section V-A, we cannot realize this encoding using pure algebraic operations, but must use a more complex "soft" decision decoding.

Lossy Case: The latter difficulty does not exist in the (lossy) quadratic-Gaussian case. Following the discussion of the lossless case, without loss of generality we assume that the rate of the X-terminal is smaller than the corresponding rate distortion function, i.e., $R_1 < R_X(D_1) = \frac{1}{2} \log \sigma_x^2 / D_1$. Consider first the case $R_2 > R_Y(D_2) = \frac{1}{2} \log \sigma_y^2 / D_2$. Quantize $\boldsymbol{Y}$ at rate R_2 into a codeword $\boldsymbol{Y}'$ with distortion $\text{Var}\,(\boldsymbol{Y}|\boldsymbol{Y}') = D_2' \leq D_2$. Assuming entropy-coded dithered quantization (ECDQ) with a "good" k-dimensional lattice [95], the quantization error $\boldsymbol{Z}' = \boldsymbol{Y}' - \boldsymbol{Y}$ is additive and white, and becomes AWGN, $\{Z^*\}$, as k goes to infinity, i.e.,

$$(\text{Var}\, Y | Y + Z^*) = D_2' \quad \text{and} \quad R_2 \to I(Y; Y + Z^*).$$

Then, use a nested lattice Wyner–Ziv code, to encode X with distortion D_1, given side information $Y' = X + Z + Z^*$ at the decoder. The rate required, $R_1 = R_{X|Y'}(D_1)$, can be written as $I(X; X + Z''|Y')$, where Z'' is an independent Gaussian random variable such that $\text{Var}\,(X|X + Z'', Y') = D_1$. Thus, the pair (R_1, R_2) lies on the boundary of the region (5.3) whenever $R_2 > R_Y(D_2)$.

Consider next the more interesting case, where both rates are smaller than the corresponding (marginal) rate distortion functions, i.e., $R_1 < R_X(D_1)$ and $R_2 < R_Y(D_2)$. Similarly to the lossless case, we can realize this case by *time-sharing* two points in the (R_1, R_2)-plain, having the form

$$(R_1 = R_X(D_1), R_2 = I(Y; Y + N_2 | X + N_1))$$

and

$$(R_1 = I(X; X + N_1 | X + N_2), R_2 = R_Y(D_2)).$$

Alternatively, we can use a *source-splitting* approach. First, quantize X at rate R_1' into a codeword $\boldsymbol{X}'$, such that R_2 is equal to the conditional rate-distortion function of Y given $\boldsymbol{X}'$

at distortion level D_2.[8] As above, we use ECDQ of sufficiently large dimension, so that $\boldsymbol{X}' = \boldsymbol{X} + \boldsymbol{Z}'$, where $\boldsymbol{Z}'$ is approximately AWGN. Then, use a nested lattice Wyner–Ziv code to encode Y at rate R_2 and distortion D_2 with side information $X + Z'$ at the decoder. Finally, use a nested lattice Wyner–Ziv code to encode X at rate R_1'' and distortion D_1, given the two side-information variables $X' = X + Z'$ and $Y = X + Z$ which are approximately jointly Gaussian by the properties of ECDQ. Unlike the binary case discussed earlier, we can convert these two side-information variables into a single sufficient statistic, and hence realize the latter stage using a "conventional" nested lattice code with algebraic encoding–decoding. See the discussion of multiple side-information terminals in Section V-A. It follows that by varying the quantization resolution at the first stage, the resulting rates $R_1 = R_1' + R_1''$ and R_2 realize the entire rate region (5.3) using nested lattice codes, as desired.

VI. Coordinated Encoding Over Mutually Interfering Channels

In this section, we apply the paradigm, observations, and results of the preceding sections to a variety of problems associated with interfering channels. In particular, we heavily rely on Section IV-B and demonstrate the power of the nested lattice coding approach in settings of single and multiple users and in multiple-element (vector) communication.

After presenting the general model in Section VI-A, in Section VI-B we address the Gaussian broadcast channel [19] and then broaden our view to discuss the framework of the broadcast approach for the block-fading channel scenario, for single [69] and multiple [68] users. Section VI-C will be devoted to the multiantenna broadcast channel recently treated in [8], [98], [99], [7]. In all these cases, a nested lattice approach is in fact a natural and appealing capacity-achieving strategy.

Two applications, though seemingly different from the above, are then shown to be fully equivalent as far as the nested lattice technique goes. Specifically, in Section VII-B, it will be concluded that the standard single-user dispersive Gaussian channel as well as the single-user multiantenna (MIMO) channel can also be treated within a similar framework.

For the sake of conciseness, we reduce the systems discussed in this section to the basic setting that has already been addressed in Section IV-B and invoke the relevant results and observations. We refer to the nested coding scheme for this setting as "dirty paper coding."

A. The General Model

The model on which we focus is the vector Gaussian channel described by

$$\boldsymbol{Y}_t = H_t \boldsymbol{X}_t + \boldsymbol{Z}_t \tag{6.1}$$

where $t = 1, 2, \ldots, n$ designates the discrete-time index and n stands for block length. Here $\boldsymbol{X}_t$ stands for the transmitted $(K \times 1)$ column vector. The Gaussian ambient noise $(M \times 1)$ column vector is denoted by $\boldsymbol{Z}_t$ and $\boldsymbol{Y}_t$ is the received $(M \times 1)$ vector. The $(M \times K)$ matrix H_t designates the matrix characterizing the MIMO channel. Unless otherwise stated, it is assumed that $H_t = H$ is a real deterministic known matrix and all signals are real valued.[9] When relevant, we treat H as a realization of a random matrix process. We also impose average power constraints on the input signals, and state the constraints according to the specific application discussed.

B. The Single-Input Gaussian Broadcast Channel

In this case, $K = 1$ and M designates the number of separate users, so that

$$Y_{t,i} = h_i X_t + Z_{t,i}, \qquad i = 1, 2, \ldots, M. \tag{6.2}$$

Let the entries of the $(M \times 1)$ vector

$$\boldsymbol{h}_t = \boldsymbol{h} = (h_1, h_2, \ldots, h_M)^T$$

(with upperscript T standing for the transpose operation) be ordered in increasing order $|h_\ell| \le |h_{\ell+1}|$. We also assume that the noise $\boldsymbol{Z}_t$ is independent and identically distributed (i.i.d.) (in t) and normalized to unit variance per component $E(Z_{t,i}^2) = 1$. This is a classical description of a Gaussian degraded broadcast channel [19], [21], the capacity region of which (assuming no common rate components), is given by the union of all rates simultaneously satisfying

$$R_i \le \frac{1}{2}\log\left(1 + \frac{h_i^2 P_i}{1 + h_i^2 \sum\limits_{j=i+1}^{M} P_j}\right), \qquad i = 1, \ldots, M \tag{6.3}$$

where the union is taken over all power assignments $\{P_i\}$, $i = 1, 2, \ldots, M$, satisfying an average power constraint

$$\sum_{i=1}^{M} P_i \le \text{SNR}. \tag{6.4}$$

Here, $E(X_t^2) \le \text{SNR}$ designates the input average power constraint, where E is the expectation operator. The classical approach [21] to achieve this region is by decomposing the transmitted signal X_t into a sum of independent components $X_{t;i}$, $i = 1, \ldots, M$, where

$$X_t = \sum_{i=1}^{M} X_{t,i} \tag{6.5}$$

and where $E(X_{t,i}^2) = P_i$. Now, $\{X_{t,i}\}_{t=1}^n$ carries the coded message information to user i, which is assumed to comprise the output of a good (capacity-achieving) Gaussian code of rate R_i. Decoding is accomplished via successive cancellation, that is, the decoder of user i can reliably decode[10] the messages of all preceding users $\{1, 2, \ldots, i-1\}$, as the channel is degraded, that is, $|h_1| \le |h_2| \le \cdots \le |h_M|$. The interference that stems from the already successfully decoded users is absolutely removed, while the interference of the users not yet decoded is accumulated and added to the ambient Gaussian noise.

We depart here from this classical approach by the way the decomposition (6.5) is interpreted and by the way the actual

[8] Note that in view of (5.3), R_2 is in general *greater* than the conditional rate-distortion function of Y given the *final* description of X; on the other hand, the conditional rate-distortion function decreases monotonically as the side information is refined. Thus, $\boldsymbol{X}'$ is *coarse* relative to the final description of X.

[9] Extensions to circularly symmetric complex signals is straightforward.

[10] It is tacitly assumed that each user is equipped with the codebooks assigned to all users that it can potentially decode.

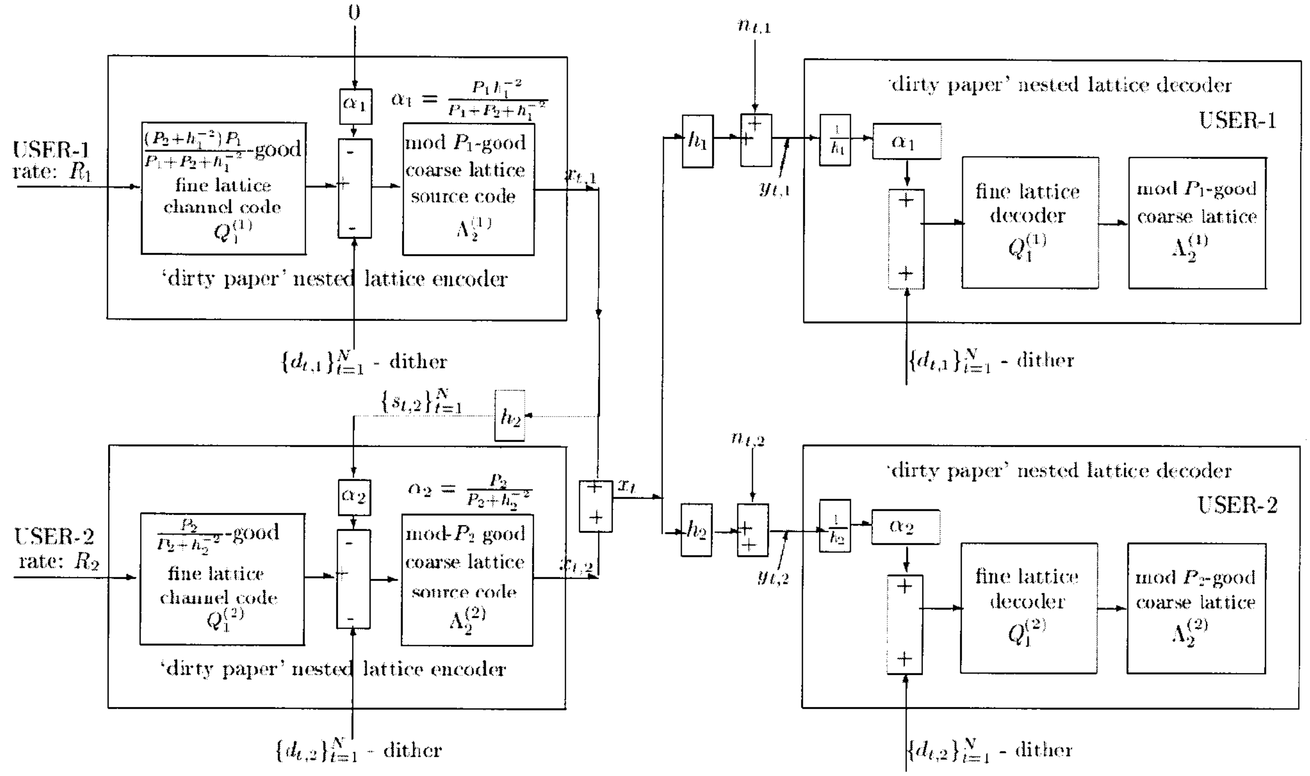

Fig. 17. Nested lattice encoding–decoding for the two-user broadcast channel.

coding–decoding strategy is constructed. The ith-user output at time t is

$$Y_{t,i} = h_i \left[X_{t,i} + \sum_{j=1}^{i-1} X_{t,j} + \sum_{j=i+1}^{M} X_{t,j} \right] + Z_{t,i}. \tag{6.6}$$

The part

$$S_{t,i} = h_i \sum_{j=1}^{i-1} X_{t,j} \tag{6.7}$$

is interpreted as an interference sequence known ahead of "time" (for all $t = 1, 2, \ldots, n$) at the transmitter. This is indeed the case at hand, as the transmitter controls the generation of all $\{X_{t,i}\}$, $i = 1, 2, \ldots, M$. The "dirty paper coding," which encompasses the nested lattice technique, as described in Section IV-B, guarantees the achievability of the rates $\{R_i\}$ as in (6.3). Note also that here the other interference term seen by user i

$$W_{t,i} = h_i \sum_{j=i+1}^{M} X_{t,j} \tag{6.8}$$

functions as an additional noise component, approaching Gaussianity as the lattice dimension grows. Caution should be exercised here as $\{X_{t,j}\}$, $j > i$ may *functionally* depend on $\{X_{t,i}\}$. Nevertheless, they are ensured to be *statistically* independent due to the dither, so we can regard $W_{t,i}$ as *additive* noise. This conforms to the fact that in the "dirty paper" setting (the Costa model) X is independent of the interfering signal S [18]. For an alternative treatment of this issue, see [8, Appendix], [98], [99].

As mentioned, the nested lattice approach yields here an algebraic binning strategy and is applied straightforwardly as described in Section IV-B. One important feature here is that the mechanism is not dependent on the Gaussianity of $\{S_{t,i}\}$ (6.7).[11] Further, also the first user $i = 1$, for which $S_{t,i} = 0$, $\forall t$, can use a capacity-approaching lattice code, as addressed in Section VII, which is indeed a special case. This provides a unifying nested lattice based approach of optimal signaling over the Gaussian broadcast channel.

In Fig. 17, the nested lattice approach is demonstrated for the $M = 2$ user broadcast channel. User $i = 1$ produces Gaussian codewords, which are decoded taking the full interference penalty of user 2 (who enjoys a better channel as $|h_2| \geq |h_1|$). This code can be based on a lattice approach, as described in Section VII-A, and so depicted. User 2 treats user 1, to whom power P_1 is assigned, as interference known beforehand at the transmitter $\{S_{t,2} = h_2 X_{t,1}\}_{t=1}^n$ and hence uses a nested lattice strategy. That is, user 2 selects a codeword from the fine lattice and then transmits an appropriate error vector based on a coarse lattice modulo operation.

The underlying interpretation giving rise to the nested lattice techniques, mentioned also in [9] and [101], conforms, in fact, to the basic insight provided by the Marton approach in deriving achievable rates for general broadcast channels [60] and evidently yields here, in the degraded channel case, the full capacity region. Specifically, Theorem 2 of Marton with the special case of W (Marton's [60] notations) chosen to be a constant, is directly related to this interpretation. The informa-

[11] Though here, Gaussianity of the marginals is preserved by the very fact that $X_{t,i}$ are so constructed as to achieve the capacity of the respective channel.

tion-carrying signal of the first user plays the role of a state sequence given beforehand to the transmitter for the second user. For this special case, the linkage to the Gelfand–Pinsker [41] setting is evident, as indeed mentioned in [41]. A well-known random binning interpretation of this special case (W-constant) of Marton's region is given in [40]. The focus in this paper, as demonstrated by the example at hand is to show specifically how in the realm of a Gaussian broadcast channel binning is efficiently implemented in terms of a nested lattice code.

Furthermore, there are some inherent advantages within the advocated setting as opposed to the conventional onion peeling at the decoder approach [19], [21]. Here, those users that experience better channels do not have to reliably decode the messages assigned to the users who experience degraded (worse) channels, and in fact the "better" users may even be fully ignorant of the codebooks assigned to those "degraded" users, a property which may be advantageous for secured communication. This occurs without affecting the achievable rate region as in (6.3).

Yet, it is not that absolutely no information is revealed about those users corresponding to the degraded channels, which are not reliably decodable within this paradigm, but rather that the amount of information is exactly the same as if i.i.d. inputs had been generated and the underlying coding strategy of those users was totally ignored. See the conclusions in [78], where the relevant result here corresponds to the case where the transmitter invests no special efforts in improving the state estimation at the receiver, and hence allocates no power to that purpose.

The ability to treat successfully the standard broadcast setting with no loss of optimality within the framework of nested lattice codes opens a variety of possibilities. Here we highlight the broadcast approach to a fading channel [69], [68]. Within this framework, in a single-user setting any channel gain realization is interpreted as a different (virtual) channel connected to a different user. In this framework of composite channels [5], the transmitter, which is not aware of the specific channel realization, is able to adapt the reliably transmitted information to the actual channel conditions. In [69], the continuous case, where the channel gain takes on real (or complex) realizations, is treated and the optimal average throughput strategy is explicitly identified.

The fact that we were able in this section to retain the Gaussian broadcast channel capacity region adhering to the nested lattice approach, makes this technique immediately applicable to the general broadcast approach to communicate over a composite channel [19]. Evidently within this application, the reliable decoded rate depends on the realization h and is a nondecreasing function of $|h|$, as all information decoded for a particular realization h_1 may also be decoded for a "better" realization h_2, where $|h_2| > |h_1|$. One may therefore wonder what, if at all, is the advantage of the nested lattice approach over the standard ordered decoding and cancellation in such a case where the information intended for some degraded gain realization is to be decoded anyhow. Note that with the nested lattice approach, all information streams are decoded in *parallel* (see Fig. 19) as opposed to serial decoding and cancellation. This strategy eliminates the error propagation associated with the standard approach, which manifests itself in the behavior of the associated error exponents. In certain cases, it may also provide some practical advantages of signal processing. Specifically, at the (usually small) cost of the additional complexity of the modulo-coarse-lattice operations at the encoder and the decoder, this "precoding" strategy saves the cost of the fine-lattice decoding of the interfering signal (directed for the bad receiver) at the good receiver. Obviously, extensions of the nested lattice technique to the broadcast approach for a multiple-access setting, where the different users communicate over fading channels [68] are also straightforward, and again rely on the very same principles as in the "dirty paper" case. This is also the case for a multiple-transmit/receive antenna, for which certain aspects of the broadcast approach are under study. Another aspect of nested lattices mentioned in Section V is its natural application to multiresolution [3] problems and successive refinement techniques [26], [65]. As indicated in [5], the successive refinement and the broadcast approach are perfectly matched to transmit over composite or compound channels where reliability (distortion) is refined as the channel realization improves. The possibility to treat both these aspects within the nested lattice paradigm demonstrates the rather broad scope of this idea.

C. The Multielement Broadcast Channel

We now turn our attention to the multiple-antenna broadcast example, which again is characterized by the basic equation (6.1). Within this representation, the vector $\boldsymbol{X}_t$ designates the power-constrained input into the K-transmitting antennas and the vector $\boldsymbol{Y}_t$ designates the associated signal received at the M antennas of the different and noncooperative users, each equipped with a single antenna. The matrix H is assumed known and fixed and it stands for the MIMO propagation coefficients. For simplicity, we assume here that $K = M$ and the matrix H has full rank K.

This model has recently been proposed and first treated in [8], [98], [99], and [74], [102], where the single-cell multiple-antenna broadcast channel and multicell single antenna per cell are studied, respectively. Subsequent efforts extending the original results of [8], [98], [99] appear in [7], [92], [103], [85], [84], and [49].

The broadcast transmission scheme advocated in [8], [98], [99] employs the lower triangular quadratic residue (QR-LQ) decomposition

$$H = GQ \tag{6.9}$$

where G is a $K \times K$ lower triangular matrix and Q is a $K \times K$ orthonormal matrix. The channel input is given by

$$\boldsymbol{X}_t = Q^* \boldsymbol{\theta}_t \tag{6.10}$$

where $\boldsymbol{\theta}_t$ is the information-carrying signal, and where superscript $*$ stands for the Hermitian transpose. The transformation (6.10) is power preserving

$$E|\boldsymbol{X}_t|^2 = E|\boldsymbol{\theta}_t|^2 = \text{SNR}. \tag{6.11}$$

This decomposition combined with (6.1) then yields

$$Y_{t,i} = g_{i,i}\,\theta_{t,i} + \sum_{j<i} g_{i,j}\,\theta_{t,j} + Z_{t,i}, \qquad \begin{array}{c} i = 1, 2, \ldots, K \\ t = 1, 2, \ldots, n \end{array}$$

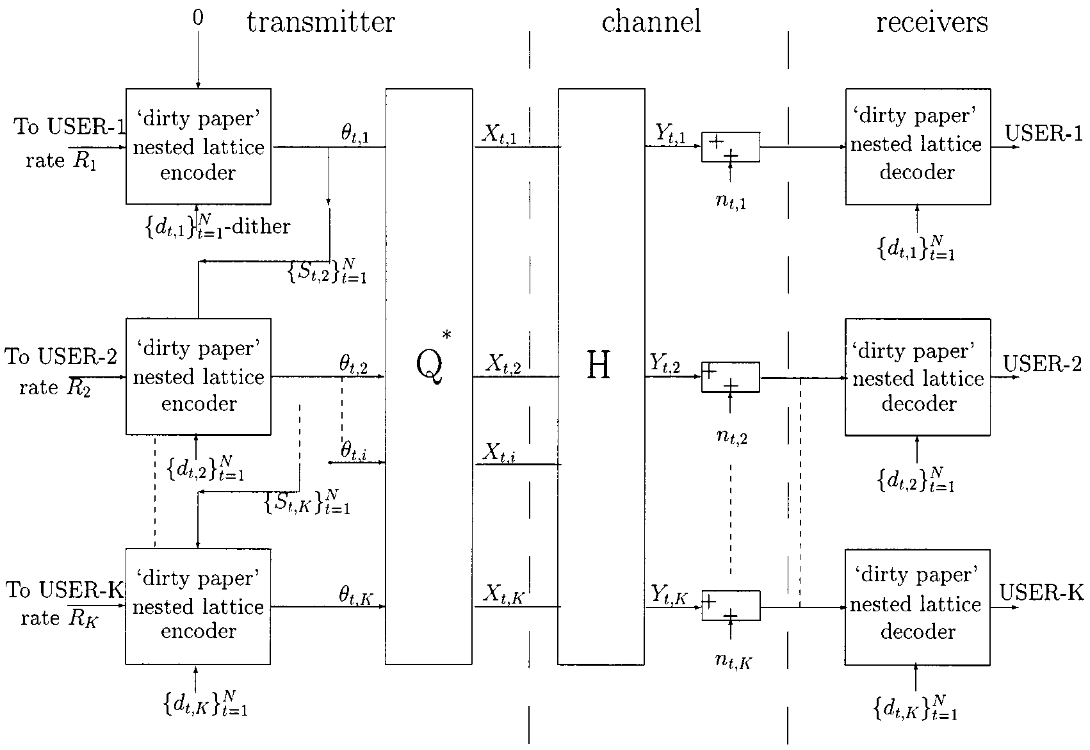

Fig. 18. The nested lattice encoding–decoding for the multiantenna broadcast channel.

(6.12)

where $g_{i,j}$ is the i, j entry of the matrix G. Again, the received signal has assumed the recognizable form of the "dirty paper" Costa setting where

$$S_{t,i}=\sum_{j<i} g_{ij}\,\theta_{t,j} \qquad \begin{array}{c} i=1,2,\ldots,K \\ t=1,2,\ldots,n \end{array} \tag{6.13}$$

plays the role of the interference sequence perfectly known beforehand of the transmitter.

The application of a nested lattice technique to this framework, as depicted in Fig. 18, is straightforward and relies again on the canonic blocks of the "dirty paper" nested lattice encoders and decoders, as presented in Section IV-B and depicted in Figs. 13–15. The resultant achievable rates of this approach, coined in [8], [98], [99] "ranked known interference (RKI)," is evidently given by

$$\bigcup_{\{P_i\}}\left\{R_i\le\frac{1}{2}\log\left(1+|g_{ii}|^2 P_i\right), \qquad i=1,2,\ldots,K\right\} \tag{6.14}$$

where the union is taken over all power assignments $P_i = E|\theta_{t,i}|^2$ such that the average power constraint (6.11)

$$\sum_{i=1}^{K} P_i \le \text{SNR} \tag{6.15}$$

is satisfied.

A generalization of the RKI approach is also suggested in [8], [98], [99], where the transformation used is

$$\boldsymbol{X}_t = Q^* B\,\boldsymbol{\theta}_t. \tag{6.16}$$

The matrix B is an upper triangular matrix satisfying $\operatorname{trace}\frac{1}{K}(BB^*) \le 1$, so as to maintain the average power constraint (6.11). The RKI approach turns then to be the special case of $B = I$. This transformation (6.16) gives rise to the familiar equation

$$Y_{t,i}=\rho_{i,i}\,\theta_{t,i}+\sum_{j<i}\rho_{i,j}\,\theta_{t,j}+\sum_{j>i}\rho_{i,j}\,\theta_{t,j}+Z_{t,i} \tag{6.17}$$

where ρ_{ij} are appropriate coefficients determined by the matrices G, Q, B. Here

$$S_{t,i}=\sum_{j<i}\rho_{ij}\,\theta_{t,j} \tag{6.18}$$

$$W_{t,i}=\sum_{j>i}\rho_{ij}\,\theta_{t,j} \tag{6.19}$$

are identified to be, respectively, the post-cursor and precursor elements as in (6.7) and (6.8). Again, the nested lattice precoding for user i eliminates the effect of $\{S_{t,i}\}_{t=1}^{n}$ known beforehand at the transmitter for $t = 1, 2, \ldots, n$, giving rise to the achievable set of rates

$$\bigcup_{\{P_i\}}\left\{R_i\le\frac{1}{2}\log\left(1+\frac{|\rho_{i,i}|^2 P_i}{1+\sum\limits_{j>i}|\rho_{i,j}|^2 P_j}\right), \qquad i=1,2,,\ldots,K\right\}. \tag{6.20}$$

The union operation in (6.20) is as in (6.14). The entries of the matrix B and the power assignment $\{P_i\}$ subjected to the average power constraint (6.15), can be chosen to optimize certain rate features as the total throughput $\sum_{i=1}^{K} R_i$ [8], [98], [99]. Indeed, for $K = 2$, it was demonstrated in [8], [98], [99] that the generalized RKI method achieves the optimal throughput in this multiple antenna, generally nondegraded, broadcast-channel setting. This optimality holds also for general K [7]. It was further demonstrated that even the basic RKI (where $B = I$) technique is asymptotically optimal in the high and low signal-to-noise ratio (SNR) regions, where in the latter case it reduces to the beamforming zero forcing technique [8], [98], [99].

The application of the nested lattice technique in this generalized RKI approach operating over the multiantenna broadcast is straightforward and depicted in Fig. 18, where Q^* is replaced by Q^*B. This application is again, in principle, the same as in the broadcast channel or the ISI case of Section VII as expression (6.17), is of the very same structure as (6.6) and (7.5), respectively. A similar decomposition to the basic RKI transformation (6.9) was suggested in [42] to combat far-end crosstalk for a discrete multitone-based system. Their approach, i.e., precoding, is however Tomlinson-like [64], [34], and as such subjected to shaping, modulo and power losses. The possible use of a Tomlinson precoder in a coordinated transmission setting has also been mentioned in [35]. The RKI technique, which interprets the LQ decomposition in terms of the "dirty paper" channel, is free of all the above degradations as described, and the nested-lattice technique which implements the full potential of the RKI (either basic or generalized) technique is indeed a natural coding/signaling strategy in this multiantenna broadcast setting. Again, the precursor (6.19) interference inherently cannot be alleviated, since symbols in the precursor are coded in order to eliminate the effect of their post-cursors, containing the current symbol. This is evident, as otherwise one could falaciously surpass the maximum achievable throughput capacity (even in the $K = 2$ case).

In [74], [75] the very same ideas are applied to a somewhat different setting of a multicell downlink, and it was shown that multicell central processing has a fundamental impact on the achievable throughput. Subsequent developments of this application based on the results in [85] are reported in [49], where also per-cell-site antenna power constraints are examined.

We have demonstrated the results here assuming that the MIMO propagation matrix H is given. Extensions to H being a realization of a random matrix process can also be treated, as is done in [8], [98], [99].

The approach of [8], [98], [99] has recently been extended in [92], [103], where it has been shown that the sum rate of the multielement broadcast channel can be achieved using the "dirty paper" principles as in [92], [103], for any number of users and any number of elements (antennas). To that end the vector generalization of Costa's approach [93] has been invoked. This observation has also been made in [85], where the rate region as in [92], [103] was interpreted in its dual setting as the union of capacity regions of an associated multiple-access channel, over all power assignments among its users subjected to a total average power constraint. Nested lattice coding as described here can immediately be used in the generalized setting as well, noting that the vector "dirty paper" setting [93] breaks up to a set of parallel scalar standard "dirty paper" channels, via the classical singular value decomposition applied to the original vector channel. This duality has also been exploited in [84] to show the throughput optimality, presenting an elegant rigorous proof. The nested lattice approach in this application, as first advocated in [8], [98], [99], is intimately and directly related to trellis precoding, as applied later in [92], [103]. Note that standard trellis precoding, when applied to the "dirty paper" configuration is optimal only at asymptotically high SNR, while the nested lattice precoding scheme is optimal throughout the whole SNR region, and that this is due to the introduction of the inflation factor α (which in fact can also be combined with the trellis precoding strategy).

We have focused on the multiple-antenna broadcast setting. However, the same model is applicable to a variety of broadcast applications as, for example, that of high-speed twisted pair wire-line communications [42]. For the sake of simplicity, we have restricted our attention to real-valued H_t and $\boldsymbol{X}_t$ in (6.1). Extensions to circularly complex valued matrices and vectors are straightforward and mainly require proper normalizations, which are already accounted for in (6.9)–(6.11). Having this extension in mind is the reason for invoking the complex notation in (6.10), (6.11), and (6.16).

VII. Nested Codes in Point-to-Point Communication

In this section, we demonstrate that the nested lattice approach presented can also be relevant to standard point-to-point problems. We point out two examples, lattice codes and decoding for the AWGN channel, and achieving the capacity of Gaussian dispersive (ISI) channels through precoding.

A. Nested Codes With Lattice Decoding for the AWGN Channel

This application turns out to be a by-product of the encoding–decoding scheme of Section IV-B (the Costa problem). We show that using nested codes in conjunction with dithering techniques, the power-constrained AWGN channel can be transformed into a modulo lattice additive noise channel having the same capacity as that of the original channel. By so doing we are able to retain the "one" in the capacity formula of the AWGN channel which was sacrificed in prior works on lattice decoding (see discussion in [53]). This allows *lattice decoding* to be optimal for all SNRs.

De Buda's Theorem [23] states that a lattice code, cut into a bounded region with second moment P, can approach arbitrarily close (in the limit of high dimension) the capacity $\frac{1}{2}\log(1 + P/N)$ of an AWGN channel at SNR $= P/N$. This result has been corrected and refined by several investigators, see [57], [83], [62], [59]. The optimality of this scheme relies upon maximum-likelihood decoding, i.e., on finding the lattice point *inside the bounded region* which is closest to the received signal. In contrast, "lattice decoding" amounts to finding the closest lattice point, ignoring the boundary of the code. Such an unconstrained search seems to save complexity, and retains codewords' symmetry, and thus attracted some special attention. However, existing lattice coding schemes with lattice decoding can transmit reliably only at rates up to $\frac{1}{2}\log P/N$. This loss of "one" in rate means significant degradation in performance for low SNR. In fact, it was conjectured [59] that lattice decoding is optimum only at high SNR, i.e., that with lattice decoding the rate $\frac{1}{2}\log P/N$ cannot be surpassed. See also discussion in [53].

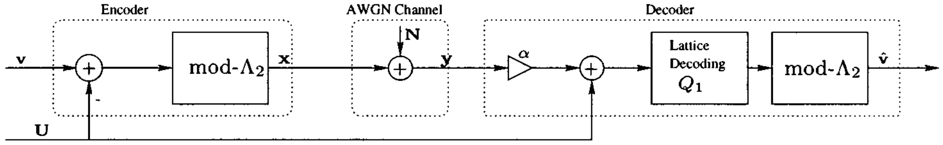

Fig. 19. Nested lattice encoding–decoding scheme for AWGN.

In [29], [30] it is shown that the encoding scheme of Section IV-B may be applied, "as is," to the ordinary (no side information) AWGN channel. This scheme, along with *lattice decoding*, preserves the one in the capacity formula. Specifically, by taking $S \equiv 0$ we may regard the encoding scheme as shown in Figs. 13 and 14 as a dithered lattice code transmitter along with a lattice decoding receiver, for the AWGN channel. Since it was shown in Section IV-B that lattice decoding may achieve the capacity of the Costa channel using the latter scheme, it follows that the same is true for the AWGN channel, regarded as a special case $(S \equiv 0)$. The resulting encoder–decoder is depicted in Fig. 19 and its equivalent channel is the same as for the Costa nested coding scheme (Fig. 15). Notice that the lattice decoding operation Q_1 does not depend on the exact distribution of the equivalent noise (4.35); rather, it is a Euclidean nearest neighbor quantizer, as if the equivalent noise was AWGN.

We note that this transmission scheme in effect transforms the original AWGN channel into a modulo-lattice additive noise one. This transformation is not strictly information lossless in the sense that it does not preserve the mutual information. However, the (information) loss goes to zero as the dimension of the code k goes to infinity. This suffices for achieving the channel's capacity, albeit may result in a suboptimal error exponent.

B. Dispersive (ISI) Channels

We now show that the classical ISI channel also falls into the general framework of Section VI. Consider the dispersive Gaussian channel, which can be expressed by

$$Y_t = \sum_j h_j \tilde{X}_{t-j} + Z_t, \qquad t = 1, 2, \ldots, n. \tag{7.1}$$

This can be viewed as a special case of (6.1), with $M = 1$ and where $H_t = H$ is an countably infinite "row vector," i.e., a sequence, with components $\{h_j\}$. In this notation $\boldsymbol{X}_t$ in (6.1) is composed of overlapped vectors of components $(\tilde{X}_t, \tilde{X}_{t-1}, \ldots, \tilde{X}_{t-j} \cdots)^T$, which are shifted by a single coordinate, with time, where $\{\tilde{X}_t\}$ stand for the scalar inputs to the ISI Gaussian channel. In this application, $\{Y_t\}$ designates the received samples and $\{h_j\}$, $j \neq 0$ are the ISI coefficients. The additive Gaussian noise samples are denoted by $\{Z_t\}$. We assume that this equation represents the sampled output of the matched filter which preserves information. Suboptimal filtering can also be represented within the setting of (7.1), as in [13].

We further assume that the ISI coefficients $\{h_j\}$ account for the transmission filter, channel time-invariant transfer function characteristics, and the receiver matched filter. The information-carrying input symbols $\{\tilde{X}_i\}$ are assumed to be i.i.d. Gaussian. The input–output mutual-information normalized per channel use equals

$$I_N(\{Y\}\colon \{\tilde{X}\}) = \frac{1}{2}\log(1+\gamma) \tag{7.2}$$

where

$$\gamma = \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(1 + \mathrm{SNR}\left|H(e^{-j\theta})\right|^2\right) d\theta\right) - 1 \tag{7.3}$$

and where

$$H(D) = \sum_j h_j D^j \tag{7.4}$$

stands for the formal D transform of the associated ISI coefficients [12]. Here, SNR designates the signal-to-noise ratio. The expression I_N may, in fact, equal the original channel capacity provided the transmission-shaping filter is selected to implement the capacity-achieving water-pouring spectrum over the original Gaussian dispersive channel. Hence, no optimality loss is incurred by assuming i.i.d. information-carrying inputs $\{\tilde{X}_i\}$ [70]. We focus here on the feedforward MMSE decision feedback (MMSE-DFE) equalizing filter [12], the output of which at time epoch t is given by

$$V_t = X_t + \sum_{j=1}^{\infty} \tilde{h}_j \tilde{X}_{t-j} + \sum_{j=1}^{\infty} \tilde{h}_{-j} \tilde{X}_{t+j} + M_t. \tag{7.5}$$

The post-cursor (causal) and precursor (anticausal) parts are designated by S_t and W_t, respectively, and are given by

$$S_t = \sum_{j=1}^{\infty} \tilde{h}_j \, \tilde{X}_{t-j} \tag{7.6}$$

$$W_t = \sum_{j=1}^{\infty} \tilde{h}_{-j} \, \tilde{X}_{t+j} \tag{7.7}$$

resembling a stationary version of the MIMO broadcast example as in (6.17). Here $\{\tilde{h}_j\}$ designates the ISI coefficients at the output of the MMSE-DFE feed-forward filter and $\{M_t\}$ stands for the corresponding filtered noise samples. This is referred to in [12] as a canonic form mainly due to the fact that mutual information is preserved on a symbol-by-symbol basis, provided an ideal DFE is available as to cancel the post-cursor (S_t) effect. This can be seen by calculating the associated SNR

$$E(X_t^2)/[E(W_t^2) + E(M_t^2)] = \gamma \tag{7.8}$$

referred to as the MMSE-DFE-U SNR, with U standing for "unbiased," meaning that in (7.5) the equivalent precursor noise term is uncorrelated with the desired signal [12]. Evidently (see [12], [104]), the sample-per-sample-wise mutual information, assuming ideal DFE $I(V_t;\, \tilde{X}_t | \tilde{X}_1,\, \tilde{X}_2,\, \ldots,\, \tilde{X}_{t-1})$ equals the full input–output mutual information $I_N(\{Y_t\}; \{\tilde{X}_t\})$ (or, as said, the optimal channel capacity), provided fully reliable feedback decisions are available. This observation motivated the introduction of a capacity-approaching coding strategy [44] where the decisions fed back are taken after decoding of a capacity-approaching code is completed, and therefore decisions are reliable. The basic scheme is sketched in Fig. 20 for the time continuous dispersive channel. The central mechanism is the interleaver/deinterleaver guaranteeing that post-cursor symbols be-

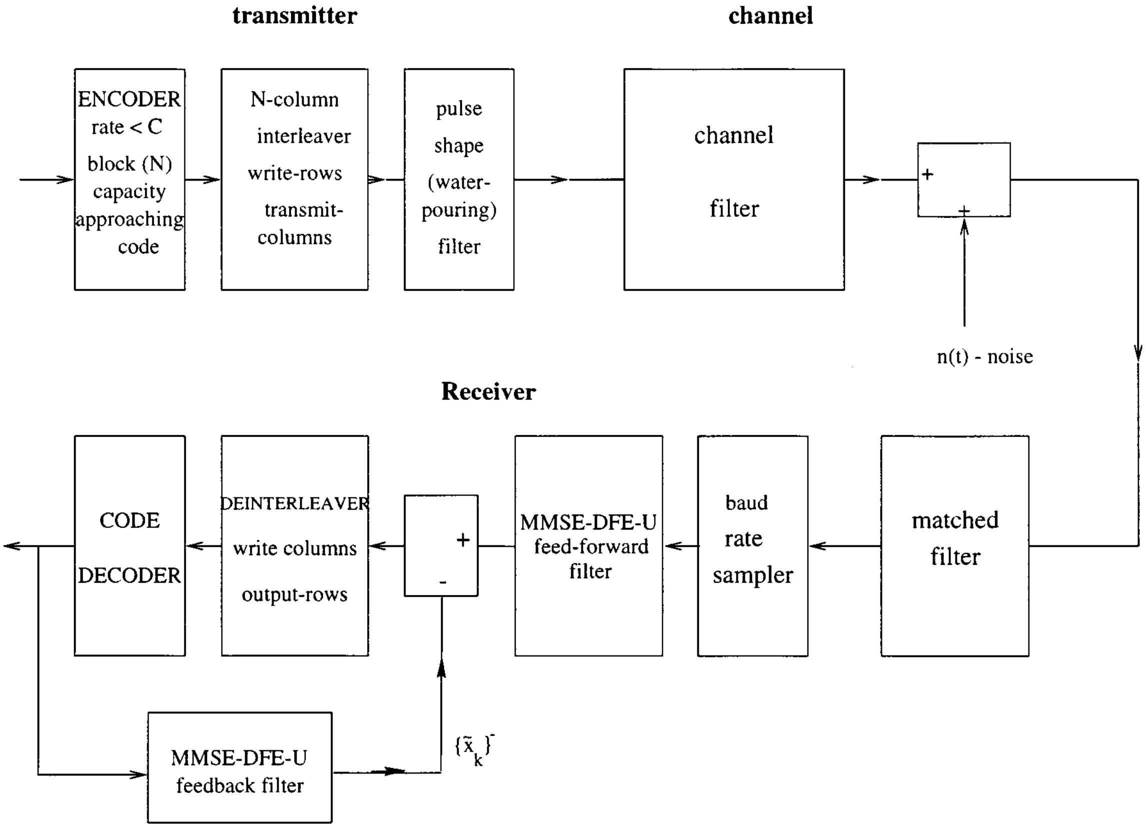

Fig. 20. The Guess–Varanasi MMSE-DFE coding strategy for the Gaussian dispersive channel.

long to different codewords, which were already decoded. This is why, in principle, when capacity-approaching codes are used, the decoding error probability at the near-capacity operational point is guaranteed to be negligible. This is in contrast to previously used DFE schemes, where the feedback filter makes use of "unreliable" symbol-by-symbol decisions. Thus, reliable post-cursor symbols are fed back and ideally removed before the current codeword gets decoded.

The scheme we advocate here uses in fact the interleaver/deinterleaver as in Fig. 20. However, the DFE part is replaced by nested lattice *precoding* as in the "dirty paper" case. Note that when codeword j, which comprises the jth row in the interleaver, is to be encoded, the post-cursor interference designated by the signals $\{S_{t,j}\}$ comprises, due to the interleaving, previous codewords. Since, as described in Section IV-B, nested lattice precoding is capable of retaining capacity over the "dirty paper" channel (and an AWGN channel as a special case), this approach achieves capacity over the regular ISI channel. This strategy is depicted in Fig. 21 for the continuous Gaussian dispersive channel. The setting is identical to that of Fig. 20, but for the coding which is replaced by the "dirty paper" encoder and the decoding by the "dirty paper" decoder, where the latter does not require any DFE loop. The blocks of the "dirty paper" encoder and decoder are shown in Figs. 13 and 14.

Evidently, precoding techniques are widely used over the ISI channel. While standard Tomlinson precoding [64] suffers inherent degradation of power loss, modulo-loss, and shaping loss [70], [87], the more sophisticated trellis precoding [33] and Laroia precoding [56] avoid the shaping and power losses but are still subjected to the modulo loss in the low SNR region. In the extreme case, where the coarse lattice in our scheme is taken to be a (Cartesian product of a) scalar quantizer, the nested precoding scheme actually becomes almost identical to combined coding and Tomlinson precoding. The crucial difference is the scaling factor α which in this case must be optimized numerically (as a function of the SNR), see [28]. Even if we take Costa's $\alpha = P/(P+N)$, the rate loss of this scheme is upper-bounded by the shaping gain, $1/2\log(2\pi e/12) \approx$ 0.254 bit, at any SNR. The Laroia precoding [56] technique which is part of the V-34 modem standard, might be viewed as a certain lattice/trellis-based precoding, which guarantees close to capacity performance at high rates (asymptotically high SNR conditions). See also [34], where lattice precoding for Tomlinson-based processing has been addressed. Nested lattice coding–decoding, as advocated here, approaches capacity at all rates and SNR values. In fact, the inflated lattice technique manifests itself in the introduction of the scalar α and the uniform dithering (see Section IV-B). These are crucial elements so as to guarantee near-capacity operation throughout the whole SNR region and not only at high SNR scenarios.

Evidently, this ISI precoding scheme can easily be generalized to the multiple-antenna (MIMO) single-user channel with the propagation matrix H (replacing the time-invariant channel filter h_j of (7.1)) available to both the transmitter and the receiver [79]. In fact, this precoding technique is immediately applicable to a MIMO, Bell Laboratories Layered Space–Time (BLAST) [38] type communication where the Cholesky-based filtering is performed at the receiver and the resultant interdata

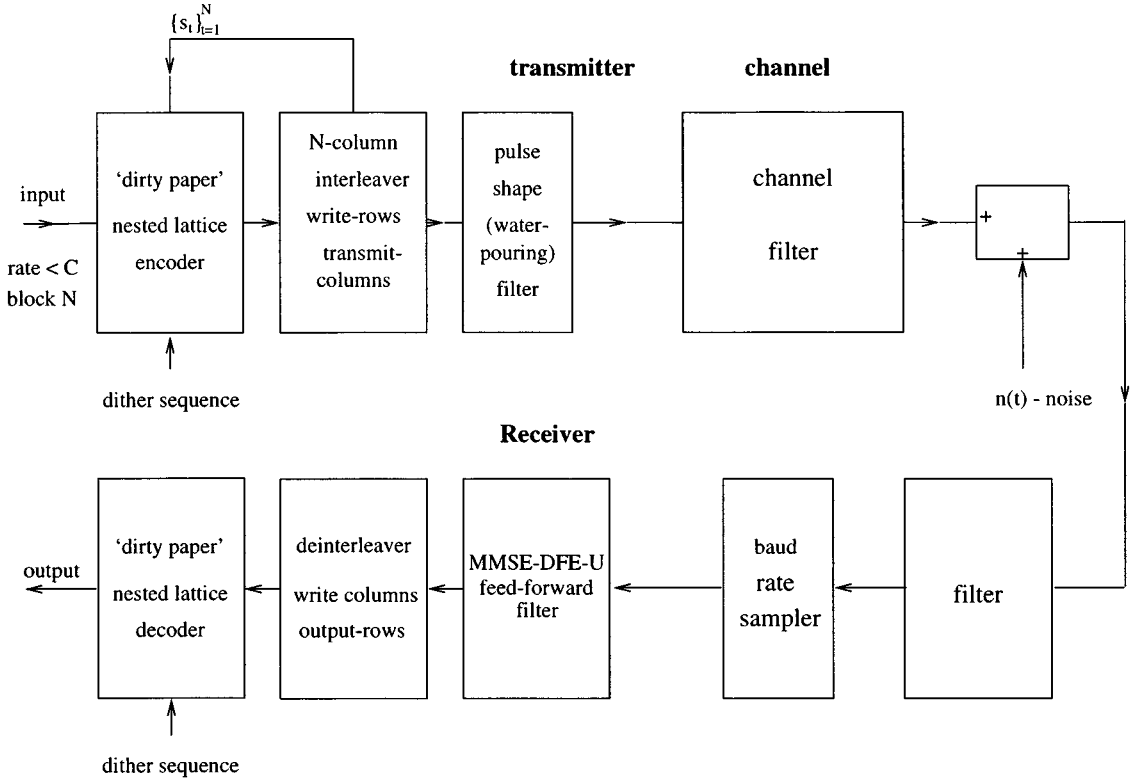

Fig. 21. The "dirty paper" capacity-approaching strategy for the Gaussian dispersive channel.

stream interference, usually coped with by post-decoding DFE [38], is absolutely eliminated by the "dirty paper" precoding as described here. A Tomlinson-type precoding for this setting has been presented in [37], [36].

A closing comment refers to the effect of the precursor designated by W_t (7.7) in both the broadcast (as presented in the previous section) and ISI settings. The interference from this part is taken in full, and that despite of the fact that the precursor is also composed, in principle, of symbols produced at the transmitter. Any mitigation of interference associated with this signal (precursor) is inherently prohibited (in the broadcast and MMSE-DFE-U ISI setting). This is evident as otherwise the achievable rate region would fallaciously surpass the ultimate broadcast channel rate region and the dispersive Gaussian channel capacity. This is a manifestation of the fact that future symbols (part of the precursor) represent codewords which are so produced as to mitigate the interference of past (post-cursor) symbols. In the MMSE-DFE ISI channel setting, this inherent limitation is another manifestation of the inherent disability to implement vector processing combined with ideal DFE, as this yields a fallacious surpass of the ultimate Shannon capacity [70], [44].

VIII. Conclusion

In this tutorial paper, we have presented the paradigm of algebraic binning in an effort to encompass a class of the rich information-theoretic settings where random binning ideas are applied and beyond. The underlying framework is the nested codes, which have recently been studied extensively in a variety of applications. The unified framework of nested structured coding encompasses also other settings, where binning seems not to be necessary, but yet, the nested approach facilitates a new angle of perspective. Multiple-description lattice quantizers [107] seem to have similar structure, but were not considered in this work.

We examine, first, noiseless source coding problems with side information available to the decoder and dual schemes of constrained point-to-point transmissions, with side information available to the transmitter only, adhering basically to Wyner's coset coding. We introduce the notion of nested codes either on a binary or continuous alphabets, and extend then the discussion to noisy side information settings, as the Wyner–Ziv rate distortion problem, and the Costa, "writing on dirty paper" and the Kuznetsov–Tsybakov–Heegard–El-Gamal defected memory problems. We note also that since the Costa problem and the problem of digital watermarking are equivalent, the nested lattice scheme also provides a capacity-achieving solution to the latter problem. The basic building blocks of nested coding are then used to address classical multiterminal problems, as correlated sources encoding–decoding, degraded Gaussian broadcast channel, and a nondegraded Gaussian multiple-antenna broadcast setting. We also study the nested coding approach on the standard point-to-point AWGN channel, and the classical dispersive Gaussian channel, and demonstrate how capacity can be achieved adhering to the basic building blocks of nested coding. It is emphasized that trellis precoding via Voronoi region coding as in [51], [33] can also be interpreted in terms of nested lattice codes and, in fact, this may serve as a possible practical construction of nested lattice codes.

As a closing remark, it is interesting to mention that nested codes were recently speculated to be the central ingredient in the accurate replication of the Genome [2].

ACKNOWLEDGMENT

The authors wish to thank David Forney, Simon Litsyn, and Sergio Verdú for valuable comments and suggestions at various stages of the work.

REFERENCES

[1] R. J. Barron, B. Chen, and G. W. Wornell, "On the duality between information embedding and source coding with side information and some applications," in *Proc. Int. Symp. Information Theory*, Washington, DC, June 2001, p. 300.

[2] G. Battail, "Is biological evolution relevant to information theory and coding," in *Int. Symp. Communication Theory and Applications (ISCTA'01)*, Ambleside, U.K., July 15–20, 2001, pp. 343–351.

[3] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.

[4] T. Berger and R. Yeung, "Multiterminal source encoding with one distortion criterion," *IEEE Trans. Inform. Theory*, vol. 35, pp. 228–236, Mar. 1989.

[5] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2693, Oct. 1998.

[6] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1987.

[7] G. Caire and S. Shamai (Shitz), "On the multiple antenna broadcast channel," in *Proc. 35th Asilomar Conf.*, Pacific Grove, CA, Oct./Nov. 2001.

[8] ——, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[9] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May 2001.

[10] A. S. Cohen and A. Lapidoth, "Generalized writing on dirty paper," in *Proc. Int. Symp. Information Theory*, Lausanne, Switzerland, June 2002, to be published.

[11] J. Chou, S. S. Pradhan, and K. Ramchandran, "On the duality between distributed source coding and data hiding," in *Proc. 33rd Asilomar Conf. Signals, Systems and Computers*, vol. 2, Pacific Grove, CA, Oct. 1999, pp. 1503–1507.

[12] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding—Part I: Equalization results, Part II: Coding results," *IEEE Trans. Commun.*, vol. 43, pp. 2582–2604, Oct. 1995.

[13] M. Cioffi and G. D. Forney, "Generalized decision-feedback equalization for packet transmission with ISI and Gaussian noise," in *Communications, Computation, Control and Signal Processing, A Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds. Boston, MA: Kluwer, 1997, pp. 79–129.

[14] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, pp. 0000–0000, June 2002. See also *Proc. Int. Symp. Information Theory*, Sorrento, Italy, June 2000, p. 48.

[15] J. H. Conway, E. M. Rains, and N. J. A. Sloane, "On the existence of similar sublattices," *Canad. J. Math*, vol. 51, no. 6, pp. 1300–1306, 1999.

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

[17] ——, "Voronoi regions of lattices, second moments of polytops, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, Mar. 1982.

[18] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.

[19] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2–14, Jan. 1972.

[20] T. M. Cover and M. Chiang, "Duality of channel capacity and rate distortion with two sided state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 0000–0000, June 2002. See also *Proc. Int. Symp. Information theory and Its Applications*, Honolulu, HI, Nov. 2000, pp. 517–520.

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

[22] I. Csiszár and J. Körner, *Information Theory—Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[23] R. de Buda, "The upper error bound of a new near-optimal code," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 441–445, July 1975.

[24] E. C. Van der Meulen, "A survey of multi-way channels in information theory," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, Jan. 1977.

[25] J. J. Eggers, J. K. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," in *IEE Colloquium: Secure Images and Image Authentication*, London, U.K., Apr. 2000.

[26] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 851–857, Nov. 1991.

[27] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice-strategies for cancelling known interference," *IEEE Trans. Inform. Theory*, submitted for publication.

[28] ——, "Capacity and lattice-strategies for cancelling known interference," in *Proc. ISITA 2000*, Honolulu, HI, Nov. 2000, pp. 681–684.

[29] U. Erez and R. Zamir, "Lattice decoding can achieve $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel," in *Proc. Int. Symp. Information Theory (ISIT2001)*, Washington, DC, June 2001, p. 300.

[30] ——, "Lattice decoding can achieve $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel," *IEEE Trans. Inform. Theory*, 2001, submitted for publication.

[31] ——, "Lattice decoded nested codes achieve the Poltyrev exponent," in *Proc. Int. Symp. Inform. Theory (ISIT2002)*, Lausanne, Switzerland, June 2002, to be published.

[32] ——, "Bounds on the ϵ-covering radius of linear codes with application to self noise in nested Wyner–Ziv coding," Dept. Elec. Eng.–Syst., Tel-Aviv Univ., Tech. Rep., Feb. 2002.

[33] M. V. Eyuboglu and G. D. Forney, Jr., "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 301–314, Mar. 1992.

[34] R. Fischer, "Precoding and signal shaping for digital transmission," Habilitationsschrift, Univ. Erlangen-Nurnberg, Nurnberg, Germany, 2001.

[35] R. Fischer, J. Huber, and G. Kamp, "Coordinated digital transmission: Theory and examples," *AEU (Elect. and Commun.)*, vol. 48, pp. 289–300, Nov./Dec. 1994.

[36] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Tomlinson–Harashima precoding in space–time transmission for low-rate backward channel," in *Proc. 4th ITG, Conf. Source and Channel Coding*, Berlin, Germany, Jan. 28–30, 2002, pp. 139–143.

[37] ——, "Space–time using Tomlinson–Harashima precoding," in *Proc. 39th Annu. Allerton Conf. Communication, Control, and Computing*. Monticello, IL: Allerton House, Oct. 3–5, 2001.

[38] G. Foschini, R. Fischer, J. Huber, and G. Kamp, "Layered space–time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Syst. Tech. J.*, pp. 41–46, Autumn 1996.

[39] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[40] A. El-Gamal and E. C. van der Muellen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 122–123, Jan. 1981.

[41] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Pered. Inform. (Probl. Inform. Transm.)*, vol. 9, no. 1, pp. 19–31, 1980.

[42] G. Ginis and J. M. Cioffi, "Vectored-DMT: A FEXT canceling modulation scheme for coordinating users," in *Proc. Int. Conf. Communications (ICC2001)*, vol. 1, Helsinki, Finland, June 2001, pp. 305–309.

[43] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.

[44] T. Guess and M. Varanasi, "A new successively decodable coding technique for intersymbol-interference channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT'00)*, Sorrento, Italy, June 2000, p. 102.

[45] C. Heegard, "Partitioned linear block codes for computer memory with 'stuck-at' defects," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 831–842, Nov. 1983.

[46] C. Heegard and T. Berger, "Rate-distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 727–734, Nov. 1985.

[47] C. Heegard and A. El-Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 731–739, Sept. 1983.

[48] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 371–377, May 1977.

[49] S. S. Jafar and A. Goldsmith, "Multiuser multicellular multiple antenna systems," in *Proc. Int. Symp. Inform. Theory (ISIT2002)*, Lausanne, Switzerland, June 2002, to be published.

[50] G. D. Forney, Jr., "Multidimensional constellation—Part II: Voronoi Constellations," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 941–958, Aug. 1989.

[51] ——, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 281–300, Mar. 1992.

[52] ——, "Coset codes—I: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, Sept. 1988.

[53] G. D. Forney, Jr., M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 820–850, May 2000.

[54] A. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 828–840, Nov. 1982.

[55] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered.. Inform.*, vol. 10, pp. 52–60, Apr.–June 1974. Translated from Russian.

[56] R. Laroia, "Coding for intersymbol interference channels—Combined coding and precoding," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1053–1061, July 1996.

[57] T. Linder, Ch. Schlegel, and K. Zeger, "Corrected proof of de Buda's theorem," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1735–1737, Sept. 1993.

[58] S. Litsyn, private communication.

[59] H. A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1767–1773, Nov. 1997.

[60] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 374–377, May 1979.

[61] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912–1923, Nov. 1997.

[62] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 409–417, Mar. 1994.

[63] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (discus): design and construction," in *Proc. IEEE Data Compression Conf.*, Snowbird, UT, Mar. 1999.

[64] J. R. Price, "Nonlinearly feedback-equalized PAM vs. capacity for noisy filter channels," in *Proc. Int. Conf. Communications (ICC'72)*, June 1972, pp. 22.12–22.17.

[65] B. E. Rimoldi, "Successive refinement of information: Characterization of the achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[66] B. Rimoldi and R. Urbanke, "Asynchronous Slepian–Wolf coding via source-splitting," in *Proc. Int. Symp. Information Theory (ISIT97)*, Ulm, Germany, June 1997, p. 271.

[67] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, Oct. 1958.

[68] S. Shamai (Shitz), "A broadcast approach for the multiple-access slow fading channel," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 128.

[69] ——, "A broadcast strategy for the Gaussian slowly fading channel," in preparation. See also *IEEE Int. Symp. Information Theory (ISIT'97)*, Ulm, Germany, June 29–July 4, 1997.

[70] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1388–1404, Sept. 1996.

[71] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Information theoretic aspects of systematic coding," in *Proc. Int. Symp. Turbo Codes and Related Topics*, ENST de Bretagne, Brest, France, Sept. 1997, pp. 40–46.

[72] ——, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 564–579, Mar. 1998.

[73] ——, "Systematic lossy source/channel coding," in *Proc. 1996 Int. Symp. Information Theory and Its Applications (ISITA'96)*, Victoria, BC, Canada, Sept. 17–20, 1996, pp. 513–516.

[74] S. Shamai (Shitz) and B. M. Zaidel, "Enhancing the cellular down-link capacity via co-processing at the transmitting end," in *Proc. IEEE Semiannu. Vehicular Technology, VTC2001 Spring Conf.*, ch. 3, sec. 36, Rhodes, Greece, May 6–9, 2001.

[75] S. Shamai (Shitz), B. M. Zaidel, and S. Verdú, "On information theoretic aspects of multi-cell wireless systems," in *Proc. 4th Int. ITG Conf. Source and Channel Coding*, Berlin, Germany, Jan. 28–30, 2002, invited paper, pp. 199–209.

[76] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[77] J. K. Su, J. J. Eggers, and B. Girod, "Channel coding and rate distortion with side information: Geometric interpretation and illustration of duality," *IEEE Trans. Inform. Theory*, submitted for publication.

[78] A. Sutivong, T. M. Cover, and M. Chiang, "Tradeoff between message and state information rates," in *Proc. 2001 IEEE Int. Symp. Information Theory (ISIT 2001)*, Washington, DC, June 24–29, 2001, p. 103.

[79] E. Teletar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun. (ETT)*, vol. 10, pp. 585–596, Nov./Dec. 1999.

[80] M. Tomlinson, "New automatic equalizer employing modulo arithmetic," *Electron. Lett.*, vol. 7, pp. 138–139, Mar. 1971.

[81] B. S. Tsybakov, "Defect and error correction," *Probl. Peredach. Inform.*, vol. 11, pp. 21–30, July–Sept. 1975. Translated from Russian.

[82] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 55–67, Jan. 1982.

[83] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, Jan. 1998.

[84] P. Viswanath and D. N. C. Tse, "Sum capacity of the multiple antenna broadcast channel," in *Int. Symp. Inform. Theory (ISIT2002)*, Lausanne, Switzerland, to be published.

[85] S. Viswanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. Int. Conf. Ccommunications 2002*, New York, NY, April 28–May 2, 2002.

[86] L. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 483–501, July 1987.

[87] R. D. Wesel and J. M. Cioffi, "Achievable rates for Tomlinson–Harashima precoding," *IEEE Trans. Inform. Theory*, vol. 44, pp. 825–831, Sept. 1998.

[88] F. M. J. Willems, "Signalling for the Gaussian channel with side information at the transmitter," in *Proc. Int. Symp. Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 348.

[89] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inform. Contr.*, vol. 38, pp. 60–80, 1978.

[90] ——, "Recent results in Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2–10, Jan. 1974.

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

[92] W. Yu and J. M. Cioffi, "Trellis precoding for the broadcast channel," in *Proc. GLOBECOM*, San Antonio, TX, Nov. 26, 2001, preprint.

[93] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang, "Writing on colored paper," in *Proc. Int. Symp. Information Theory (ISIT2001)*, Washington, DC, June 24–29, 2001, p. 302.

[94] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, vol. 45, pp. 106–117, Jan. 1999.

[95] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1152–1159, July 1996.

[96] ——, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.

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

[98] G. Caire and S. Shamai (Shitz), "On achievable rates in multiple-antenna broadcast downlink," in *Proc. 38 Allerton Conf. Communication, Control and Computing*, Allerton House, Monticello, IL, Oct. 4–6, 2000.

[99] G. Caire and S. Shamai (Shitz), "On achievable rates in multiple-antenna broadcast downlink," in *Proc. 2001 IEEE Int. Symp. Information Theory*, Washington, DC, June 24–29, 2001, p. 147.

[100] G. D. Forney, "On the duality of coding and quantizing," in *DIMACS Ser. Discr. Math. Theory Comp. Sci.*, 1993, vol. 14.

[101] J. Kusuma and K. Ramachandran, "Communicating by cosets and applications to broadcast," in *Proc. Conf. Information Sciences and Systems*, Princeton, NJ, Mar. 20–22, 2000.

[102] S. Servetto, "Quantization with side information: Lattice codes, asymptotics, and applications in wireless networks," *IEEE Trans. Inform Theory*. See also *Proc. Data Compression Conf (DCC00)*, Snowbird, UT, Mar. 2000, pp. 510–519, submitted for publication.

[103] W. Yu and J. M. Cioffi, "The sum capacity of a Gaussian vector broadcast channel," *IEEE Trans. Inform. Theory*, submitted for publication.

[104] T. Guess and M. Varanassi, "An information-theoretic derivation of the MMSE decision-feedback equalizer," in *Proc. 36th Allerton Conf.*, Monticello, IL, Sept. 1998, pp. 318–327.

[105] R. Zamir, "The rate loss in the Wyner–Ziv problem," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2073–2084, Nov. 1996.

[106] F. Fu and R. Yeung, "On the capacity and error-correcting codes of write-efficient memories," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2299–2314, Nov. 2000.

[107] V. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, "Multiple description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1718–1734, July 2001.