

## MIT Open Access Articles

*Joint Wyner-Ziv/Dirty-Paper Coding by Modulo-Lattice Modulation*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Kochman, Y., and R. Zamir. "Joint Wyner-Ziv/Dirty-Paper Coding by Modulo-Lattice Modulation." *Information Theory, IEEE Transactions on* 55.11 (2009): 4878-4889. © 2009 IEEE

**As Published:** <http://dx.doi.org/10.1109/tit.2009.2030488>

**Publisher:** Institute of Electrical and Electronics Engineers

**Persistent URL:** <http://hdl.handle.net/1721.1/52460>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# Joint Wyner–Ziv/Dirty-Paper Coding by Modulo-Lattice Modulation

Yuval Kochman, *Member, IEEE*, and Ram Zamir, *Senior Member, IEEE*

**Abstract**—The combination of source coding with decoder side information (the Wyner–Ziv problem) and channel coding with encoder side information (the Gel’fand–Pinsker problem) can be optimally solved using the separation principle. In this work, we show an alternative scheme for the quadratic-Gaussian case, which merges source and channel coding. This scheme achieves the optimal performance by applying a modulo-lattice modulation to the analog source. Thus, it saves the complexity of quantization and channel decoding, and remains with the task of “shaping” only. Furthermore, for high signal-to-noise ratio (SNR), the scheme approaches the optimal performance using an SNR-independent encoder, thus it proves for this special case the feasibility of universal joint source–channel coding.

**Index Terms**—Analog transmission, broadcast channel, joint source/channel coding, minimum mean-squared error (MMSE) estimation, modulo lattice modulation, unknown signal-to-noise ratio (SNR), writing on dirty paper, Wyner–Ziv (WZ) problem.

## I. INTRODUCTION

CONSIDER the quadratic-Gaussian joint source/channel coding problem for the Wyner–Ziv (WZ) source [1] and Gel’fand–Pinsker channel [2], as depicted in Fig. 1. In the Wyner–Ziv setup, the source is jointly distributed with some side information (SI) known at the decoder. In the Gaussian case, the WZ-source sequence  $S_k$  is given by

$$S_k = Q_k + J_k \quad (1)$$

where the unknown source part,  $Q_k$ , is Gaussian independent and identically distributed (i.i.d.) with variance  $\sigma_Q^2$ , while  $J_k$  is an arbitrary SI sequence known at the decoder. In the Gel’fand–Pinsker setup, the channel transition distribution depends on a state that serves as encoder SI. In the Gaussian case, known as the dirty-paper channel (DPC) [3], the DPC output,  $Y_k$ , is given by

$$Y_k = X_k + Z_k + I_k \quad (2)$$

Manuscript received January 21, 2008; revised December 12, 2008. Current version published October 21, 2009. This work was supported by the Israeli Science Foundation (ISF) under Grant 1259/07, and by the Advanced Communication Center (ACC). The work of the first author was also supported by a fellowship of the Yitzhak and Chaya Weinstein Research Institute for Signal Processing at Tel-Aviv University. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Seattle, WA, July 2006.

Y. Kochman was with the Department Electrical Engineering–Systems, Tel-Aviv University, Tel-Aviv, Israel. He is now with the Signals, Information and Algorithms Laboratory, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139 USA (e-mail: yuvalko@mit.edu).

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

Communicated by H.-A. Loeliger, Associate Editor for Coding Techniques.

Color version of Figure 6 in this paper is available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2009.2030488

where  $X_k$  is the channel input, the unknown channel noise,  $Z_k$ , is Gaussian i.i.d. with variance  $N$ , while  $I_k$  is an arbitrary interference, known at the encoder. When referring to  $I_k$  and  $J_k$ , we use the terms interference and SI interchangeably, since they may be seen either as external components added to the source and to the channel noise, or as known parts of these entities. From here onward we use the bold notation to denote  $K$ -dimensional vectors, i.e.,

$$\mathbf{X} = [X_1, \dots, X_k, \dots, X_K].$$

The sequences  $\mathbf{Q}$ ,  $\mathbf{J}$ ,  $\mathbf{Z}$ , and  $\mathbf{I}$  are all mutually independent, hence the channel noise  $\mathbf{Z}$  is independent of the channel input sequence  $\mathbf{X}$ . The encoder is some function of the source vector that may depend on the channel SI vector as well

$$\mathbf{X} = f(\mathbf{S}, \mathbf{I}) \quad (3)$$

and must obey the power constraint

$$\frac{1}{K} E\{\|\mathbf{X}\|^2\} \leq P \quad (4)$$

where  $\|\cdot\|$  denotes the Euclidean norm. The decoder is some function of the channel output vector that may depend on the source SI vector as well

$$\hat{\mathbf{S}} = g(\mathbf{Y}, \mathbf{J}) \quad (5)$$

and the reconstruction quality performance criterion is the mean-squared error (MSE):

$$D = \frac{1}{K} E\{\|\hat{\mathbf{S}} - \mathbf{S}\|^2\}. \quad (6)$$

The setup of Fig. 1 described above is a special case of the joint WZ-source and Gel’fand–Pinsker channel setting. Thus, by Merhav and Shamai [4], Shannon’s separation principle holds. Therefore, a combination of optimal source and channel codes can approach the optimum distortion  $D^{\text{opt}}$ , satisfying

$$R_{\text{WZ}}(D^{\text{opt}}) = C_{\text{DPC}} \quad (7)$$

where  $R_{\text{WZ}}(D)$  is the WZ-source rate–distortion function and  $C_{\text{DPC}}$  is the DPC capacity. However, the optimality of “digital” separation-based schemes comes at the price of large delay and complexity. Moreover, they suffer from lack of robustness: if the channel signal-to-noise ratio (SNR) turns out to be lower than expected, the resulting distortion may be very large, while if the SNR is higher than expected, there is no improvement in the distortion [5], [6].

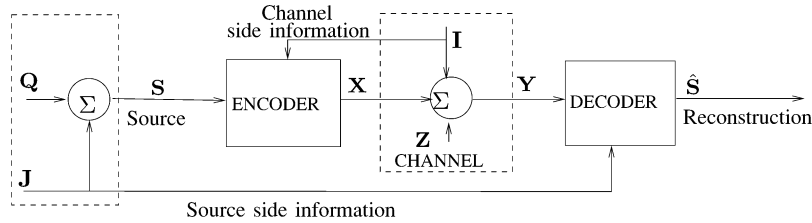


Fig. 1. The Wyner–Ziv/dirty-paper coding problem.

In the special case of white Gaussian source and channel without SI ( $\mathbf{I} = \mathbf{J} = \mathbf{0}$ ), it is well known that analog transmission is optimal [7]. In that case, the encoding and decoding functions

$$\begin{aligned} X_k &= \beta S_k \\ \hat{S}_k &= \frac{\alpha}{\beta} Y_k \end{aligned} \quad (8)$$

are mere scalar factors, where  $\beta$  is a “zoom-in” factor chosen to satisfy the channel power constraint and  $\alpha$  is the channel minimum mean-squared error (MMSE) coefficient, also known as the Wiener coefficient. This scheme achieves the optimal distortion (7) while having low complexity (two multiplications per sample), zero delay, and *full robustness*: only the receiver needs to know the channel SNR, while the transmitter is completely ignorant of that. Such a perfect matching of the source to the channel, which allows scalar/analog coding, only occurs under very special conditions [8].

In the quadratic-Gaussian setting, if both SI components  $\mathbf{I}$  and  $\mathbf{J}$  were available in both ends, one could “eliminate” these components by subtracting  $\mathbf{J}$  at the encoder, then using analog transmission (8) for the source unknown part  $\mathbf{Q}$ , and then at the decoder subtracting  $\mathbf{I}$  and adding again  $\mathbf{J}$ . In the Wyner–Ziv/dirty-paper problem we cannot use such simple elimination, since subtracting  $\mathbf{I}$  at the encoder or subtracting  $\mathbf{J}$  at the decoder inflict a power loss. In fact, scalar/analog coding is *not* feasible in this case [4]. Interestingly, still  $R_{\text{WZ}}(D)$  is just the Gaussian rate–distortion function for the unknown source part  $\mathbf{Q}$  [9], while  $C_{\text{DPC}}$  is just the additive white Gaussian noise (AWGN) capacity for the channel noise  $\mathbf{Z}$  [3]. We see, then, that perfect interference cancellation is required, though it is not achievable by analog coding.

In this work, we propose a scheme for the joint Wyner–Ziv/dirty-paper problem that takes a middle path, i.e., a “semi-analog” solution. Like the digital solution, it is optimal (in the sense of (7)) for any fixed SNR. At the same time, it partially gains the complexity advantages of analog transmission, as well as its robustness: it allows a good compromise between the performance at different SNRs, and becomes SNR-independent at the limit of high SNR. This last result proves that joint source–channel coding can be asymptotically universal with respect to the channel SNR—a special case of the long-standing problem of universal joint source–channel coding.

The scheme we present subtracts the channel interference  $\mathbf{I}$  at the encoder, then uses again subtraction of the source known part  $\mathbf{J}$  at the decoder; while, as explained above, these operations by themselves cannot achieve optimality—we show that the use of modulo-lattice arithmetic at both ends of the scheme

achieves an *equivalent single-letter channel* with  $\mathbf{I} = \mathbf{J} = \mathbf{0}$ . Since the processing is applied to the analog signal, without using any information-bearing code, we call this approach *modulo-lattice modulation* (MLM).

Modulo-lattice codes were suggested as a tool for SI source and channel problems; see [10], [11], where a lattice is used for shaping of a digital code (which may itself have a lattice structure as well, yielding a nested lattice structure). Modulo-lattice transmission of an analog signal in the WZ setting was first introduced in [12], in the context of joint source/channel coding with bandwidth expansion, i.e., when there are several channel uses per each source sample. Here, we generalize and formalize this approach, and apply it to SI problems. In a preliminary version of this work [13], we used the MLM scheme as a building block in *analog matching* of colored sources to colored channels. Later, Wilson *et al.* [14], [15] used transmission of an analog signal modulo a *random* code to arrive at similar results. Recently, MLM was used in network settings for computation over the Gaussian multiple-access channel [16] or for coding for the colored Gaussian relay network [17].

The rest of the paper is organized as follows: In Section II, we present preliminaries about multidimensional lattices, and discuss the existence of lattices that are asymptotically suitable for joint WZ/DPC coding. In Section III, we present the joint WZ/DPC scheme and prove its optimality. In Section IV, we examine the scheme in an unknown SNR setting and show its asymptotic robustness. Finally, Section V discusses complexity reduction issues.

## II. BACKGROUND: GOOD SHAPING LATTICES FOR ANALOG TRANSMISSION

Before we present the scheme, we need some definitions and results concerning multidimensional lattices. Let  $\Lambda$  be a  $K$ -dimensional lattice, defined by the generator matrix  $G \in \mathbb{R}^{K \times K}$ . The lattice includes all points  $\{\mathbf{l} = G \cdot \mathbf{i} : \mathbf{i} \in \mathbb{Z}^K\}$  where  $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ . The nearest neighbor quantizer associated with  $\Lambda$  is defined by

$$Q(\mathbf{x}) = \arg \min_{\mathbf{l} \in \Lambda} \|\mathbf{x} - \mathbf{l}\|$$

where  $\|\cdot\|$  denotes the Euclidian norm and ties are broken in a systematic manner. Let the basic Voronoi cell of  $\Lambda$  be

$$\mathcal{V}_0 = \{\mathbf{x} : Q(\mathbf{x}) = \mathbf{0}\}.$$

The second moment of a lattice is given by the variance of a uniform distribution over the basic Voronoi cell

$$\sigma^2(\Lambda) = \frac{1}{K} \cdot \frac{\int_{\mathcal{V}_0} \|\mathbf{x}\|^2 d\mathbf{x}}{\int_{\mathcal{V}_0} d\mathbf{x}}. \quad (9)$$

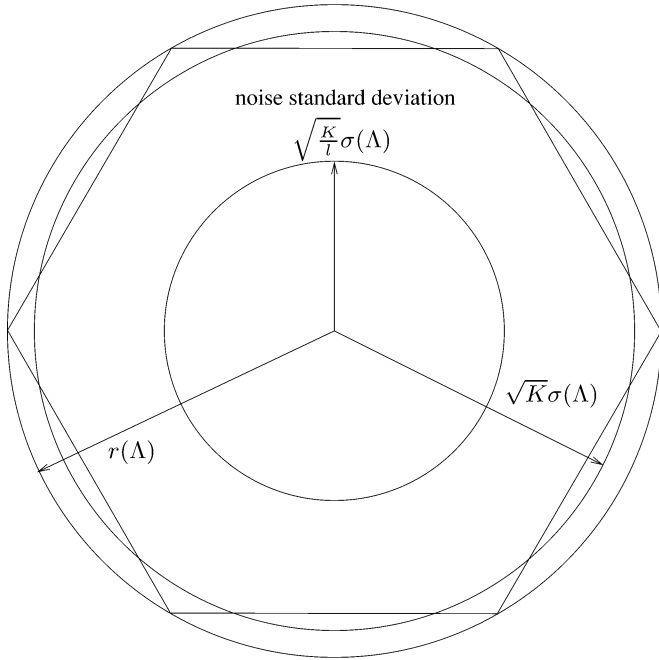


Fig. 2. Lattice covering radius, lattice second moment, and Gaussian noise, following (12) and (14).

The modulo-lattice operation is defined by

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

By definition, this operation satisfies the “distributive law”

$$[\mathbf{x} \bmod \Lambda + \mathbf{y}] \bmod \Lambda = [\mathbf{x} + \mathbf{y}] \bmod \Lambda. \quad (10)$$

The covering radius of a lattice is given by

$$r(\Lambda) = \max_{\mathbf{x} \in \mathcal{V}_0} \|\mathbf{x}\|. \quad (11)$$

For a dither vector  $\mathbf{d}$ , the dithered modulo-lattice operation is

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

If the dither vector  $\mathbf{D}$  is independent of  $\mathbf{x}$  and uniformly distributed over the basic Voronoi cell  $\mathcal{V}_0$ , then  $\mathbf{Y} = [\mathbf{x} + \mathbf{D}] \bmod \Lambda$  is uniformly distributed over  $\mathcal{V}_0$  as well, and is independent of  $\mathbf{x}$  [18]. Consequently, the second moment of  $\mathbf{Y}$  per element is  $\sigma^2(\Lambda)$ .

The loss factor  $L(\Lambda, p_e)$  of a lattice with respect to (w.r.t.) Gaussian noise at error probability  $p_e$  is defined as follows. Let  $\mathbf{Z}$  be a zero-mean Gaussian i.i.d. vector with element variance equal to the lattice second moment  $\sigma^2(\Lambda)$ . Then

$$L(\Lambda, p_e) = \min \left\{ l : \Pr \left\{ \frac{\mathbf{Z}}{\sqrt{l}} \notin \mathcal{V}_0 \right\} \leq p_e \right\}. \quad (12)$$

See Fig. 2 for graphical representation. This factor is, in fact, the product of two better known quantities, the lattice normalized second moment [19] and volume-to-noise ratio [20], see Appendix I. For small enough  $p_e$ , it is at least one. By [21, Theorem 5], there exists a sequence of lattices (simultaneously good for source and channel coding; see more on this in Appendix I)

which possesses a vanishing loss at the limit of high dimension, i.e.,

$$\lim_{p_e \rightarrow 0} \lim_{K \rightarrow \infty} L(\Lambda_K, p_e) = 1. \quad (13)$$

Moreover, there exists a sequence of such lattices that is also good for covering, i.e., defining

$$\tilde{L}(\Lambda) = \frac{r^2(\Lambda)}{K \cdot \sigma^2(\Lambda)} \quad (14)$$

where  $r(\Lambda)$  was defined in (11), the sequence also satisfies:<sup>1</sup>  $\lim_{K \rightarrow \infty} \tilde{L}(\Lambda_K) = 1$ . However, for this work we need a slightly modified result, which allows the replacement of the Gaussian noise by a combination of Gaussian and “self-noise” components. To that end, we define for any  $0 \leq \alpha \leq 1$  the  $\alpha$ -mixture noise as

$$\mathbf{Z}_\alpha = \sqrt{1 - (1 - \alpha)^2} \mathbf{W} - (1 - \alpha) \mathbf{D}$$

where  $\mathbf{W}$  is Gaussian i.i.d. with element variance  $\sigma^2(\Lambda)$ , and  $\mathbf{D}$  is uniform over  $\mathcal{V}_0$  and independent of  $\mathbf{W}$ . Note that since  $\frac{1}{K} \|\mathbf{D}\|^2 = \sigma^2(\Lambda)$ , the resulting mixture also has average per-element variance  $\sigma^2(\Lambda)$ . We redefine the loss factor w.r.t. this mixture noise as

$$L(\Lambda, p_e, \alpha) = \min \left\{ l : \Pr \left\{ \frac{\mathbf{Z}_\alpha}{\sqrt{l}} \notin \mathcal{V}_0 \right\} \leq p_e \right\}. \quad (15)$$

Note that this definition reduces to (12) for  $\alpha = 1$ . Using this definition, we have the following, which is a direct consequence of [22], as explained in Appendix I.

**Proposition 1: (Existence of good lattices)** For any error probability  $p_e > 0$ , and for any  $0 \leq \alpha \leq 1$ , there exists a sequence of  $K$ -dimensional lattices  $\Lambda_K$  satisfying

$$\lim_{p_e \rightarrow 0} \lim_{K \rightarrow \infty} L(\Lambda_K, p_e, \alpha) = 1 \quad (16)$$

and

$$\lim_{K \rightarrow \infty} \tilde{L}(\Lambda_K) = 1. \quad (17)$$

Note that since by definition,  $L(\Lambda_K, p_e, \alpha)$  is nonincreasing in  $p_e$ , it follows that for any  $p_e > 0$  this sequence of lattices satisfies

$$\limsup_{K \rightarrow \infty} L(\Lambda_K, p_e, \alpha) \leq 1. \quad (18)$$

### III. MODULO-LATTICE WZ/DPC CODING

We now present the modulo-lattice joint source/channel coding scheme for the SI problem of Fig. 1, based upon good lattices in the sense discussed above. As explained in the Introduction, the quadratic-Gaussian rate-distortion function (RDF) of the WZ source (1) is equal to the RDF of the source  $Q_k$  (without the known part  $J_k$ ), given by

$$R_{\text{WZ}}(D) = \frac{1}{2} \log \frac{\sigma_Q^2}{D}. \quad (19)$$

<sup>1</sup>Note that  $\tilde{L}(\Lambda_K) \geq 1$ , since the second moment is the expected value of the squared norm of points which lie inside a ball of radius  $r(\Lambda)$ .

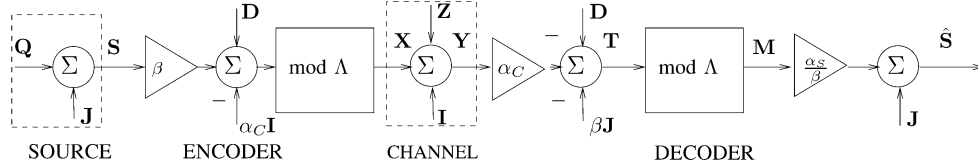


Fig. 3. Analog Wyner-Ziv/dirty-paper coding scheme:  $\mathbf{S}$  = source,  $\hat{\mathbf{S}}$  = reconstruction,  $\mathbf{Z}$  = channel noise,  $\mathbf{I}$  = interference known at the encoder,  $\mathbf{J}$  = source component known at the decoder,  $\mathbf{D}$  = dither.

Similarly, the capacity of the Gaussian DPC (2) is equal to the AWGN capacity (without the interference  $I_k$ )

$$C_{\text{DPC}} = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right). \quad (20)$$

Recalling that the separation principle holds for this problem [4], the optimum distortion (7) is thus given by

$$D^{\text{opt}} = \frac{N}{P+N} \sigma_Q^2 \quad (21)$$

which equals the optimum distortion for the source  $Q_k$  passed through a channel with noise  $Z_k$ , without SI.

We show how to approach  $D^{\text{opt}}$  using the joint source/channel coding scheme depicted in Fig. 3. In this scheme, the  $K$ -dimensional encoding and decoding functions (3), (5) are given by

$$\mathbf{X} = [\beta \mathbf{S} + \mathbf{D} - \alpha_C \mathbf{I}] \bmod \Lambda \quad (22a)$$

$$\hat{\mathbf{S}} = \frac{\alpha_S}{\beta} \{ [\alpha_C \mathbf{Y} - \mathbf{D} - \beta \mathbf{J}] \bmod \Lambda \} + \mathbf{J} \quad (22b)$$

respectively, where the second moment (9) of the lattice is  $\sigma^2(\Lambda) = P$ , and the dither vector  $\mathbf{D}$  is uniformly distributed over  $\mathcal{V}_0$  and independent of the source and of the channel. The channel power constraint is satisfied automatically by the properties of dithered lattice quantization discussed in Section II. The factors  $\alpha_S$ ,  $\alpha_C$ , and  $\beta$  will be chosen in the sequel. For optimum performance,  $\beta$  which is used at the encoder will depend upon the variance of the source unknown part, while  $\alpha_C$  used at the decoder will depend upon the channel SNR. It is assumed, then, that both the encoder and the decoder have full knowledge of the source and channel statistics; we will break with this assumption in the next section.

The following theorem gives the performance of the scheme, in terms of the lattice parameters  $L(\cdot, \cdot, \cdot)$  in (15) and in  $\tilde{L}(\cdot)$  (14), and the quantities

$$\alpha_0 \triangleq \frac{P}{P+N} \quad (23a)$$

$$\tilde{\alpha} \triangleq \max \left( \alpha_0 - \frac{L(\Lambda, p_e, \alpha_0) - 1}{L(\Lambda, p_e, \alpha_0)}, 0 \right). \quad (23b)$$

We will also use these quantities in the sequel to specify the choice of factors  $\alpha_S$ ,  $\alpha_C$ , and  $\beta$ .

**Theorem 1: (Performance with any lattice)** For any lattice  $\Lambda$  and any error probability  $p_e > 0$ , there exists a choice of factors  $\alpha_C, \alpha_S, \beta$  such that the system of (22) (depicted in Fig. 3) satisfies

$$D \leq L(\Lambda, p_e, \alpha_0) D^{\text{opt}} + p_e D^{\text{max}}$$

where the optimum distortion  $D^{\text{opt}}$  was defined in (21), and

$$D^{\text{max}} = 4\sigma_Q^2 \left( 1 + \frac{\tilde{L}(\Lambda)}{\tilde{\alpha}} \right). \quad (24)$$

Note that for a good enough lattice  $\tilde{\alpha}$  is positive, thus  $D^{\text{max}}$  is finite.

We prove this theorem in the sequel. As a direct corollary from it, taking  $p_e$  to be an arbitrarily small probability and using the properties of good lattices (17) and (18), we have the following asymptotic optimality result.<sup>2</sup>

**Theorem 2: (Large-dimension optimality)** Let  $D(\Lambda_K)$  be the distortion achievable by the system of (22) with a lattice from a sequence  $\{\Lambda_K\}$  that is simultaneously good for source and channel coding in the sense of Proposition 1. Then for any  $\epsilon > 0$ , there exists a choice of factors  $\alpha_C, \alpha_S$ , and  $\beta$  such that

$$\limsup_{K \rightarrow \infty} D(\Lambda_K) \leq D^{\text{opt}} + \epsilon.$$

For proving Theorem 1 we start with a lemma, showing equivalence in probability to a real-additive noise channel (see Fig. 4(b)). The equivalent additive noise is

$$\mathbf{Z}_{\text{eq}} = \alpha_C \mathbf{Z} - (1 - \alpha_C) \mathbf{X} \quad (25)$$

where  $\mathbf{Z}$  and  $\mathbf{X}$  are the physical channel input and AWGN, respectively. By the properties of the dithered modulo-lattice operation, the physical channel input  $\mathbf{X}$  is uniformly distributed over  $\mathcal{V}_0$  and independent of the source. Thus,  $\mathbf{Z}_{\text{eq}}$  is indeed additive and has per-element variance

$$\sigma_{\text{eq}}^2 = \alpha_C^2 N + (1 - \alpha_C)^2 P. \quad (26)$$

**Lemma 1: (Equivalent additive noise channel)** Fix some  $p_e > 0$ . In the system defined by (1), (2), and (22), the decoder modulo output  $\mathbf{M}$  (see Fig. 3) satisfies

$$\mathbf{M} = \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \text{ with probability at least } (1 - p_e) \quad (27)$$

provided that

$$\beta^2 \sigma_Q^2 + \sigma_{\text{eq}}^2 \leq \frac{P}{L(\Lambda, p_e, \alpha_C)}, \quad (28)$$

where  $\mathbf{Z}_{\text{eq}}$ , defined in (25), is independent of  $\mathbf{Q}$  and  $\mathbf{J}$  and has per-element variance  $\sigma_{\text{eq}}^2$  (26), and  $L(\cdot, \cdot, \cdot)$  was defined in (15).

<sup>2</sup>The explicit derivation of  $D^{\text{max}}$  is not necessary for proving Theorem 2; see Appendix II-B.

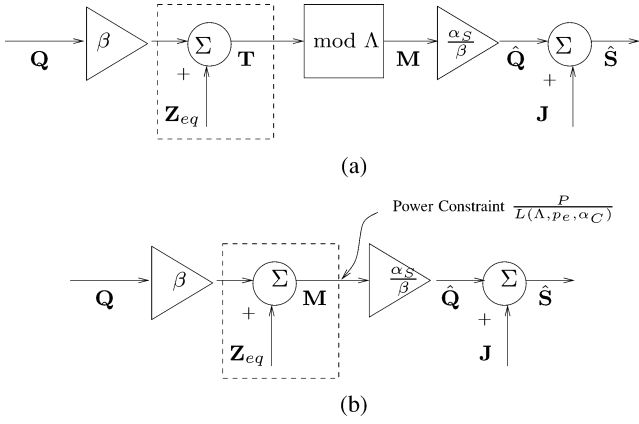


Fig. 4. Equivalent channels for the WZ/WDP coding scheme. (a) Equivalent modulo-lattice channel. (b) Equivalent real-additive noise channel with probability (w.p.)  $(1 - p_e)$ .

*Proof:* We will first prove equivalence to the channel of Fig. 4(a)

$$\mathbf{M} = [\beta\mathbf{Q} + \mathbf{Z}_{\text{eq}}] \bmod \Lambda \quad (29)$$

where  $\mathbf{Z}_{\text{eq}}$  was defined in (25). To that end, let  $\mathbf{T} = \alpha_C \mathbf{Y} - \mathbf{D} - \beta \mathbf{J}$  denote the input of the decoder modulo operation (see (22b) and Fig. 3). Combine (2) and (22a) to assert

$$\begin{aligned} \mathbf{T} &= \alpha_C (\mathbf{X} + \mathbf{Z} + \mathbf{I}) - \mathbf{D} - \beta \mathbf{J} \\ &= [\beta \mathbf{S} + \mathbf{D} - \alpha_C \mathbf{I}] \bmod \Lambda + \mathbf{Z}_{\text{eq}} + \alpha_C \mathbf{I} - \mathbf{D} - \beta \mathbf{J}. \end{aligned}$$

Now, using (1) and the “distributive law” (10)

$$\mathbf{T} \bmod \Lambda = [\beta \mathbf{Q} + \mathbf{Z}_{\text{eq}}] \bmod \Lambda$$

and since  $\mathbf{T} = \mathbf{M} \bmod \Lambda$ , we establish (29). Now we note that

$$\begin{aligned} \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} &= \beta \mathbf{Q} + \alpha_C \mathbf{Z} - (1 - \alpha_C) \mathbf{X} \\ &\triangleq \sqrt{1 - (1 - \alpha_C)^2 \mathbf{W} - (1 - \alpha_C) \mathbf{X}} \end{aligned}$$

where  $\mathbf{W}$  is Gaussian i.i.d.,  $\mathbf{X}$  is uniform over the basic cell  $\mathcal{V}_0$  of the lattice  $\Lambda$ , and the total variance (per element) is given by the left-hand side (LHS) of (28). By the definition of  $L(\cdot, \cdot, \cdot)$ , we have that

$$\mathbf{T} = \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \in \mathcal{V}_0 \quad (30)$$

w.p. at least  $(1 - p_e)$ . Substituting this in (29), we get (27).  $\square$

By this lemma, as long as (28) holds, the whole system is equivalent w.p.  $(1 - p_e)$  to the channel depicted in Fig. 4(b)

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{J} + \frac{\alpha_S}{\beta} \mathbf{Z}_{\text{eq}} + \alpha_S \mathbf{Q} \\ &= \mathbf{S} + \frac{\alpha_S}{\beta} \mathbf{Z}_{\text{eq}} - (1 - \alpha_S) \mathbf{Q}. \end{aligned} \quad (31)$$

This channel equivalence holds for any choice of dimension  $K$ , lattice  $\Lambda$ , and factors  $\alpha_C$ ,  $\alpha_S$ , and  $\beta$ , as long as (28) holds.

For the proof of Theorem 1 we make the following choice (using the parameters of (23)):

$$\alpha_C = \alpha_0 \quad (32a)$$

$$\beta^2 = \tilde{\alpha} \frac{P}{\sigma_Q^2} \quad (32b)$$

$$\alpha_S = \frac{\tilde{\alpha} P}{\tilde{\alpha} P + \alpha_0 N}. \quad (32c)$$

It will become evident in the sequel, that  $\alpha_C$  and  $\alpha_S$  are the MMSE (Wiener) coefficients for estimating  $\mathbf{X}$  from  $\mathbf{X} + \mathbf{Z}$  and  $\mathbf{Q}$  from  $\mathbf{Q} + \frac{\mathbf{Z}_{\text{eq}}}{\beta}$ , respectively, while  $\beta$  is the maximum zooming factor that allows to satisfy (28) with equality, whenever possible.

*Proof of Theorem 1:* For calculating the achievable distortion, first note that by the properties of MMSE estimation

$$\sigma_{\text{eq}}^2 = \alpha_C N = \alpha_0 N.$$

Using this, it can be verified that our choice of  $\beta$  satisfies (28), thus (31) holds w.p.  $(1 - p_e)$ . Denoting by  $D^{\text{correct}}$  and  $D^{\text{incorrect}}$  the distortions conditioned on the event that (31) holds or does not hold, respectively, we have

$$\begin{aligned} D &= (1 - p_e) D^{\text{correct}} + p_e D^{\text{incorrect}} \\ &\leq D^{\text{correct}} + p_e D^{\text{incorrect}}. \end{aligned} \quad (33)$$

In Appendix II-A, we prove that the second term satisfies  $D^{\text{incorrect}} \leq D^{\text{max}}$ . For the first term, we have

$$\begin{aligned} D^{\text{correct}} &= \frac{1}{K} E \left\{ \left\| \frac{\alpha_S}{\beta} \mathbf{Z}_{\text{eq}} - (1 - \alpha_S) \mathbf{Q} \right\|^2 \right\} \\ &\stackrel{(a)}{=} \alpha_S \frac{\sigma_{\text{eq}}^2}{\beta^2} = \frac{\sigma_Q^2 \sigma_{\text{eq}}^2}{\beta^2 \sigma_Q^2 + \sigma_{\text{eq}}^2} \\ &= \frac{D^{\text{opt}}}{1 - \alpha_0 + \tilde{\alpha}} \\ &= \min(L(\Lambda, p_e, \alpha_C) D^{\text{opt}}, \sigma_Q^2) \\ &\leq L(\Lambda, p_e, \alpha_C) D^{\text{opt}} \end{aligned}$$

where (a) stems from the properties of MMSE estimation. Summing both conditional distortion terms, the proof is completed  $\square$

As mentioned in the Introduction, a recent work [15] derives a similar asymptotic result, replacing the shaping lattice of our scheme by a *random* shaping code. Such a choice is less restrictive since it is not tied to the properties of good Euclidean lattices, though it leads to higher complexity due to the lack of structure. The use of lattices also allows analysis in finite dimension as in Theorem 1 and in Section V. Furthermore, structure is essential in network joint source/channel settings; see e.g., [16]. Finally, the dithered lattice formulation allows to treat any interference signals (not necessarily Gaussian), see Remark 2 in the sequel.

We conclude this section by the following remarks, intended to shed more light on the significance of the results above.

1. **Optimal decoding.** The decoder we described is *not* the MMSE estimator of  $\mathbf{S}$  from  $\mathbf{Y}$ . This is for two reasons: First, the decoder ignores the probability of incorrect lattice decoding. Second, since  $\mathbf{Z}_{\text{eq}}$  is not Gaussian, the modulo-lattice operation w.r.t. the lattice Voronoi cells is not equivalent to maximum-likelihood estimation of the lattice point (see [22] for a similar discussion in the context of channel coding). Consequently, for any finite dimension the decoder can be improved. We shall discuss further the issue of working with finite-dimensional lattices in Section V.

2. **Universality w.r.t.  $\mathbf{I}$  and  $\mathbf{J}$ .** None of the scheme parameters depend upon the nature of the channel interference  $\mathbf{I}$  and source known part  $\mathbf{J}$ . Consequently, the scheme is adequate for arbitrary (individual) sequences. This has no effect on the asymptotic performance of Theorem 2, but for finite-dimensional lattices the scheme may be improved, e.g., if the interference signals are known to be Gaussian with low enough variance. A similar argument also holds when the source or channel statistics is not perfectly known, see Section IV in the sequel.

3. **Non-Gaussian Setting.** If the source unknown part  $\mathbf{Q}$  or the channel noise  $\mathbf{Z}$  are not Gaussian, the optimum quadratic-Gaussian distortion  $D^{\text{opt}}$  may still be approached using the MLM scheme, though it is no longer the optimum performance for the given source and channel.

4. **Asymptotic choice of parameters.** In the limiting case, where  $L(\Lambda, p_e, \alpha_0) \rightarrow 1$ , we have that  $\alpha_S = \tilde{\alpha} = \alpha_0$  in (32), i.e., the choice of parameters approaches

$$\alpha_C = \alpha_S = \frac{P}{P+N} = \alpha_0 \quad (34a)$$

$$\beta^2 = \alpha_0 \frac{P}{\sigma_Q^2}. \quad (34b)$$

5. **Properties of the equivalent additive-noise channel.** With high probability, we have the equivalent real-additive noise channel of (31) and Fig. 4(b). This differs from the modulo-additivity of the lattice strategies of [22], [23]: Closeness of point under a modulo arithmetic does not mean closeness under a difference distortion measure. The condition (28) forms an output-power constraint: No matter what the noise level of the channel is, its output must have a power of no more than  $P$ ; this replaces the input-power constraint of the physical channel. Furthermore, by the lattice quantization noise properties [18], the “self-noise” component  $(1 - \alpha_C)\mathbf{X}$  in (25) is asymptotically Gaussian i.i.d., and consequently so is the equivalent noise  $\mathbf{Z}_{\text{eq}}$ . Thus, the additive equivalent channel (31) is asymptotically an *output-power constrained AWGN channel*.

6. **Noise margin.** The additivity in (31) is achieved through leaving a “noise margin.” The condition (28) means that the sum of the (scaled) unknown source part and equivalent noise should “fit into” the lattice cell (see (30)). Consequently, the unknown source part  $\mathbf{Q}$  is inflated to a power strictly smaller than the lattice power  $P$ . In the limit of infinite dimension, when the choice of parameters becomes (34), this power becomes  $\beta^2 \sigma_Q^2 = \alpha_0 P$ . In comparison, it is shown in [23] that in a lattice solution to

a digital SI problem, if the information-bearing code (fine lattice) occupies a portion of power  $\gamma P$  with any  $\alpha_0 \leq \gamma \leq 1$ , capacity is achieved.<sup>3</sup> This freedom, however, has to do with the modulo-additivity of the equivalent channel; in our joint source/channel setting, necessarily  $\gamma = \alpha_0$ .

7. **Comparison with analog transmission.** Finally, consider the similarity between our asymptotic AWGN channel and the optimal analog transmission scheme without SI (8): Since we have “eliminated from the picture” the SI components  $\mathbf{I}$  and  $\mathbf{J}$ , we are left with the transmission of the source unknown component through an equivalent additive noise channel. As mentioned above, the unknown source part  $\mathbf{Q}$  is only adjusted to power  $\alpha_0 P$  (in the limit of high dimension), while in (8) the source  $\mathbf{S}$  is adjusted to power  $P$ ; but since the equivalent noise  $\mathbf{Z}_{\text{eq}}$  has variance  $\alpha_0 N$ , the equivalent channel has an SNR of  $P/N$ , just as the physical channel.

#### IV. TRANSMISSION UNDER UNCERTAINTY CONDITIONS

We now turn to the case where either the variance of the channel noise  $N$ , or the variance of the source unknown part  $\sigma_Q^2$ , are unknown at the encoder.<sup>4</sup> In Section IV-A we assume that  $\sigma_Q^2$  is known at both sides, but the channel SNR is unknown at the encoder. We show that in the limit of high SNR, optimality can still be approached. In Section IV-B, we address the general SNR case, as well as the case of unknown  $\sigma_Q^2$ ; for that, we adopt an alternative broadcast-channel point of view.

For convenience, we present our results in terms of the channel SNR

$$\text{SNR} \triangleq \frac{P}{N} \quad (35)$$

and the achieved signal-to-distortion ratio (SDR)

$$\text{SDR} \triangleq \frac{\sigma_Q^2}{D}. \quad (36)$$

Denoting the theoretically optimal SDR as  $\text{SDR}^{\text{opt}}$ , (21) becomes

$$\text{SDR}^{\text{opt}} = 1 + \text{SNR}. \quad (37)$$

Our achievability results in this section are based upon application of the MLM scheme, generally with a suboptimal choice of parameters due to the uncertainty. We only bring asymptotic results, using high-dimensional “good” lattices. We present, then, the following lemma, using the definition

$$\beta_0^2 = \frac{P}{\sigma_Q^2}. \quad (38)$$

*Lemma 2:* Let  $\text{SDR}(\Lambda_K)$  be the distortion achievable by the system of (22) with a lattice from a sequence  $\{\Lambda_K\}$  that is good

<sup>3</sup>In [24], a similar observation is made, and a code of power  $\alpha_0 P$  is presented as a preferred choice, since it allows easy iterative decoding between the information-bearing code and the coarse lattice.

<sup>4</sup>We do not treat uncertainty at the decoder, since  $N$  can be learnt, while the major insight into the matter of unknown  $\sigma_Q^2$  is gained already by assuming uncertainty at the encoder.

in the sense of Proposition 1. For any choice of factors  $\alpha_C$ ,  $\alpha_S$ , and  $\beta$

$$\begin{aligned} & \liminf_{K \rightarrow \infty} \text{SDR}(\Lambda_K) \\ & \geq \beta^2 \left\{ (1 - \alpha_S)^2 \beta^2 + \alpha_S^2 \left[ \frac{\alpha_C^2}{\text{SNR}} + (1 - \alpha_C)^2 \right] \beta_0^2 \right\}^{-1} \end{aligned} \quad (39)$$

provided that

$$\frac{\beta^2}{\beta_0^2} + \frac{\alpha_C^2}{\text{SNR}} + (1 - \alpha_C)^2 < 1. \quad (40)$$

*Proof:* This is a direct application of Lemma 1 and of (18). First we fix some  $p_e > 0$ , and note that (40) is equivalent to (28). The SDR of the equivalent channel (31), at the limit  $L(\Lambda_K, p_e, \alpha_C) \rightarrow 1$  is then given by (39). Then for  $p_e \rightarrow 0$ , the effect of decoding errors vanishes, as shown in Appendix II-B.  $\square$

Note, that by substituting the asymptotically optimal choice of parameters (34) in (39), the limit becomes  $\text{SDR}^{\text{opt}}$ .

#### A. Asymptotic Robustness for Unknown SNR

Imagine that we know that  $\text{SNR} \geq \text{SNR}_0$ , for some specific  $\text{SNR}_0$ , and that  $\sigma_Q^2$  is known. Suppose that we set the scheme parameters such that the correct decoding condition (40) holds for  $\text{SNR}_0$ . Since the variance of the equivalent noise can only decrease with the SNR, correct lattice decoding will hold for any  $\text{SNR} \geq \text{SNR}_0$ , and we are left with the equivalent additive-noise channel where the resulting SDR is a strictly decreasing function of the SNR. We use this observation to derive an asymptotic result, showing that for high SNR, a *single* encoder can approach optimality simultaneously for all actual SNR. To that end, we replace the choice given in (32), which leads to optimality at one SNR, by the high-SNR choice  $\alpha_C = \alpha_S = 1$ , where  $\beta$  is chosen to ensure correct decoding even at the minimal  $\text{SNR}_0$ .

**Theorem 3: (Robustness at high SNR)** For a sequence of encoding–decoding schemes indexed by  $K = 1, 2, \dots$  and a given SNR, let  $\text{SDR}_K(\text{SNR})$  denote the SDR (36) achieved by the  $K$ th scheme. We further denote as  $\text{SDR}^{\text{opt}}(\text{SNR})$  the optimum performance (37). For any  $\epsilon > 0$ , there exists a fixed (SNR-independent) sequence of schemes that satisfies

$$\liminf_{K \rightarrow \infty} \text{SDR}_K(\text{SNR}) \geq (1 - \epsilon) \text{SDR}^{\text{opt}}(\text{SNR}) \quad (41)$$

for all sufficiently large (but finite) SNR.

In other words, the theorem states that (41) holds for all  $\text{SNR} \geq \text{SNR}_0(\epsilon)$ , where  $\text{SNR}_0(\epsilon)$  is finite for all  $\epsilon > 0$ . A limit of a *sequence* of schemes is needed in the theorem, rather than a single scheme, since for any single scheme we have  $p_e > 0$ , thus the effect of incorrect decoding cannot be neglected in the limit  $\text{SNR} \rightarrow \infty$  (meaning that the convergence in Lemma 2 is not uniform). If we restricted our attention to SNRs bounded by some arbitrarily high value, a single scheme would be sufficient.

*Proof:* We use a sequence of MLM schemes with a sequence of good lattices in the sense of Proposition 1. If  $\alpha_C = 1$ , then any

$$\beta^2 < \frac{\text{SNR}_0 - 1}{\text{SNR}_0} \cdot \beta_0^2$$

where  $\beta_0$  is given by (38), satisfies the condition (40) for  $\text{SNR}_0$ , thus for any  $\text{SNR} \geq \text{SNR}_0$ . Here we assume that  $\text{SNR}_0 > 1$ , without loss of generality (w.l.o.g.) since we can always choose  $\text{SNR}_0(\epsilon)$  of the theorem accordingly. With this choice and with  $\alpha_S = 1$ , we have by Lemma 2 that the SDR may approach (for any  $\text{SNR} \geq \text{SNR}_0$ )

$$\begin{aligned} \frac{\beta^2}{\beta_0^2} \text{SNR} &= \frac{\text{SNR}_0 - 1}{\text{SNR}_0} \cdot \text{SNR} \\ &= \frac{\text{SNR}_0 - 1}{\text{SNR}_0} \cdot \frac{\text{SNR}}{\text{SNR} + 1} \cdot \text{SDR}^{\text{opt}} \\ &\geq \frac{\text{SNR}_0 - 1}{\text{SNR}_0 + 1} \cdot \text{SDR}^{\text{opt}}. \end{aligned}$$

Now take  $\epsilon = 1 - \frac{\text{SNR}_0 - 1}{\text{SNR}_0 + 1}$ . Since  $\lim_{\text{SNR}_0 \rightarrow \infty} \epsilon = 0$ , one may find  $\text{SNR}_0$  for any  $\epsilon > 0$  as required  $\square$

Note that we have here also a fixed decoder; if we are only interested in a fixed encoder we can adjust  $\alpha_S$  at the decoder and reduce the margin from optimality.

#### B. Joint Source/Channel Broadcasting

Abandoning the high SNR assumption, we can no longer simultaneously approach the optimal performance (37) for multiple SNRs. However, in many cases we can still do better than a separation-based scheme. In order to demonstrate that, we choose to alternate our view to a *broadcast* scenario, where the same source needs to be transmitted to multiple decoders, each one with different conditions; yet all the decoders share the same channel interference  $\mathbf{I}$ , see Fig. 5. The variation of the source SI component  $\mathbf{J}$  between decoders means that the source has two decompositions

$$\mathbf{S} = \mathbf{Q}_1 + \mathbf{J}_1 = \mathbf{Q}_2 + \mathbf{J}_2 \quad (42)$$

and we define the per-element variances of the unknown parts as  $\sigma_1^2$  and  $\sigma_2^2$ , respectively. Note that this variation does not imply any uncertainty from the point of view of the MLM encoder, as long as  $\sigma_1^2 = \sigma_2^2$ ; see [25] for a similar observation in the context of source coding. We denote the SNRs at the decoders as  $\text{SNR}_1 \leq \text{SNR}_2$ , and find achievable corresponding SDR  $\{\text{SDR}_1, \text{SDR}_2\}$  pairs. It will become evident from the exposition, that this approach is also good for a continuum of possible SNRs.

We start from the case  $\sigma_1^2 = \sigma_2^2$ , for which we have the following.

**Theorem 4:** In the broadcast WZ/DPC channel of Fig. 5 with  $\sigma_1^2 = \sigma_2^2$ , the signal-to-distortions pair

$$\left\{ 1 + \frac{\bar{\alpha} \cdot \text{SNR}_1}{\alpha_C^2 + (1 - \alpha_C)^2 \text{SNR}_1}, 1 + \frac{\bar{\alpha} \cdot \text{SNR}_2}{\alpha_C^2 + (1 - \alpha_C)^2 \text{SNR}_2} \right\}$$



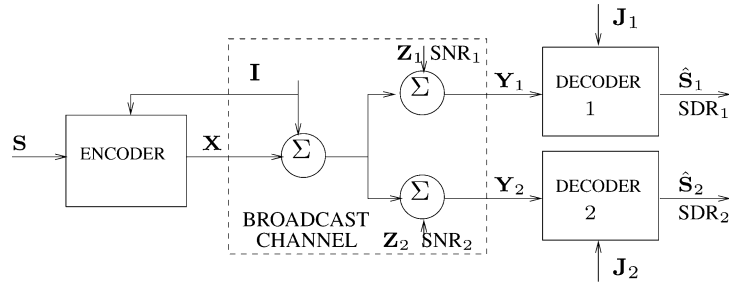


Fig. 5. A broadcast presentation of the uncertainty problem.

where

$$\bar{\alpha} = \alpha_C \left( 2 - \frac{\text{SNR}_1 + 1}{\text{SNR}_1} \alpha_C \right) \quad (43)$$

can be approached for any  $0 < \alpha_C \leq \min(1, \frac{2 \cdot \text{SNR}_1}{1 + \text{SNR}_1})$ . In addition, if there is no channel interference ( $\mathbf{I} = \mathbf{0}$ ), then the pair

$$\left\{ 1 + \text{SNR}_1, 1 + \frac{\text{SNR}_1(1 + \text{SNR}_2)}{1 + \text{SNR}_1} \right\}$$

can be approached as well.

*Proof:* As in the proof of Theorem 3, we use Lemma 2 with a choice of  $\beta$  which allows correct decoding in the lower SNR. For the first part of the theorem, fix any  $\alpha_C$  according to the theorem conditions, and choose any

$$\beta^2 < \bar{\alpha} \frac{P}{\sigma_Q^2}$$

where  $\bar{\alpha}$  was defined in (43), in order to satisfy (40). In each decoder, optimize  $\alpha_S$  in (39) to approach the desired distortion. For the second part of the theorem, if there is no channel interference, the encoder is  $\alpha_C$ -independent, thus each decoder can work with a different  $\alpha_C$  value. We can therefore make the encoder and the first decoder optimal for  $\text{SNR}_1$ , while the second decoder only suffers from the choice of  $\beta$  at the encoder. Again, we substitute in (39) to arrive at the desired result.  $\square$

By standard time-sharing arguments, the achievable SDR regions include the convex hull (in the distortions plane) defined by these points and the trivial  $\{1 + \text{SNR}_1, 1\}$  and  $\{1, 1 + \text{SNR}_2\}$  points. Fig. 6 demonstrates these regions, compared to the ideal (unachievable) region of simultaneous optimality for both SNRs, and the separation-based region achieved by the concatenation of successive-refinement source code (see, e.g., [26]) with broadcast channel code [27] (about the suboptimality of this combination without SI, see, e.g., [28]). It is evident, that in most cases the use of the MLM scheme significantly improves the SDR tradeoff over the performance offered by the separation principle, and that the scheme approaches simultaneous optimality where both SNRs are high, as promised by Theorem 3. Note that, unlike the separation-based approach, the MLM approach also offers reasonable SDRs for intermediate SNRs. Moreover, note that this region is achievable when no assumption is made about the statistics of  $\mathbf{I}$  and  $\mathbf{J}$ . If these interferences are not very strong compared to  $P$  and  $\sigma_Q^2$ , respectively, then one may

further extend the achievable region by allowing some residual interference.

To conclude, we briefly discuss the case where  $\sigma_1^2 \neq \sigma_2^2$ . We define the SDR of each decoder relative to its own variance, and ask what are the achievable SDRs for a pair of SNRs, which may be equal or different. Assume here the simple case, where there is no channel interference, i.e.,  $\mathbf{I} = \mathbf{0}$ . In this case, the encoder only needs to agree upon  $\beta$  with the decoders, thus (by Lemma 2) we may approach for  $n = 1, 2$

$$\text{SDR}_n = 1 + \frac{\beta^2}{\beta_{\text{opt},n}^2} \text{SNR}_n \quad (44)$$

where  $\beta_{\text{opt},n}$  is the optimum choice of  $\beta$  for  $\text{SNR}_n$  according to (34). It follows, that if the two decoders require the same value of  $\beta$ , they may both approach the theoretically optimal distortion. This translates to the optimality condition

$$\sigma_1^2 \frac{1 + \text{SNR}_1}{\text{SNR}_1} = \sigma_2^2 \frac{1 + \text{SNR}_2}{\text{SNR}_2}.$$

This scenario was presented in [29], where simultaneous optimality using hybrid digital/analog schemes was proven under a different condition

$$\frac{\sigma_1^2}{\text{SNR}_1} = \frac{\sigma_2^2}{\text{SNR}_2}.$$

Both conditions reflect the fact that better source conditions (lower  $\sigma_Q^2$ ) can compensate for worse channel conditions (lower SNR). It follows from the difference between the conditions, that for some parameter values the MLM scheme outperforms the approach of [29], thus extending the achievable SDRs region.

## V. DISCUSSION: DELAY AND COMPLEXITY

We have presented the joint source/channel MLM scheme, proven its optimality for the joint WZ/DPC setting with known SNR, and shown its improved robustness over a separation-based scheme. We now discuss the potential complexity and delay advantages of our approach relative to separation-based schemes, first considering the complexity at high dimension and then suggesting a scalar variant.

Consider a separation-based solution, with source and channel encoder/decoder pairs. An optimal channel coding scheme typically consists of two codes: an information-bearing code and a shaping code, both of which require a nearest-neighbor search at the decoder. An optimal source coding scheme also consists of both a quantization code and

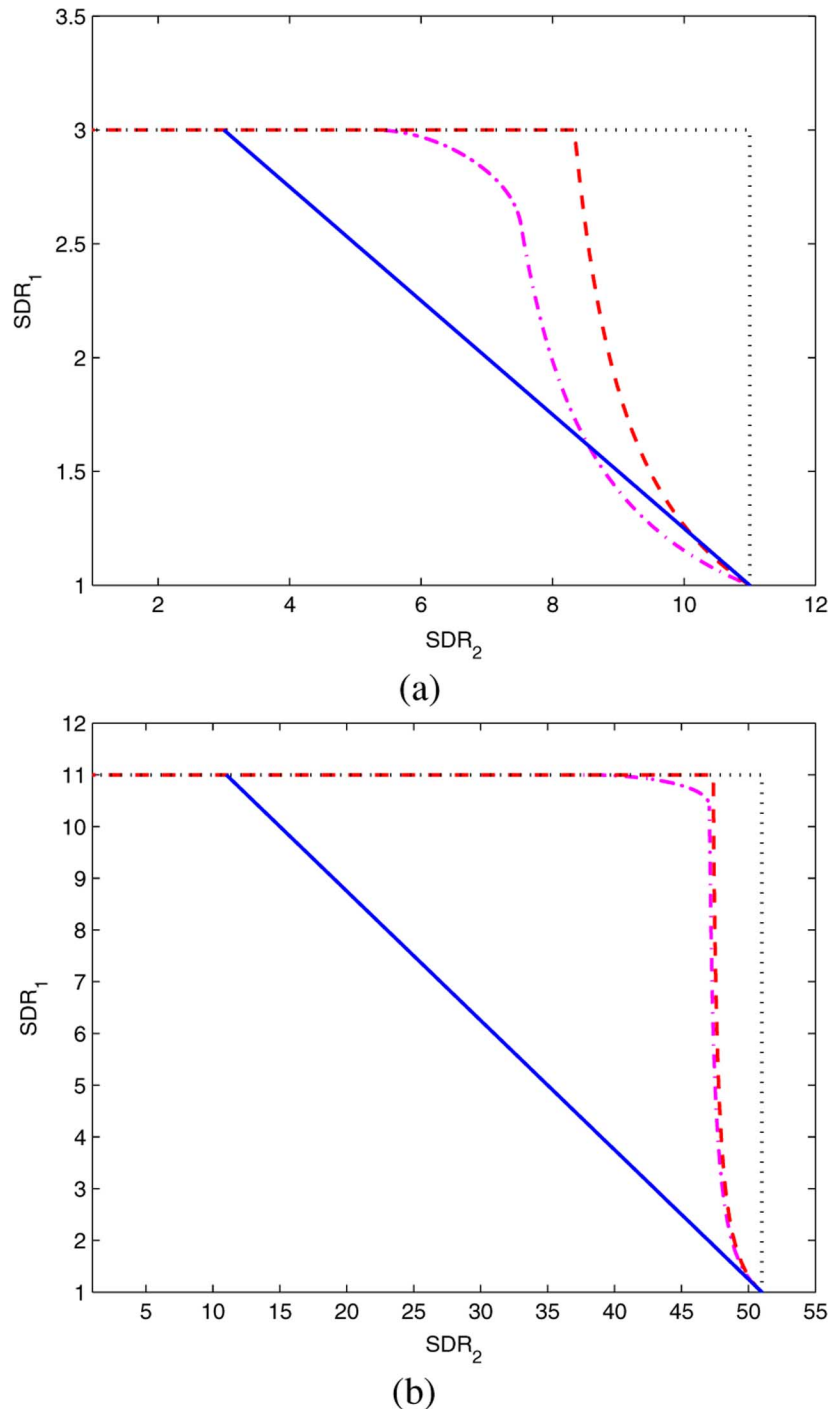


Fig. 6. Broadcast performance. Solid line: Achievable by separation for arbitrary  $\mathbf{I}$  and  $\mathbf{J}$ . Dash-dotted line: Achievable by MLM for arbitrary  $\mathbf{I}$  and  $\mathbf{J}$ . Dashed line: Achievable by MLM for arbitrary  $\mathbf{J}$ , with  $\mathbf{I} = \mathbf{0}$ . Dotted line: Outer bound of ideal matching to both SNRs (achievable by analog transmission when  $\mathbf{I} = \mathbf{J} = \mathbf{0}$ ). (a)  $\text{SNR}_1 = 2$ ,  $\text{SNR}_2 = 10$ , (b)  $\text{SNR}_1 = 10$ ,  $\text{SNR}_2 = 50$ .

a shaping code in order to achieve the full vector quantization gain (see, e.g., [30]), thus, two nearest-neighbor searches are needed at the encoder. The MLM approach omits the information-bearing channel code and the quantization code, and merges the channel and source shaping codes into one. It is convenient to compare this approach with the nested lattices approach to channel and source coding with SI [10], since in that approach both the channel and source information bearing/shaping code pairs are materialized by nested lattices. In comparison, our scheme requires only a single lattice (parallel to the

coarse lattice of nested schemes), and in addition the source and channel lattices collapse into a single one.

There is a price to pay, however: For the WZ problem, the coarse lattice should be good for channel coding, while for the writing on dirty paper (WDP) problem the coarse lattice should be good for source coding [10]. The lattice used for MLM needs to be simultaneously good for source *and* channel coding (see Appendix I). While the existence of such lattices in the high-dimension limit is assured by [21], in finite dimension the lattice that is best in one sense is not necessarily best in the other sense

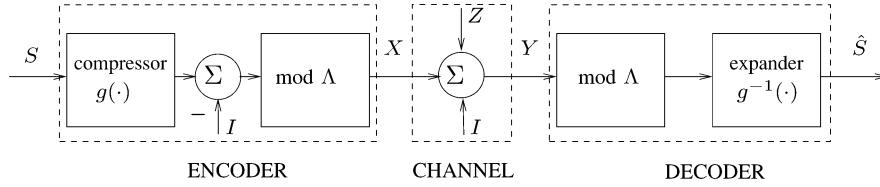


Fig. 7. Scalar MLM/comparing scheme for joint source–channel coding over a high-SNR dirty-paper channel:  $S$  = source,  $\hat{S}$  = reconstruction,  $Z$  = channel noise,  $I$  = interference known at the encoder,  $g(\cdot)$  = companding function.

[19], resulting in a larger implementation loss. Quantitatively, whereas for source coding the lattice should have a low normalized second moment, and for channel coding it should have a low volume-to-noise ratio, for joint source–channel coding the product  $L(\Lambda, p_e)$  (12) should be low<sup>5</sup> (see Appendix I). The study of such lattices is currently under research. Exact comparison of schemes in high dimension will involve studying the achieved *joint source–channel excess distortion exponent* (see [31] for a recent work about this exponent in the Gaussian setting).

From a practical point of view, the question of a low-dimensional scheme is very important, since it implies both low complexity and low delay. One may ask then, what can be achieved using low-dimensional lattices, e.g., a scalar lattice? The difficulty, however, is that in low dimensions a low probability of incorrect decoding  $p_e$  implies a high loss factor  $L(\Lambda, p_e)$ , thus the distortion promised by Theorem 1 grows. Some improvement may be achieved by using an optimal decoder rather than the one described in this work (see Remark 1 at the end of Section III), an issue which is left for further research. A recent work [32] suggests an alternative, for the case of channel interference only ( $\mathbf{J} = \mathbf{0}$ ), by also changing the encoder: The scalar zooming factor  $\beta$  of the MLM scheme is replaced by nonlinear companding of the signal; see Fig. 7. At high SNR, the distortion loss of such a scalar MLM scheme with optimal companding comparing to (7) is shown to be

$$\frac{D^{\text{companding}}}{D^{\text{opt}}} = \frac{\sqrt{3}\pi}{2} \cong 4.3 \text{ dB.}$$

In comparison, the loss of a separation-based scalar scheme, consisting of a scalar quantizer and a scalar (uncoded) channel constellation, is *unbounded* in the limit  $\text{SNR} \rightarrow \infty$ . This is since in a separation-based scheme the mapping of quantized source values to channel inputs is arbitrary; consequently, keeping the loss bounded implies that the error probability must go to zero in the high-SNR limit, and the gap of a scalar constellation from capacity grows.

#### APPENDIX I MEASURES OF GOODNESS OF LATTICES

In this appendix, we discuss measures of goodness of lattices for source and channel coding, and their connection with the loss factor relevant to our joint source/channel scheme.

<sup>5</sup>In Theorem 1, we show that the figure of merit is  $L(\Lambda, p_e, \alpha)$  (15) for some  $\alpha < 1$ , but for reasonably high SNR it seems that the effect of self-noise should not be too dominant, so we can set  $\alpha = 1$ .

When a lattice is used as a quantization codebook in the quadratic Gaussian setting, the figure of merit is the lattice *normalized second moment*

$$G(\Lambda) \triangleq \frac{\sigma^2(\Lambda)}{V(\Lambda)^{\frac{2}{K}}} \quad (45)$$

where the cell volume is  $V(\Lambda) = \int_{\mathcal{V}_0} d\mathbf{x}$ . By the iso-perimetric inequality  $G(\Lambda) \geq G_K^*$ , where  $G_K^*$  is the normalized second moment of a ball with the same dimension  $K$  as the lattice. This quantity satisfies  $G_K^* \geq \frac{1}{2\pi e}$ , with asymptotic equality in the limit of large dimension. A sequence of  $K$ -dimensional lattices is said to be *good for MSE quantization* if

$$\lim_{K \rightarrow \infty} G(\Lambda_K) = \frac{1}{2\pi e} \quad (46)$$

thus it asymptotically achieves the minimum possible lattice second moment for a given volume.

When a lattice is used as an AWGN channel codebook, the figure of merit is the lattice *volume-to-noise ratio* at a given error probability  $1 > p_e > 0$  (see, e.g., [20], [22])

$$\mu(\Lambda, p_e) \triangleq \frac{V(\Lambda)^{\frac{2}{K}}}{\sigma_Z^2} \quad (47)$$

where  $\sigma_Z^2$  is the maximum variance (per element) of a white Gaussian vector  $\mathbf{Z}$  having an error probability

$$\Pr\{\mathbf{Z} \notin \mathcal{V}_0\} \leq p_e.$$

For any lattice,  $\mu(\Lambda, p_e) \geq \mu_K^*(p_e)$ , where  $\mu_K^*(p_e)$  is the volume-to-noise ratio of a ball with the same dimension  $K$  as the lattice. For any  $1 > p_e > 0$ ,  $\mu_K^*(p_e) \geq 2\pi e$ , with asymptotic equality in the limit of large dimension. A sequence of  $K$ -dimensional lattices is *good for AWGN channel coding* if

$$\lim_{p_e \rightarrow 0} \lim_{K \rightarrow \infty} \mu(\Lambda_K, p_e) = 2\pi e \quad (48)$$

thus it possesses the property of having a minimum possible cell volume such that the probability of an i.i.d. Gaussian vector of a given power to fall outside the cell vanishes.

Combining the definitions (45) and (47), we see that the loss factor  $L(\Lambda, p_e)$  (12) satisfies

$$L(\Lambda, p_e) = G(\Lambda) \cdot \mu(\Lambda, p_e).$$

Furthermore, the existence of a good sequence of lattices in the sense of (13) is assured by the existence of a sequence that simultaneously satisfies (46) and (48), which was shown in [21, Theorem 5].

Proposition 1 is implicit in the proof of [22, Theorem 5]. It is based upon the existence of lattices that are simultaneously good for AWGN channel coding and for covering [21], where goodness for covering also implies goodness for MSE quantization;

for such lattices, it is shown that the mixture noise cannot be much worse than a Gaussian noise of the same variance. Later, it was shown in [33] that, for such lattices, for small enough error probability  $p_e$ , the introduction of self-noise actually reduces the loss factor, i.e.,  $L(\Lambda, p_e, \alpha) \leq L(\Lambda, p_e, 1)$ .

## APPENDIX II

### THE EFFECT OF DECODING FAILURE ON THE DISTORTION

With probability  $p_e$ , correct lattice decoding fails, i.e., (30) does not hold. These events contribute to the total distortion a portion of

$$\tilde{D} \triangleq p_e \cdot D^{\text{incorrect}} \quad (49)$$

where  $D^{\text{incorrect}}$  is the distortion given a decoding failure, as in the proof of Theorem 1. In this appendix, we quantify this effect: In the first part, we show that  $D^{\text{max}}$  of (24) is a (rather loose) bound on  $D^{\text{incorrect}}$ , thus completing the proof of Theorem 1. In the second part, we show directly that  $\tilde{D}$  must vanish in the limit of small  $p_e$ , without resorting to an explicit bound on  $D^{\text{incorrect}}$ .

In both parts we use the observation that

$$\hat{\mathbf{S}} - \mathbf{S} = \hat{\mathbf{Q}} - \mathbf{Q} \quad (50)$$

where  $\hat{\mathbf{Q}} \triangleq \frac{\alpha_S}{\beta} [\beta \mathbf{Q} + \mathbf{Z}_{\text{eq}}] \bmod \Lambda$ , see also Fig. 4(b). We note that although  $\hat{\mathbf{Q}}$  is unbounded, we always have that

$$\frac{\beta}{\alpha_S} \hat{\mathbf{Q}} \in \mathcal{V}_0. \quad (51)$$

#### A. A Bound on the Conditional Distortion for Any Lattice

In order to complete the proof of Theorem 1, we now bound  $D^{\text{incorrect}}$  of (33) by  $D^{\text{max}}$  of (24)

$$\begin{aligned} D^{\text{incorrect}} &= \frac{1}{K} E\{ \|\hat{\mathbf{S}} - \mathbf{S}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} \\ &= \frac{1}{K} E\{ \|\hat{\mathbf{Q}} - \mathbf{Q}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} \\ &\leq \frac{2}{K} (E\{ \|\hat{\mathbf{Q}}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} \\ &\quad + E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \}) \end{aligned} \quad (52)$$

where the inequality follows from assuming maximizing the  $(-1)$  correlation coefficient and then applying the Cauchy-Schwartz inequality. We shall now bound these two terms. For the first one, recalling the definition of the covering radius (11), we bound the conditional expectation by the maximum possible value

$$\begin{aligned} E\{ \|\hat{\mathbf{Q}}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} &\leq \max(\|\hat{\mathbf{Q}}\|^2) \\ &= \frac{\alpha_S^2 \cdot r^2(\Lambda)}{\beta^2} \\ &\leq \frac{r^2(\Lambda)}{\beta^2}. \end{aligned} \quad (53)$$

For the second term, we have

$$\begin{aligned} E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} &\leq E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} \notin \mathcal{V}_0 \} \\ &\leq E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} \notin \mathcal{B}_0 \} \end{aligned}$$

where  $\mathcal{B}_0$  is the circumsphere of  $\mathcal{V}_0$ , of radius  $r(\Lambda)$ . It follows that

$$E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} \leq \sigma_Q^2 E\{ V | V > v_0 \}$$

where  $V \sim \chi_K^2$  and  $v_0 \triangleq \frac{r^2(\Lambda)}{\beta^2 \sigma_Q^2}$ . This conditional expectation is given by

$$E\{ V | V > v_0 \} = \frac{\mathcal{Q}(\frac{K}{2} + 1, v_0 \frac{K}{2})}{\mathcal{Q}(\frac{K}{2}, v_0 \frac{K}{2})} \leq v_0 + 2$$

where  $\mathcal{Q}(\cdot, \cdot)$  is the regularized incomplete Gamma function, and the inequality can be shown by means of calculus. This gives the bound on the second term

$$E\{ \|\mathbf{Q}\|^2 | \beta \mathbf{Q} + \mathbf{Z}_{\text{eq}} \notin \mathcal{V}_0 \} \leq \frac{r^2(\Lambda)}{\beta^2} + 2K\sigma_Q^2.$$

Substituting this and (53) in (52), we have that

$$D^{\text{incorrect}} \leq 4 \left( \frac{r^2(\Lambda)}{K\beta^2} + \sigma_Q^2 \right).$$

Recalling the choice of  $\beta$  in (32b) and the definition of  $\tilde{L}(\cdot, \cdot)$  in (14), the bound  $D^{\text{max}}$  follows.

#### B. Asymptotic Effect of Decoding Failures

In this part, we follow the claims used by Wyner in the source coding context to establish [9, eq. (5.2)], to see that  $\lim_{p_e \rightarrow 0} \tilde{D} = 0$ , where  $\tilde{D}$  was defined in (49), without using the explicit bound derived in Appendix II-A. This serves as a simpler proof of Theorem 2; moreover, it also applies to a non-optimal choice of parameters, thus it allows the analysis of performance under uncertainty conditions.

Denoting the decoding failure event by  $\varepsilon$  and its indicator by  $I_\varepsilon$ , and recalling (50), we rewrite the contribution to the distortion as

$$\tilde{D} = E\{ I_\varepsilon \cdot (\hat{\mathbf{Q}} - \mathbf{Q})^2 \}.$$

For any value of the source unknown part  $\mathbf{Q}$ , the distortion is bounded by

$$d(\mathbf{Q}) \triangleq \sup_{\hat{\mathbf{Q}}} (\hat{\mathbf{Q}} - \mathbf{Q})^2.$$

The expectation  $E\{d(\mathbf{Q})\}$  is finite, since  $\mathbf{Q}$  is Gaussian and  $\hat{\mathbf{Q}}$  is bounded (see (51)). We now have that

$$\tilde{D} \leq E\{ I_\varepsilon \cdot d(\mathbf{Q}) \}.$$

Using a simple lemma of Probability Theory [9, Lemma 5.1], since  $E\{d(\mathbf{Q})\}$  is finite, this expectation approaches zero as  $p_e$ , the probability of the event  $\varepsilon$ , goes to zero.

## ACKNOWLEDGMENT

The authors wish to thank Uri Erez for helping to make some of the connections which led to this work.

## REFERENCES

- [1] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Info. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [2] S. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Pered. Inform. (Probl. Inf. Transm.)*, vol. 9, no. 1, pp. 19–31, 1980.
- [3] M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
- [4] N. Merhav and S. Shamai (Shitz), "On joint source-channel coding for the Wyner-Ziv source and the Gelfand-Pinsker channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2844–2855, Nov. 2003.
- [5] J. Ziv, "The behavior of analog communication systems," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 5, pp. 587–594, Sep. 1970.
- [6] M. D. Trott, "Unequal error protection codes: Theory and practice," in *Proc. Information Theory Workshop*, Haifa, Israel, Jun. 1996, p. 11.
- [7] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 4, pp. 558–567, Oct. 1965.
- [8] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [9] A. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inf. Contr.*, vol. 38, pp. 60–80, 1978.
- [10] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.
- [11] R. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1159–1180, May 2003.
- [12] Z. Reznic, M. Feder, and R. Zamir, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3778–3788, Aug. 2006.
- [13] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," in *Proc. IEEE Int. Symp. Information Theory (ISIT-2006)*, Seattle, WA, Jul. 2006, pp. 1539–1543.
- [14] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," in *Proc. Information Theory Workshop*, Lake Tahoe, CA, Sep. 2007, pp. 299–308.
- [15] M. Wilson, K. Narayanan, and G. Caire, "Joint source channel coding with side information using hybrid digital analog codes," *IEEE Trans. Inf. Theory* [Online]. Available: <http://arxiv.org/abs/0802.3851>, submitted for publication
- [16] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.
- [17] Y. Kochman, A. Khina, U. Erez, and R. Zamir, "Rematch and forward for parallel relay networks," in *Proc. IEEE Int. Symp. Information Theory (ISIT-2008)*, Toronto, ON, Canada, Jul. 2008, pp. 767–771.
- [18] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.
- [19] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer-Verlag, 1988.
- [20] G. D. Forney, Jr., M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.
- [21] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.
- [22] U. Erez and R. Zamir, "Achieving  $1/2 \log(1 + \text{SNR})$  on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.
- [23] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.
- [24] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai (Shitz), "Superposition coding for side-information channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.
- [25] J. K. Wolf, "Source coding for a noiseless broadcast channel," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 2004, pp. 666–671.
- [26] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 851–857, Nov. 1991.
- [27] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.
- [28] B. Chen and G. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 881–890, Jul. 1998.
- [29] D. Gunduz, J. Nayak, and E. Tuncel, "Wyner-Ziv coding over broadcast channels using hybrid digital/analog transmission," in *Proc. IEEE Int. Symp. Information Theory (ISIT-2008)*, Toronto, ON, Canada, Jul. 2008, pp. 1543–1547.
- [30] T. Lookabaugh and R. M. Gray, "High resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1020–1033, Sep. 1989.
- [31] Y. Zhong, F. Alajaji, and L. Campbell, "On the excess distortion exponent for memoryless Gaussian source-channel pairs," in *Proc. IEEE Int. Symp. Information Theory (ISIT-2006)*, Seattle, WA, Jul. 2006, pp. 2139–2143.
- [32] I. Leibowitz, "The Ziv-Zakai Bound at High Fidelity, Analog Matching, and Companding," Master's thesis, Tel-Aviv Univ., Tel-Aviv, Israel, Nov. 2007.
- [33] T. Liu, P. Moulin, and R. Koetter, "On error exponents of modulo lattice additive noise channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 454–471, Feb. 2006.

**Yuval Kochman** (S'07–M'09) was born in Tel-Aviv, Israel, in 1971. He received the B.Sc. and M.Sc. degrees in electrical engineering (both *cum laude*) from Tel-Aviv University in 1993 and 2003, respectively. He has recently completed his studies towards the Ph.D. degree at the Department of Electrical Engineering–Systems at Tel-Aviv University.

He is currently a Postdoctoral Associate at the Signals, Information and Algorithms Laboratory at the Massachusetts Institute of Technology (MIT), Cambridge. He has been working in the areas of radar and communications. His research interests include information theory, digital communications, data compression and signal processing.

**Ram Zamir** (S'89–M'95–SM'99) was born in Ramat-Gan, Israel, in 1961. He received the B.Sc., M.Sc. (*summa cum laude*) and D.Sc. (with distinction) degrees from Tel-Aviv University, Tel-Aviv, Israel, in 1983, 1991, and 1994, respectively, all in electrical engineering.

In the years 1994–1996, he spent a postdoctoral period at Cornell University, Ithaca, NY, and at the University of California, Santa Barbara. In 2002, he spent a Sabbatical year at the Massachusetts Institute of Technology (MIT), Cambridge. Since 1996, he has been with the department of Electrical Engineering–Systems at Tel Aviv University. He has been consulting in the areas of radar and communications, mainly in developing algorithms and in the design of signals, and has been teaching information theory, data compression, random processes, communications systems and communications circuits at Tel Aviv University. His research interests include information theory, communication and remote sensing systems, and signal processing.

Dr. Zamir received the Israel Ministry of Communications Award in 1993, and the Wolfson Post-Doctoral Research Award in 1994, and visited the Technical University of Budapest, Budapest, Hungary, under a joint program of the Israeli and Hungarian academies of science in summer 1995. He served as an Associate Editor for Source Coding in the IEEE TRANSACTIONS ON INFORMATION THEORY (2001–2003), and headed the Information Theory Chapter of the Israeli IEEE society (2000–2005).