

# The Case for Structured Random Codes in Network Communication Theorems

Bobak Nazer and Michael Gastpar  
University of California, Berkeley  
Wireless Foundations Research Center, Dept. of EECS  
Berkeley, CA 94720-1770, USA  
Email: bobak, gastpar@eecs.berkeley.edu

**Abstract**—In the Shannon-theoretic analysis of joint source-channel coding problems, achievability is usually established via a two-stage approach: The sources are compressed into bits, and these bits are reliably communicated across the noisy channels. Random coding arguments are the backbone of both stages of the proof. This strategy not only establishes the optimal performance for stationary ergodic point-to-point problems, but also for a number of simple network situations, such as independent sources that are communicated with respect to separate fidelity criteria across a multiple-access channel. Beyond such simple cases, for general networks, unstructured random coding arguments are not sufficient. This was first realized for source coding by Körner and Marton, who showed that for a distributed source coding problem where one only needs to recover a function of the sources random linear codes are necessary. The goal of this note is to extend this insight to pure channel coding as well as to joint source-channel coding problems, such as the problem of reliable computation over a multiple-access channel and a multiple-access network with relays. This version corrects an error that appears in the published version.

## I. INTRODUCTION

Random coding arguments are at the foundation of most rate-distortion and channel capacity achievability proofs, and the combination of the two is at the foundation of most joint source-channel coding achievability results. For the channel capacity side, the basic idea (for a multiterminal problem) is as follows. First, choose several random variables with an appropriate joint distribution. Then, generate high-dimensional codebooks with entries drawn i.i.d. according to this joint distribution. Finally, analyze the error performance of the codebooks in expectation and use this to show the existence of at least one good fixed set of codebooks. Amazingly, this method takes us quite far in network information theory. It has been successfully used to give the capacity region of the multiple-access channel [1], [2], stochastically degraded broadcast channel [3], and physically degraded relay channel [4], just to name a few. However, an elegant multiterminal problem developed by Körner and Marton showed that random code constructions are not always sufficient [5]; structured random codes, such as linear or lattice codes, may be required

on the achievability side of the proof. This key insight is the inspiration for this paper.

Structured random codes are usually considered to shed light on issues related to practical constructions. Given a capacity theorem, it is often of interest to demonstrate the existence of a capacity-achieving linear code to show that the storage requirements for the codebook need not be exponential in the blocklength. Although linearity is not enough to reduce the complexity of maximum-likelihood decoding, it often enables many complexity-saving reductions in approximate decoding algorithms.

As mentioned earlier, Körner and Marton found a simple many-help-one distributed source coding problem where random linear codes are needed to access the full rate region in [5]. In their consideration, a central decoder wants to reconstruct the parity of two correlated sources seen by separate encoders. By employing the same linear code at each encoder and summing codewords at the decoder, the parity can be reconstructed using rates too low to recover the individual sources. If the sources are independent, they must be sent in their entirety to the decoder.

One of the best known situations where joint source-channel coding improves performance is the reliable transmission of correlated sources over a multiple-access channel (MAC) [6], [7]. Cover, El Gamal, and Salehi use random coding arguments to give an achievable rate region in [6]. Their joint source-channel scheme uses the source correlations to create channel input probability distributions unavailable to a separation-based scheme. Exploiting the source correlations in this fashion is sometimes known as *collaborative gain*.

In [8], [9], an uncoded joint source-channel scheme is shown to be optimal (and significantly better than separation) for estimating a remote source from multiple observations. Although at a first glance, the scheme seems to benefit only from the correlations between the observations, it also exploits an ideal structural match between the channel, a Gaussian MAC, and the sufficient statistic, the sum of the observations. In extension of these results, we here illustrate that for general network joint source-channel coding problems, structured codes will be necessary to establish optimal performance. A similar conclusion was reached in [10] where lattices are used to achieve rates inaccessible to random codes for a MAC with

This work was supported by the National Science Foundation under CAREER Grant CCF-0347298 and NeTS-ProWin Grant CCF-0627024 as well a Graduate Research Fellowship.

interference known at the transmitters. Here, the paradigmatic example is the problem of reliable computation over MACs.

## II. DEFINITIONS

*Definition 1:* Choose an alphabet  $\mathcal{X}$ . An  $(n, R)$  code,  $\mathcal{C}$ , is a set of  $2^{nR}$  distinct length- $n$  vectors in  $\mathcal{X}^n$ . Each vector,  $\mathbf{c} \in \mathcal{C}$ , is referred to as a *codeword*.

We now give an informal definition of what we mean by a random coding argument. Unfortunately, it is quite difficult to give a formal definition that includes all current techniques (such as superposition coding) but excludes roundabout ways of constructing a set of linear codes by complicated thinning arguments.

Given that we are not restricted to using a single random codebook at each terminal, it is hard to bound the performance of all possible random coding arguments in expectation. For the scope of this paper, we limit ourselves to comparisons to the best known random coding argument for a particular problem class. For example, for a distributed lossless compression problem, we compare to the performance of random binning in expectation.

*Informal Definition 2:* A *random coding argument* consists of generating random codebooks for use at each encoder and evaluating the error performance of these codebooks in expectation under a decoding rule such as maximum-likelihood or joint typicality. Given that they perform well in expectation it is then argued that at least one good fixed codebook must exist. For instance, to achieve point-to-point channel capacity, we draw a random  $(n, R)$  code with every element of every codeword generated i.i.d. according to the capacity-achieving distribution  $p(x)$ .

Several powerful generalizations of the single random codebook construction have been studied including:

- Block Markov Coding [4], [11]
- Superposition Coding [3], [12]
- Compress-Forward [4]

With these tools in hand, most of the currently known achievability results of network information theory can be derived.

We will use the below definition for structured codes over discrete alphabets.

*Definition 3:* Choose a discrete alphabet  $\mathcal{X}$  and a function  $f : \mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X}$ . An  $(n, R)$  *structured code*,  $\mathcal{C}_f$ , is a set of  $2^{nR}$  codewords of length  $n$  such that given any two codewords,  $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}_f$ , applying  $f(\cdot)$  symbolwise to the the codewords results in another codeword: In other words,  $\mathcal{C}_f$  is closed under  $f(\cdot)$ . A *structured random code* is just a structured code drawn according to some distribution (which is not constrained to be i.i.d.)

As we will see, this definition is not quite sufficient for structured codes over the reals that must satisfy a power constraint. Furthermore, we do not use any non-linear structured codes in this paper and so it is not clear that the above definition is appropriate for non-linear functions.

### A. Linear Codes

Linear codes are the most commonly used structured codes as they often enable many complexity reductions in the encoding and decoding algorithms.

*Definition 4:* Let  $\mathbb{F}$  be a finite field. An  $(n, R)$  linear code,  $\mathcal{C}$ , is a structured code on  $\mathbb{F}^n$  that is closed under the additive operation of  $\mathbb{F}$ . That is, given any two codewords,  $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$ , adding the codewords symbolwise results in another codeword. A *random linear code* is just a linear code drawn according to some distribution. For every linear code there is at least one matrix  $\mathbf{H} \in \mathbb{F}^{nR \times n}$  such that each codeword,  $\mathbf{c} \in \mathcal{C} \subset \mathbb{F}^n$ , can be written as  $\mathbf{xH}$  for some  $\mathbf{x} \in \mathbb{F}^{nR}$ .

### B. Lattice Codes

Using linear codes over the reals is slightly more tricky. A similar notion of linearity is available for the reals in the form of a lattice code. An additional complication is the fact that for such channels, input constraints are imposed (such as a power constraint in the case of the standard AWGN channels). In other words, since a lattice has infinite extent, we cannot simply use every point as a codeword.

*Definition 5:* An  $n$ -dimensional *lattice*,  $\Lambda$ , is a set of points in  $\mathbb{R}^n$  such that if  $\mathbf{x}, \mathbf{y} \in \Lambda$ , then  $\mathbf{x} + \mathbf{y} \in \Lambda$ , and if  $\mathbf{x} \in \Lambda$ , then  $-\mathbf{x} \in \Lambda$ . A lattice can always be written in terms of a generator matrix  $\mathbf{G} \in \mathbb{R}^{n \times n}$ :

$$\Lambda = \{\mathbf{x} = \mathbf{zG} : \mathbf{z} \in \mathbb{Z}^n\}, \quad (1)$$

where  $\mathbb{Z}$  represents the integers.

*Definition 6:* An  $(n, R)$  lattice code,  $\mathcal{C}$ , is a code with elements taken from the intersection of some  $n$ -dimensional lattice  $\Lambda$  and a convex  $n$ -dimensional shape  $T$  (such as a power constraint), that is,  $\mathcal{C} = \Lambda \cap T$  and  $|\mathcal{C}| = 2^{nR}$ . Note that given any two codewords  $\mathbf{c}_1$  and  $\mathbf{c}_2$  in  $\mathcal{C}$ , their sum is an element of the intersection of  $\Lambda$  and the direct sum of  $T$  with itself.

*Definition 7:* A *lattice quantizer* is a map,  $Q : \mathbb{R}^n \rightarrow \Lambda$ , that sends a point,  $\mathbf{x}$ , to the nearest lattice point in Euclidean distance:

$$\mathbf{x}_q = Q(\mathbf{x}) = \arg \min_{\mathbf{l} \in \Lambda} \|\mathbf{x} - \mathbf{l}\|_2. \quad (2)$$

Of course, non-linear structured codes may be quite useful. Unfortunately, we are not currently aware of any non-linear structured code constructions that surpass purely random codes in a class of communications problems.

A great deal of work has gone into showing that linear and lattice codes are sufficient for many channel coding and source coding problems. In the following section, we will briefly review some of these results.

## III. STRUCTURED CODES FOR POINT-TO-POINT PROBLEMS

We will now review some of the previous work on the existence of capacity-achieving structured codes for classical point-to-point problems.

For additive noise channels, structured codes can achieve rates all the way up to capacity. Work in this area began with Elias' proof that there exist binary linear codes which are good

for channel coding over the binary symmetric channel (BSC) [13]. In fact, Elias' method extends to a much larger class of channels. If we assume the alphabet is over a finite field and the uniform input distribution is capacity-achieving (which is the case for additive noise), then a matrix with elements chosen i.i.d. and uniformly from the finite field will suffice. This is captured in the following lemma from Problem 2.1.11 in [14].

*Lemma 1:* Let  $\mathbf{w} \in \mathbb{F}^k$  be the message and let the channel output be given by  $\mathbf{y} = \mathbf{x} + \mathbf{z}$  where  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{F}^n$  and  $\mathbf{z}$  is an i.i.d. sequence. Then the capacity of the channel is given by  $C = \log |\mathbb{F}| - H(Z)$  and can be achieved with a linear code  $\mathbf{G}^{k \times n}$  so that  $\mathbf{x} = \mathbf{w}\mathbf{G}$ . Specifically, for any  $\epsilon > 0$  and  $n$  large enough,  $\Pr(c(\mathbf{y}) \neq \mathbf{w}) < \epsilon$  where  $c(\cdot)$  is the maximum-likelihood (ML) estimate of  $\mathbf{w}$ . Note that  $k \log |\mathbb{F}| < nC$  is required to stay below the capacity.

Linear codes can also be used for compressing any discrete alphabet source so long as the rate is higher than the source entropy. In fact, they can reach any point in the Slepian-Wolf rate region for distributed compression [15].

For additive white Gaussian noise (AWGN) channels, showing that lattice codes are sufficient to reach capacity was considerably more challenging. An AWGN point-to-point channel has an output  $Y \in \mathbb{R}$  which can be written as:

$$Y = X + Z, \quad (3)$$

where  $X$  is the channel input and  $Z$  is i.i.d. Gaussian noise with variance  $N$ . Unlike the discrete alphabet case, AWGN channel encoders are usually subject to a power constraint of the form:

$$\frac{1}{n} \sum_{i=1}^n x_i^2 \leq P. \quad (4)$$

The capacity of an AWGN channel is well-known to be:

$$C = \frac{1}{2} \log \left( 1 + \frac{P}{N} \right). \quad (5)$$

As lattices have infinite extent (and thus violate the power constraint), much effort was focused on finding lattices that when intersected with an  $n$ -dimensional ball of radius  $\sqrt{nP}$  centered at 0 form a good code. Urbanke and Rimoldi showed that such lattices indeed exist in [16]. Further work by Erez and Zamir has focused on proving that decoding to the closest lattice point also achieves capacity [17]. We will not concern ourselves with reduced decoding complexity. However, we will make use of a related result from Erez, Litsyn, and Zamir [18] that there exist lattices which are simultaneously good for AWGN channel coding and Gaussian source coding.

Given an i.i.d. Gaussian source with variance  $\sigma_S^2$  the rate distortion function for mean-squared error is:

$$R(D) = \frac{1}{2} \log \left( \frac{\sigma_S^2}{D} \right). \quad (6)$$

The following lemma from [18] captures how lattices can be used for channel coding and source coding:

*Lemma 2:* Let  $\mathbf{s}$  be a length- $n$  Gaussian vector with variance  $\sigma_S^2$ . For  $n$  large enough, there exists a lattice,  $\Lambda$ , such

that quantizing  $\mathbf{s}$  to the lattice points inside  $\Lambda \cap B_0(\sqrt{n\sigma_S^2})$  is sufficient for recovery at mean-squared error (or distortion)  $D$ . Furthermore, the number of points in  $\Lambda \cap B_0(\sqrt{n\sigma_S^2})$  satisfies:

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\Lambda \cap B_0(\sqrt{n\sigma_S^2})| = \frac{1}{2} \log \left( \frac{\sigma_S^2}{D} \right). \quad (7)$$

Transmitting points from the same lattice (up to inflation) intersected with a power sphere  $B_0(\sqrt{nP})$  followed by maximum-likelihood decoding is sufficient for achieving any rate below the capacity of an AWGN channel with power constraint  $P$  and noise variance  $N$ .

#### IV. DISTRIBUTED COMPUTATION

When we are interested in computing a function of distributed sources, structured codes can be extremely useful. In essence, an appropriately chosen structured code will commute with respect to the desired function. The function can then be applied to codewords instead of the original sources. This technique can reduce the required rates as perfect reconstruction of the sources is no longer required.

##### A. Körner-Marton Problem

Körner and Marton found a simple distributed compression problem with an elegant structured coding solution in [5]. Their work was (to the best of our knowledge) the first situation where structured codes were required to complete the achievability proof.

Let the vector source  $(S_1, S_2)$  be generated iid from the following joint probability distribution function (pdf):

$$\Pr(S_1 = 0, S_2 = 0) = \Pr(S_1 = 1, S_2 = 1) = \frac{1-p}{2} \quad (8)$$

$$\Pr(S_1 = 0, S_2 = 1) = \Pr(S_1 = 1, S_2 = 0) = \frac{p}{2}. \quad (9)$$

A simple calculation will show that  $S_1$  and  $S_2$  have uniform marginal distributions. This joint pdf is the only one for which the scheme presented is known to be optimal [7]. We would like to reconstruct the mod-2 sum,  $U = S_1 \oplus S_2$ , at the decoder with vanishing probability of error. More formally, we would like to find the set of rates  $R_1$  and  $R_2$  such that there exist two encoders and a decoder:

$$\mathcal{E}_j : \{0, 1\}^n \rightarrow \{0, 1\}^{nR_j} \quad j = 1, 2 \quad (10)$$

$$\mathcal{D} : \{0, 1\}^{nR_1} \times \{0, 1\}^{nR_2} \rightarrow \{0, 1\}^n, \quad (11)$$

such that the probability of error for recovering  $U$  goes to 0 in the blocklength:

$$\hat{\mathbf{u}} = \mathcal{D}(\mathcal{E}_1(\mathbf{s}_1), \mathcal{E}_2(\mathbf{s}_2)) \quad (12)$$

$$\lim_{n \rightarrow \infty} P(\hat{\mathbf{u}} \neq \mathbf{u}) = 0. \quad (13)$$

##### 1) Optimal Rate Region:

*Theorem 1 (Körner-Marton):* The rate region for distributed compression of  $U = S_1 \oplus S_2$  is given by the following constraints:

$$R_1 > h_B(p) \quad \text{and} \quad R_2 > h_B(p), \quad (14)$$

where  $h_B(p)$  is the usual binary entropy function.

*Proof: (Achievability.)* Choose a linear source code,  $\mathbf{G} \in \{0, 1\}^{n \times nR}$  with rate  $R > h_B(p)$  that is sufficient for losslessly compressing  $U$ . Have each encoder apply this code to its observed source to get  $\mathbf{w}_1 = \mathbf{s}_1 \mathbf{G}$  and  $\mathbf{w}_2 = \mathbf{s}_2 \mathbf{G}$ . These codewords are sent to the decoder which computes  $\mathbf{w}_1 \oplus \mathbf{w}_2 = \mathbf{s}_1 \mathbf{G} \oplus \mathbf{s}_2 \mathbf{G} = \mathbf{u} \mathbf{G}$ . Since  $\mathbf{G}$  was chosen for recovering  $U$ , decoding is successful.

*(Converse.)* Consider the relaxation where the decoder has full knowledge of  $S_2$  and we would like to jointly encode  $S_1$  and  $U$  to losslessly reconstruct  $U$  at the decoder. Note that any scheme that accomplishes this also gives the decoder a lossless reconstruction of  $S_1$ . Thus, it can be shown that for joint encoding,  $R \geq H(S_1, U|S_2) = H(U|S_2) = H(U) = h_B(p)$  is required for a vanishing probability of error. This implies that for separate encoding of  $S_1$  and  $U$ ,  $R_1 + R_U \geq h_B(p)$ . Similarly, we can get that  $R_2 + R_U \geq h_B(p)$ . Setting  $R_U = 0$  gives the desired result. ■

2) *Performance of Best Known Random Code:* The best known random coding strategy is random binning at each terminal: The encoder for  $S_1$  randomly and uniformly assigns each length- $n$  sequence  $\mathbf{s}_1$  into one of  $2^{nR_1}$  bins. The encoder for  $S_2$  does the same using  $2^{nR_2}$  bins. These bin indices are transmitted to the decoder. If  $R_1 + R_2 > 1 + h_B(p)$ ,  $R_1 > h_B(p)$ , and  $R_2 > h_B(p)$ , then the decoder can completely reconstruct  $\mathbf{s}_1$  and  $\mathbf{s}_2$ , thus  $\mathbf{u}$  follows by taking the mod-2 sum. However, if lower rates ( $R_1, R_2$ ) are used,  $\mathbf{u}$  cannot be recovered. To give the intuition of the proof, note that nearly all the probability is concentrated in approximately  $2^{n(1+h_B(p))}$  pairs of sequences. Suppose that  $\mathbf{s}_1^*$  and  $\mathbf{s}_2^*$  are the sequences seen at the encoders, that they are assigned bins  $i_1$  and  $i_2$  respectively, and that  $\mathbf{u}^*$  is the mod-2 sum. There are exactly  $2^n - 1$  sequence pairs with the same mod-2 sum. Essentially all other sequence pairs must end up with different index pairs, or it will not be possible to reliably recover  $\mathbf{u}^*$ . However, the probability of the latter event is small.

## B. Computation over Multiple-Access Channels

In the standard multiple-access problem, the decoder must recover the messages sent by each encoder. Suppose now that we are only interested in recovering a function of the transmitted messages. If the multiple-access channel (MAC) is simply a deterministic function of its inputs, then clearly it can be used as a reliable computational unit for computing that function. However, if noise is also injected, then some form of coding is required to compute reliably.

In [19], we gave a class of strategies for computing functions over noisy MACs. As it turns out, structured codes are an essential part of the code construction. We briefly describe two distributed linear computation problems below and refer the interested reader to [19] for a more comprehensive study.

1) *Discrete Case:* First we consider sending linear functions over a discrete linear MAC.

Let  $\mathbb{F}$  be a finite field and let the vector source  $(S_1, S_2, \dots, S_M) \in \mathbb{F}^M$  be generated i.i.d. from some joint pdf. We would like to reconstruct the linear function,  $U = \alpha_1 S_1 + \alpha_2 S_2 + \dots + \alpha_M S_M$ , at the decoder with vanishing

probability of error. Each source is seen by an encoder with channel input  $X_j \in \mathbb{F}$  for  $j = 1, 2, \dots, M$ . A discrete linear MAC is given by  $Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_M X_M + Z$  where  $Z \in \mathbb{F}$  is drawn i.i.d. according to some pdf.

More formally, we would like to find the highest computation rate  $R$  such that there exist  $M$  encoders and a decoder:

$$\mathcal{E}_j : \mathbb{F}^{nR} \rightarrow \mathbb{F}^n \quad j = 1, 2, \dots, M \quad (15)$$

$$\mathcal{D} : \mathbb{F}^n \rightarrow \mathbb{F}^{nR}, \quad (16)$$

such that the probability of error for recovering  $U$  goes to 0 in the blocklength:

$$\lim_{n \rightarrow \infty} P(\hat{\mathbf{u}} \neq \mathbf{u}) = 0, \quad (17)$$

where  $\hat{\mathbf{u}} = \mathcal{D}(\mathbf{y})$ .

*Theorem 2:* The capacity for computing  $U$  over a discrete linear MAC is given by:

$$C = \frac{\log |\mathbb{F}| - H(Z)}{H(U)}. \quad (18)$$

In fact, a similar result holds for transmitting several (possibly correlated) linear functions. See [19] for more details. We use the term *computation coding* to refer to the strategy employed here.

*Proof: (Achievability.)* Choose a matrix  $\mathbf{H}$  that is appropriate for compressing  $U$ . Similarly, choose a good point-to-point channel coding matrix  $\mathbf{G}$  for overcoming noise  $Z$ . At each encoder we set  $\mathbf{x}_j = \beta_j^{-1} \alpha_j \mathbf{H} \mathbf{G}$ . After the linear operation performed by the channel, we get:

$$\mathbf{y} = \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_M \mathbf{x}_M + \mathbf{z} \quad (19)$$

$$\mathbf{y} = \mathbf{u} \mathbf{H} \mathbf{G} + \mathbf{z}. \quad (20)$$

The design of  $\mathbf{G}$  allow us to recover from the additive noise and the design of  $\mathbf{H}$  allows us to recover  $\mathbf{u}$  with a vanishing probability of error. *(Converse.)* For this class of MACs, we can simply allow the encoders to completely collaborate and get a tight upper bound. This reduces our problem to a point-to-point problem and we can invoke the separation theorem to get a converse:  $RH(U) < \log |\mathbb{F}| - H(Z)$ . ■

2) *Gaussian Case:* The natural extension of the discrete problem considered above to the continuous case is transmitting the sum of Gaussian sources over a Gaussian MAC at the minimal mean-squared error. When the source and channel bandwidths are equal, then uncoded transmission is optimal. However, given more channel uses than source symbols, we would like to continue to use the additive property of the MAC to our advantage. Below we give an achievable scheme for refining the sum over many channel uses. Unfortunately, at this time we do not have a tight lower bound on the distortion.

Each encoder,  $\mathcal{E}_j$ , sees an independent identically distributed (i.i.d.) Gaussian sequence  $\{S_j[i]\}_{i=1}^k$  with mean 0 and variance  $\sigma_S^2$ . For every  $k$  source symbols, we are allotted  $n = \ell k + r$  channel uses where  $\ell, r \in \mathbb{Z}_+$  and  $r < k$ .

$$\mathcal{E}_j : \mathbb{R}^k \rightarrow \mathbb{R}^n. \quad (21)$$

The encoders must satisfy average power constraints:

$$\frac{1}{n} \sum_{i=1}^n x_j[i]^2 \leq P \quad \forall j \in \{1, 2, \dots, M\}. \quad (22)$$

The channel output is just the sum of the channel inputs plus independent Gaussian noise:

$$Y[i] = \sum_{j=1}^M X_j[i] + Z[i], \quad (23)$$

where  $\{Z[i]\}_{i=1}^n$  is an i.i.d. Gaussian sequence with mean 0 and variance  $\sigma_Z^2$ .

Our goal is to reconstruct the sum of the sources,  $U = S_1 + S_2 + \dots + S_M$ , at the decoder with the lowest possible distortion. Distortion is measured by the usual mean-squared error criterion:

$$D = \frac{1}{k} \sum_{i=1}^k E[(U_i - \hat{U}_i)^2]. \quad (24)$$

*Theorem 3:* The distortion for sending  $k$  sums of independent Gaussian sources over a Gaussian MAC with  $n = \ell k$ ,  $\ell \in \mathbb{Z}_+$  channel uses is bounded above and below as follows:

$$M\sigma_S^2 \left( \frac{N}{N+MP} \right)^\ell \leq D \leq \sigma_S^2 \left( \frac{MN}{N+MP} \right)^\ell. \quad (25)$$

*Proof:* (Upper Bound.) We first use uncoded transmission to communicate our observation sequences across the channel to get an MMSE estimate  $\hat{\mathbf{u}}$  of the sum  $\mathbf{u}$  at distortion  $M\sigma_S^2 \left( \frac{N}{N+MP} \right)$ . Now we employ our lattice-based scheme from Theorem 3 in [19] to refine this estimate of our sum with the remaining  $(\ell - 1)k$  channel uses. The lattice scheme essentially works as follows. We choose a lattice  $\Lambda$  in  $\mathbb{R}^k$  that is good for both source coding and channel coding using the results of [18]. For a block of  $k$  channel uses, each encoder transmits:

$$\mathbf{x}_m = [\gamma \mathbf{s}_j + \mathbf{d}_j] \bmod \Lambda \quad (26)$$

where  $\mathbf{d}_j$  is a random vector available as common randomness which is used as a dither and  $\gamma$  is a constant in  $\mathbb{R}$ . For details, see [19]. The decoder combines the received signal  $\mathbf{y}$  with the previous estimate  $\hat{\mathbf{u}}$  to get a new estimate  $\hat{\hat{\mathbf{u}}}$ :

$$\hat{\hat{\mathbf{u}}} = \beta \left[ \alpha \mathbf{y} - \left( \sum_{j=1}^M \mathbf{d}_j + \gamma \hat{\mathbf{u}} \right) \right] \bmod \Lambda + \hat{\mathbf{u}} \quad (27)$$

where  $\alpha$  and  $\beta$  are appropriately chosen constants in  $\mathbb{R}$ . This process is iterated until we have expended all  $\ell k$  channel uses to get the desired distortion.

(Lower Bound.) Using steps from the converse to the multiple-access problem (see [20, pp. 399-407]) as well as the independence of the sources, we can get that  $I(X_1, X_2, \dots, X_M; Y) \leq \frac{1}{2} \log \left( 1 + \frac{MP}{N} \right)$ . It is also clear that the rate distortion function for jointly compressing the sum is given by  $R_U(D) = \frac{1}{2} \log \left( \frac{M\sigma_S^2}{D} \right)$ . By applying the data processing inequality, we get the desired lower bound. ■

3) *Performance of Best Known Random Code:* For the problems considered above, random coding arguments perform quite poorly. As in the Körner-Marton problem, the best performance one can hope for is complete reconstruction of the sources at the decoder followed by computing the sum. This reduces the rate at which functions are computed proportionally to the number of users if the sources are independent.

For the discrete problem considered in Section IV-B.1, the random coding argument is as follows. Each encoder generates  $2^{nR}$  codewords according to the uniform distribution on  $\mathbb{F}$  (which is also the capacity-achieving distribution). Suppose the vectors  $\mathbf{s}_1^*, \mathbf{s}_2^*, \dots, \mathbf{s}_M^*$  are seen at each encoder and the desired function is given by  $\mathbf{u}^*$ . Note that nearly all the probability is concentrated in approximately  $2^{nH(S_1, S_2, \dots, S_M)}$  sequences. Of these sequences, approximately  $2^{nH(S_1, S_2, \dots, S_M|U)}$  will also result in the same desired function  $\mathbf{u}^*$ . For a joint typicality decoder, the probability that an incorrect set of vectors is found to be jointly typical with the channel output is approximately  $2^{-nI(X_1, X_2, \dots, X_M; Y)}$ . Note that  $I(X_1, X_2, \dots, X_M; Y) = \log |\mathbb{F}| - H(Z)$  for the chosen input distribution. Since  $H(S_1, S_2, \dots, S_M) > H(S_1, S_2, \dots, S_M|U)$ , the probability that an incorrect function  $\tilde{\mathbf{u}}$  will be jointly typical with the channel output  $\mathbf{y}$  will go to 1 unless we choose  $R$  such that  $R < (\log |\mathbb{F}| - H(Z))/H(S_1, S_2, \dots, S_M)$ . This is equivalent to requiring complete reconstruction of the sources.

For the Gaussian problem, a similar argument shows that random coding arguments can only achieve distortion  $D_{\text{SEP}} = M\sigma_S^2 (N/(N+MP))^{\ell/M}$ . Again, this is equivalent to reconstructing each source at the decoder with distortion  $\frac{D}{M}$  and then adding them up.

## V. RELAY-TYPE PROBLEM

So far we have seen that structured random codes offer gains when we are interested in computing a function of sources in a distributed fashion. We now demonstrate that computation over MACs can be quite useful in analyzing more traditional channel coding problems.

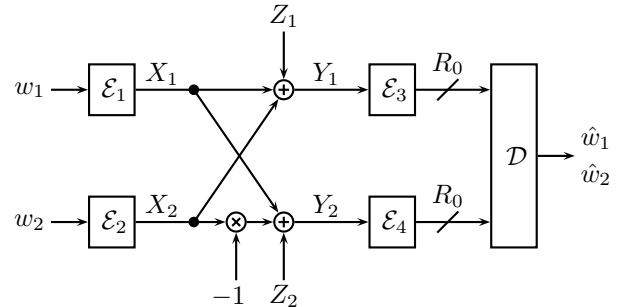


Fig. 1. Sum-Difference Relay Channel

Two separate users would like to send independent messages  $w_1, w_2 \in \{1, 2, \dots, 2^{nR}\}$  to a decoder across a multiple-access channel with relays (see Figure 1). Encoder 1 produces

channel input  $X_1^n \in \mathbb{R}^n$  with average power constraint  $P$  and Encoder 2 must produce channel input  $X_2^n \in \mathbb{R}^n$  with the same constraint. One relay terminal sees the sum of these signals plus noise:  $Y_1 = X_1 + X_2 + Z_1$ . The other terminal sees the difference plus noise:  $Y_2 = X_1 - X_2 + Z_2$ . Finally each relay has a noiseless bit pipe to the decoder with rate  $R_0$ . We would like to determine the maximum achievable symmetric rate  $R$  to the decoder for a given value of  $R_0$ .

*Theorem 4:* For the two relay problem described above, the following symmetric rate is achievable using lattice codes:

$$R_{\text{LAT}} = \min \left( \frac{1}{2} \log \left( \frac{1}{2} + \frac{P}{N} \right), R_0 \right). \quad (28)$$

We now provide an outline of the proof. For a more detailed study, see [21]. Choose  $\mathbf{s}_1$  and  $\mathbf{s}_2$  to be i.i.d. length- $k$  Gaussian vectors with mean 0 and variance  $\sigma_S^2$ . We will use the channel network  $n = k\ell$  times where  $\ell \in \mathbb{Z}_+$  will be specified later. Use Theorem 3 to pick a code for sending the sum  $\mathbf{u} = \mathbf{s}_1 + \mathbf{s}_2$  to Encoder 3 at distortion  $D \triangleq \sigma_S^2 \left( \frac{2N}{N+2P} \right)^\ell$ . Due to the symmetry of the underlying lattice code and the negative sign on the lower path, Encoder 4 will be able to reconstruct the difference  $\mathbf{v} = \mathbf{s}_1 - \mathbf{s}_2$  at distortion  $D$  as well.

In order to send the sum and the difference to the final decoder, we will need to requantize them. Pick a Gaussian source code for compressing a variance  $2\sigma_S^2$  source to distortion  $D_0 = 2\sigma_S^2 2^{-2\ell R_0}$ . By the triangle inequality, this requantization step will cause the total distortion for  $\mathbf{u}$  and  $\mathbf{v}$  to be at most  $D_0 + D$ . At the decoder we estimate  $\mathbf{s}_1$  by  $\hat{\mathbf{s}}_1 = \frac{1}{2}(\hat{\mathbf{u}} + \hat{\mathbf{v}})$  and  $\mathbf{s}_2$  by  $\hat{\mathbf{s}}_2 = \frac{1}{2}(\hat{\mathbf{u}} - \hat{\mathbf{v}})$ . It can be checked that these give us the original sources to within distortion  $\max \left( 4\sigma_S^2 \left( \frac{2N}{N+2P} \right)^\ell, 2D_0 \right)$ .

Finally, we note that if we consider supersymbols of length  $n$  and fix  $\mathcal{E}_3$  and  $\mathcal{E}_4$ , then we can invoke the separation theorem for a MAC. It can then be shown that the rate achievable by each encoder  $R = \min \left( \frac{1}{2} \left( \log \left( \frac{1}{2} + P/N \right) \right), R_0 \right) - \frac{1}{2} \log(4)$ . Thus, for  $\ell$  large enough, we can achieve the desired rates.

The best known random coding strategies are known as decode-and-forward (DF) and compress-and-forward (CF) with Gaussian codebooks [22]. The symmetric rates per user for each of these strategies are:

$$R_{\text{DF}} = \min \left( \frac{1}{4} \log \left( 1 + \frac{2P}{N} \right), R_0 \right) \quad (29)$$

$$R_{\text{CF}} = \frac{1}{2} \log \left( 1 + \frac{2P(2^{2R_0} - 1)}{2P + N2^{2R_0}} \right). \quad (30)$$

It is easy to see that neither of these strategies can reach the performance of the lattice-based strategy for sufficiently high SNR. In the published version we claimed that  $\frac{1}{2} \log \left( 1 + \frac{P}{N} \right)$  was achievable but there was an error in the proof.

## VI. DISCUSSION AND OPEN PROBLEMS

In this paper, we have shown that random coding arguments are not always sufficient to establish optimal performance in networked joint source-channel coding problems. That is, structured random codes can give better achievable

performance than standard random code constructions. The paradigmatic case is the problem of reliable computation over multiple-access channels, for which we have developed codes. These codes can then be used in a somewhat modular fashion in larger problems, such as the problem of multicasting over a set of point-to-point and multi-access links (“network coding” with multi-access components) [19], [21], [23].

## REFERENCES

- [1] R. Ahlswede, “Multi-way communication channels,” in *Proceedings of the 2nd ISIT, Prague*, pp. 23–52, Publishing House of the Hungarian Academy of Sciences, 1971.
- [2] H. Liao, *Multiple access channels*. PhD thesis, University of Hawaii, Honolulu, 1972.
- [3] T. M. Cover, “Comments on broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2524–2530, October 1998.
- [4] T. M. Cover and A. A. E. Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inform. Theory*, vol. 25, pp. 572–584, September 1979.
- [5] J. Körner and K. Marton, “How to encode the modulo-two sum of binary sources,” *IEEE Trans. Inform. Theory*, vol. 25, pp. 219–221, March 1979.
- [6] T. Cover, A. El Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 648–657, November 1980.
- [7] R. Ahlswede and T. S. Han, “On source coding with side information via a multiple-access channel and related problems in multi-user information theory,” *IEEE Trans. Inform. Theory*, vol. 29, pp. 396–412, May 1983.
- [8] M. Gastpar and M. Vetterli, “Source-channel communication in sensor networks,” in *2nd Int Workshop on Info Proc in Sensor Networks (IPSN '03)* (L. J. Guibas and F. Zhao, eds.), Lecture Notes in Computer Science, (New York, NY), pp. 167–177, Springer, April 2003.
- [9] M. Gastpar, “Uncoded transmission is exactly optimal for a simple Gaussian “sensor” network,” in *2nd Annual ITA, UCSD*, (La Jolla, CA), January 2007.
- [10] T. Philosof, A. Khisti, U. Erez, and R. Zamir, “Lattice strategies for the dirty multiple access channel,” in *IEEE ISIT 2007, Nice, France*.
- [11] T. M. Cover and C. S. K. Leung, “An achievable rate region for the multiple-access channel with feedback,” *IEEE Trans. Inform. Theory*, vol. 27, pp. 292–298, May 1981.
- [12] T. M. Cover, “Broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 18, pp. 2–14, January 1972.
- [13] P. Elias, “Coding for noisy channels,” *IRE Convention Record*, vol. 4, pp. 37–46, 1955.
- [14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1982.
- [15] I. Csiszár, “Linear codes for sources and source networks: Error exponents, universal coding,” *IEEE Trans. Inform. Theory*, vol. 28, pp. 585–592, July 1982.
- [16] R. Urbanke and B. Rimoldi, “Lattice codes can achieve capacity on the AWGN channel,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 273–278, January 1998.
- [17] U. Erez and R. Zamir, “Achieving  $\frac{1}{2} \log(1 + \text{SNR})$  on the AWGN channel with lattice encoding and decoding,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 2293–2314, October 2004.
- [18] U. Erez, S. Litsyn, and R. Zamir, “Lattices which are good for (almost) everything,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 3401–3416, October 2005.
- [19] B. Nazer and M. Gastpar, “Computation over multiple-access channels,” *IEEE Transactions on Information Theory*, vol. 53, pp. 3498–3516, October 2007.
- [20] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.
- [21] B. Nazer and M. Gastpar, “The case for structured random codes in network capacity theorems,” Submitted to European Trans. Telecom.
- [22] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Transactions on Information Theory*, vol. 51, pp. 3037–3063, September 2005.
- [23] B. Nazer and M. Gastpar, “Lattice coding increases multicast rates for Gaussian multiple-access networks,” in *45th Annual Allerton Conference*, (Monticello, IL), September 2007.