Integer-Forcing Source Coding

Or Ordentlich and Uri Erez, *Member, IEEE*

Abstract—Integer-Forcing (IF) is a new framework, based on compute-and-forward, for decoding multiple integer linear combinations from the output of a Gaussian multiple-input multiple-output channel. This work applies the IF approach to arrive at a new low-complexity scheme, IF source coding, for distributed lossy compression of correlated Gaussian sources under a minimum mean squared error distortion measure. All encoders use the same nested lattice codebook. Each encoder quantizes its observation using the fine lattice as a quantizer and reduces the result modulo the coarse lattice, which plays the role of binning. Rather than directly recovering the individual quantized signals, the decoder first recovers a full-rank set of judiciously chosen integer linear combinations of the quantized signals, and then inverts it. In general, the linear combinations have smaller average powers than the original signals. This allows to increase the density of the coarse lattice, which in turn translates to smaller compression rates. We also propose and analyze a one-shot version of IF source coding, that is simple enough to potentially lead to a new design principle for analog-to-digital converters that can exploit spatial correlations between the sampled signals.

I. INTRODUCTION

The distributed lossy compression problem, depicted in Figure 1, consists of multiple distributed encoders and one decoder. The encoders have access to correlated observations which they try to describe to the decoder with minimum rate and minimum distortion [1]–[3]. This problem naturally arises in numerous scenarios. For instance, consider a sensor network where multiple sensors that observe correlated random variables are connected via finite rate links to a central processor, but not to one another, and have to describe their observations to the central processor with minimum distortion. As another example, consider two competing television channels that cover the same event and have to broadcast their programs to the same end-users (that may choose which channel to watch and therefore need to be able to recover both programs with low distortion). Although the distributed lossy compression problem is usually classified as a pure source-coding problem, it is also an important building block in network channel coding problems. For instance, multiple relays may observe correlated signals that describe the messages transmitted by the different encoders in the network. The relays can compress-and-forward these signals further down the network in order to ultimately help the decoder recover the transmitted messages.

A special case that received considerable attention is that of distributed lossy compression of jointly Gaussian random
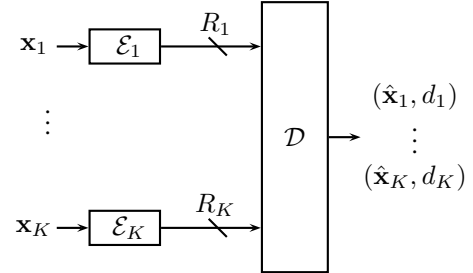

Fig. 1. The distributed source coding problem. The kth encoder $\mathcal{E}_k$ has access to the vector $\mathbf{x}_k$ that contains n i.i.d. realizations of the random variable x_k. It encodes $\mathbf{x}_k$ to an index taking values in $1, \ldots, 2^{nR_k}$. The sources $x_1, \ldots, x_K$ are assumed correlated and the encoders are not allowed to cooperate. The decoder's goal is to produce estimates of each $\mathbf{x}_k$ with average distortions d_k using the K indices it received from the encoders.

variables under a quadratic distortion measure. The best known achievable scheme is that of Berger and Tung [1], [2], although some examples where Berger-Tung compression can be outperformed are known [4]–[6]. In the Gaussian case, the Berger-Tung approach reduces to each encoder compressing its source using a standard point-to-point quantizer, followed by Slepian-Wolf [7] encoding. For the quadratic Gaussian case with $K = 2$, Wagner *et al.* [8] proved that this approach is optimal.

The importance of the quadratic-Gaussian distributed lossy compression problem has motivated researchers to design low-complexity encoding schemes that approach the performance of the Berger-Tung inner bound. This line of work was pioneered in [9], [10] and remains an active area of research, see, e.g., [11]–[13] and references therein. However, at a high level, the existing approaches for distributed source coding are either notably asymmetric in the rates they require from the encoders, as they rely on the lattice-based implementation of Wyner-Ziv coding [13], [14] and successive Wyner-Ziv coding [11], or specifically tailored to predefined correlation characteristics of the sources [10]. In general, the rate requirements in schemes that are based on Wyner-Ziv coding can be symmetrized by time-sharing between different compression/decompression orders [13]. Nevertheless, schemes using time-sharing have a few drawbacks. First, it requires the encoders and the decoders to use a larger number of codebooks, which complicates implementation. Second, it requires coordination between the distributed encoders, which is less crucial when time-sharing is not used. Finally, the compression block must be at least as long as the number of operation points that are time-shared.

In this work we propose a novel framework, *integer-forcing source coding*, for distributed lossy compression with *symmetric* rate and distortion requirements for all encoders.

This scheme does not incorporate time-sharing. As in previous works, our approach is based on standard quantization followed by lattice-based binning. However, in contrast to previous works, in the proposed framework the decoder first uses the bin indices for recovering linear combinations with integer coefficients of the quantized signals, and only then recovers the quantized signals themselves. The decoder is free to optimize the full-rank set of integer-valued coefficients such as to best exploit the correlations between the quantized signals. Choosing these coefficients appropriately results in performance that is close to that of a joint typicality decoder, with a substantially smaller computational burden. In fact, the only operations performed by the encoders are quantization and lattice-binning which corresponds to nearest neighbor decoding, whereas the decoder is only required to perform matrix multiplications and nearest neighbor decoding operations.

An important feature of the proposed approach is that it allows the system designer to trade-off performance and complexity. At one extreme, integer-forcing (IF) source coding can be implemented using high-dimensional nested lattices that have near-optimum quantization and channel coding performance. At the other extreme, IF source coding can be implemented with the low-complexity one-dimensional scaled integer lattice $\mathbb{Z}$, used as a quantizer as well as a channel code. Surprisingly, the rate loss from using the $1D$ lattice rather than "good" high-dimensional nested lattices, amounts to about 2 bits per sample per encoder, at any distortion level. At high resolution, where the compression rate is high, this loss of 2 bits is insignificant.

Implementing the $1D$ version of IF source coding only requires each encoder to reduce its observation modulo the lattice $2^R \Delta \mathbb{Z}$ and then quantize the obtained signal onto $\Delta \mathbb{Z}$, for some $\Delta > 0$ which depends on the required distortion. This simple operation can actually be implemented using an analog-to-digital converter (ADC).[1] The observation that at high resolution $1D$ IF source coding does not lose much w.r.t. the asymptotic performance achieved by Berger-Tung's compression may challenge the current paradigm of ADC design - rather than sample each source at a high rate and then compress it, why not sample at the compression rate to begin with? An idea in a similar spirit lies at the heart of compressed sensing [16], where the *number of samples* required to reconstruct a sparse signal is reduced according to its sparseness level. Here, the *number of sampled bits* required for reconstructing a source is reduced towards the source's rate-distortion function. The power consumption of an ADC depends on the number of bits it produces per second [17]. If the front end of the ADC includes an analog modulo operation, the ADC will need less quantization levels, i.e., less bits. Thus, if analog modulo reduction can be implemented efficiently, the IF approach may potentially lead to a more efficient ADC architectures.

IF source coding can be seen as the source coding dual of IF equalization [18]. IF equalization is a low complexity receiver architecture for the Gaussian MIMO channel. The IF receiver

first decodes integer linear combinations of the transmitted codewords, which is possible if all transmitted codewords were taken from the same linear code [19], and then solves these linear combinations for the transmitted codewords. In IF source coding, all encoders first quantize their observations to the desired distortion level, and then reduce them modulo the same lattice Λ.[2] The decoder receives the quantized modulo reduced signals. In order to form estimates of the original signals with the desired distortion level, it has to figure out what was the effect of the modulo reduction on each observation. Rather than doing this directly, it first tries to figure out what is the effect of reducing K linear combinations with integer-valued coefficients of the original signals modulo Λ, and only then extract the desired effects. See Figure 2.

The rest of the paper is organized as follows. In Section II we formally define the distributed lossy compression problem at hand, and introduce the performance benchmark we use throughout the paper which is based on the Berger-Tung inner bound. Basic lattice definitions and figures of merit are recalled in Section III, where standard results on lattice quantization are also reviewed. The IF source coding scheme is presented in Section IV, and the performance limits of the scheme are derived for the asymptotic case of high-dimensional "good" nested lattice codebooks. In Section V, a comparison between the performance of IF source coding and other known coding schemes is given for several scenarios. Applications of IF source coding to several communication problems that are not restricted to pure lossy compression are also given. In particular, we study the performance of a compress-and-forward scheme for relay networks where the compression is performed via IF source coding. We also study the problem of distributively transmitting K correlated Gaussian random variables over K parallel AWGN channels, and show that IF source coding can improve over standard approaches. In Section VI we describe and analyze the one-shot version of IF source coding, where the scaled $1D$ integer lattice is used for quantization and channel coding.

Notation. We denote scalars by lowercase letters, vectors by boldface lowercase letters and matrices by boldface uppercase letters, e.g., x, $\mathbf{x}$ and $\mathbf{X}$. Column vectors usually represent the spatial dimension whereas row vectors represent the time dimension. For example $\mathbf{x} = [x_1 \cdots x_K]^T \in \mathbb{R}^{K \times 1}$ may represent a Gaussian vector of correlated random variables, whereas $\mathbf{x}_k \in \mathbb{R}^{1 \times n}$ may represent n i.i.d. realizations of the random variable x_k. We denote the Euclidean norm of a vector by $\| \cdot \|$ and the absolute value of the determinant of a square matrix by $| \cdot |$. All variables in the paper are real-valued and all logarithms are to the base 2.

II. PROBLEM STATEMENT

We consider a distributed source coding setting with K encoding terminals and one decoder. Each of the K encoders has access to a vector $\mathbf{x}_k \in \mathbb{R}^n$ of n i.i.d. realizations of the random variable x_k, $k = 1, \ldots, K$. The random vector

---

[1]The analog modulo operation is actually already implemented, to some extent, in a class of ADCs called *folding ADCs* [15].

[2]If the quantization is performed by the $1D$ lattice $\Lambda_f = \Delta \mathbb{Z}$ and the coarse lattice used for binning is $\Lambda = 2^R \Delta \mathbb{Z}$, where 2^R is a positive integer, the order of the modulo and quantization operations can be switched.
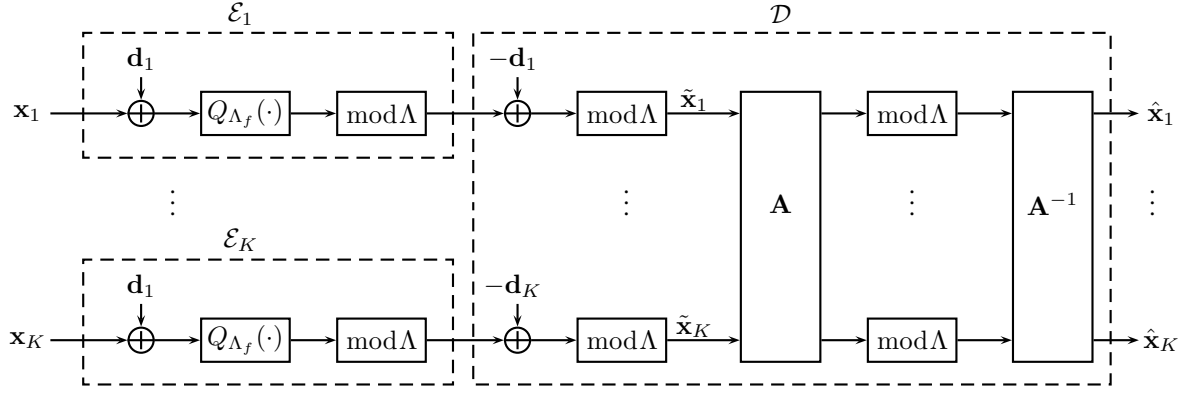
Fig. 2. A schematic overview of the integer-forcing source coding framework with the nested lattice pair $\Lambda \subset \Lambda_f$. Each encoder adds a dither $\mathbf{d}_k$ uniformly distributed over the Voronoi region of the fine lattice Λ_f and statistically independent of all other quantities, quantizes the dithered signal onto Λ_f and reduces the result modulo the coarse lattice Λ. The encoding rate is $\frac{1}{n} \log(\mathrm{Vol}(\Lambda)/\mathrm{Vol}(\Lambda_f))$. The decoder subtracts back the dithers and reduces the results modulo Λ (this $\mathrm{mod}\,\Lambda$ reduction is actually not necessary and is only illustrated for didactic purposes). Then, the decoder multiplies the signals by a full-rank integer matrix $\mathbf{A} \in \mathbb{Z}^{K \times K}$, reduces the results $\mathrm{mod}\,\Lambda$ and multiplies by $\mathbf{A}^{-1}$ to form the estimates $\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_K$.

$\mathbf{x} = [x_1 \; \cdots \; x_K]^T$ is assumed Gaussian with zero mean and covariance matrix

$$\mathbf{K_{xx}} \triangleq \mathbb{E}(\mathbf{xx}^T).$$

Each encoder maps its observation $\mathbf{x}_k$ to an index using the encoding function

$$\mathcal{E}_k \; : \; \mathbb{R}^n \to \; \{1, \ldots, 2^{nR_k}\},$$

and sends the index to the decoder.

The decoder is equipped with K decoding functions

$$\mathcal{D}_k \; : \; \{1, \ldots, 2^{nR_1}\} \times \cdots \times \{1, \ldots, 2^{nR_K}\} \; \to \; \mathbb{R}^n,$$

for $k = 1, \ldots, K$. Upon receiving K indices, one from each terminal, the decoder generates estimates

$$\hat{\mathbf{x}}_k = \mathcal{D}_k \left(\mathcal{E}_1(\mathbf{x}_1), \ldots, \mathcal{E}_K(\mathbf{x}_K) \right), \quad k = 1, \ldots, K.$$

A rate-distortion vector $(R_1, \ldots, R_K, d_1, \ldots, d_K)$ is achievable if there exist encoding functions $\mathcal{E}_1, \ldots, \mathcal{E}_K$ and decoding functions $\mathcal{D}_1, \ldots, \mathcal{D}_K$ such that

$$\frac{1}{n}\mathbb{E}\left(\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2\right) \leq d_k, \tag{1}$$

for all $k = 1, \ldots, K$. Let $\mathbf{X} \triangleq [\mathbf{x}_1^T \; \cdots \; \mathbf{x}_K^T]^T$. A conditionally *unbiased* rate-distortion vector $(R_1, \ldots, R_K, d_1, \ldots, d_K)$ is achievable if in addition to (1), the condition

$$\mathbb{E}(\hat{\mathbf{x}}_k | \mathbf{X}) = \mathbf{x}_k, \quad k = 1, \ldots, K \tag{2}$$

is satisfied for any realization of $\mathbf{X}$. Note that this condition is equivalent to

$$\mathbb{E}(\mathbf{x}_k - \hat{\mathbf{x}}_k | \mathbf{X}) = 0, \quad k = 1, \ldots, K.$$

Although condition (2) is not as common in the literature as condition (1), in this paper we restrict attention to the conditionally unbiased case, i.e., we impose condition (2). Several applications of interest require the estimates formed by the decoder to be conditionally unbiased. For instance, consider a communication scenario where distributed antenna terminals observe noisy linear combinations of the signals transmitted by several encoders and want to forward a compressed version

of these signals to a central processor that needs to decode the transmitted messages. In such a scenario it is most convenient to treat the quantization noise as an additive one, meaning that it is statistically independent of the signals that are being quantized. This amounts to requiring condition (2). Moreover, when the conditionally unbiased requirement (2) is not essential to the application at hand, one can always perform minimum mean-squared estimation of $\mathbf{X}$ from $\hat{\mathbf{X}}$ and further reduce the MSE distortion.

We further focus on the symmetric case where $R_1 = \cdots = R_K = R$ and $d_1 = \cdots = d_k = d$. We do this for three reasons. First, such a symmetry constraint naturally arises in many applications, where the coding burden has to be equally split between the distributed encoders. Second, this allows for a simpler description of the proposed coding scheme and the rate-distortion region it achieves. Finally, in an asymmetric setting there exist several examples where structured binning outperforms the standard approach of Berger-Tung compression [4]–[6]. Focusing on the symmetric case eliminates the possibility of such examples that are, to some extent, skewed towards using structured binning. Nevertheless, we stress that the scheme proposed in this paper is not restricted to the symmetric case, and can be easily extended to achieve asymmetric rate-distortion vectors by using a more complicated *chain* of nested lattices, rather than the nested lattice *pair* we use in the sequel.

Finding the full rate-distortion region, i.e., the set of all achievable rate-distortion vectors, for the described setup is an open problem for $K > 2$. For $K = 2$, Wagner *et al.* [8] showed that the Berger-Tung approach is optimal. This approach consists of quantizing each source using standard single-source rate-distortion theory with a Gaussian test channel, and then using Slepian-Wolf encoding for compressing the quantization indices. For $K > 2$ it is now known that the Berger-Tung approach does not attain the full rate-distortion region (see e.g. [5]). However, to the best of our knowledge, it is not known whether the Berger-Tung inner bound is loose for the symmetric case. In the absence of a better known coding scheme, we take the symmetric rate from Berger-Tung's inner

bound as our benchmark. More specifically, the sum-rate in Berger-Tung's inner bound is given by

$$\sum_{k=1}^{K} R_k \geq I(\mathbf{x}; \mathbf{u}), \qquad (3)$$

where $\mathbf{u} = [u_1 \cdots u_K]^T$ is a vector of auxiliary random variables that satisfy the set of Markov chains

$$u_k - x_k - (\{x_j, u_j\}_{j \neq k})$$

and such that there exist functions $\hat{x}_k(u_1, \ldots, u_K)$ satisfying $\mathbb{E}(x_k - \hat{x}_k)^2 < d_k$ for all $k = 1, \ldots, K$. Optimizing over $\mathbf{u}$ is a difficult task. A common and natural choice in the quadratic-Gaussian case is taking

$$u_k = \alpha_k x_k + w_k, \quad k = 1, \ldots, K, \qquad (4)$$

where $w_1, \ldots, w_K$ are statistically independent zero mean Gaussian random variables that are also independent of $\mathbf{x}$, and $\alpha_1, \ldots, \alpha_K$ are some constants [20]. Such a choice was shown to be optimal for $K = 2$ [8], but may be suboptimal for larger dimensions. Since we are after conditionally unbiased estimates for the K components of $\mathbf{x}$, we set $\alpha_k = 1$, $w_k \sim \mathcal{N}(0, d)$ and $\hat{x}_k(u_1, \ldots, u_K) = u_k$ for all $k = 1, \ldots, K$. Substituting this choice in (3) gives

$$\sum_{k=1}^{K} R_k \geq \frac{1}{2} \log \frac{|\mathbf{K}_{\mathbf{xx}} + d\mathbf{I}|}{|d\mathbf{I}|}$$
$$= \frac{1}{2} \log \left| \mathbf{I} + \frac{1}{d} \mathbf{K}_{\mathbf{xx}} \right|. \qquad (5)$$

This sum-rate is achievable using Berger-Tung compression. In this paper we are interested in the symmetric rate-distortion region. To this end, we take (5) normalized by K as our benchmark

$$R_{\text{bench}}^{\text{BT}}(d) \triangleq \frac{1}{2K} \log \left| \mathbf{I} + \frac{1}{d} \mathbf{K}_{\mathbf{xx}} \right|. \qquad (6)$$

Note that $R_{\text{bench}}^{\text{BT}}(d)$ is not a lower bound on the minimal symmetric rate-distortion function achieved by Berger-Tung compression, as our choice of $\mathbf{u}$ is not necessarily the best one. It is also not an upper bound on the minimal symmetric rate-distortion function achieved by Berger-Tung compression, as the symmetric rate with our choice of $\mathbf{u}$ may not be dominated by the sum-rate constraint.

III. PRELIMINARIES

In this section we recall several lattice properties that will be useful in the sequel and review the concept of dithered lattice quantization.

A lattice Λ is a discrete subgroup of $\mathbb{R}^n$ which is closed under reflection and real addition. We denote the nearest neighbor quantizer associated with the lattice Λ by

$$Q_\Lambda(\mathbf{y}) = \underset{\mathbf{t} \in \Lambda}{\arg\min} \|\mathbf{y} - \mathbf{t}\|. \qquad (7)$$

The basic Voronoi region of Λ, denoted by $\mathcal{V}$, is the set of all points in $\mathbb{R}^n$ which are quantized to the zero vector, where

ties in (7) are broken in a systematic manner. The modulo operation returns the quantization error w.r.t. the lattice,

$$[\mathbf{y}] \bmod \Lambda = \mathbf{y} - Q_\Lambda(\mathbf{y})$$

and satisfies the property

$$[a[\mathbf{y}] \bmod \Lambda] \bmod \Lambda = [a\mathbf{y}] \bmod \Lambda \qquad (8)$$

for any $a \in \mathbb{Z}$ and $\mathbf{y} \in \mathbb{R}^n$. This property will be used extensively in the sequel. The second moment of Λ is defined as

$$\sigma^2(\Lambda) = \frac{1}{n} \frac{1}{\text{Vol}(\mathcal{V})} \int_{\mathbf{u} \in \mathcal{V}} \|\mathbf{u}\|^2 d\mathbf{u},$$

where $\text{Vol}(\mathcal{V})$ is the volume of $\mathcal{V}$. The effective radius of a lattice $r_{\text{eff}}(\Lambda)$ is defined as the radius of an n-dimensional ball whose volume equals $\text{Vol}(\mathcal{V})$.

The lattice Λ can be used for quantizing continuous sources. In particular, an encoder which is interested in conveying a source $\mathbf{y} \in \mathbb{R}^n$ to a decoder can compute $Q_\Lambda(\mathbf{y})$, which is a lattice point in Λ, and send a description of this point to the decoder. The quantization error of such a scheme is $\mathbf{e} = \mathbf{y} - Q_\Lambda(\mathbf{y})$, which is a deterministic function of $\mathbf{y}$. Recall that in this paper we are interested in encoder/decoder pairs that produce conditionally unbiased estimates of the source, which is clearly not the case for a standard lattice quantizer. This may be overcome by allowing the encoder and decoder to use common randomness. Let $\mathbf{d} \sim \text{Unif}(\mathcal{V})$ be a random dither vector uniformly distributed over $\mathcal{V}$ and statistically independent of $\mathbf{y}$, known to both the encoder and the decoder. The *dithered lattice quantizer* associated with the lattice Λ computes $Q_\Lambda(\mathbf{y} + \mathbf{d})$ and sends a description of the obtained lattice point to the decoder. The decoder produces the estimate

$$\hat{\mathbf{y}} = Q_\Lambda(\mathbf{y} + \mathbf{d}) - \mathbf{d}$$
$$= \mathbf{y} + Q_\Lambda(\mathbf{y} + \mathbf{d}) - (\mathbf{y} + \mathbf{d})$$
$$= \mathbf{y} - [\mathbf{y} + \mathbf{d}] \bmod \Lambda.$$

The Crypto Lemma [21, Lemma 1] ensures that the estimation error $-[\mathbf{y} + \mathbf{d}] \bmod \Lambda$ is statistically independent of $\mathbf{y}$ and is uniformly distributed over $\mathcal{V}$. The symmetry of the Voronoi region $\mathcal{V}$ guarantees that the estimation error has the same distribution as $\mathbf{d}$ and has zero mean. Thus, $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{d}$ in distribution, and is a conditionally unbiased estimate of $\mathbf{y}$. Clearly, the average MSE distortion attained by dithered lattice quantization is given by

$$\frac{1}{n} \mathbb{E} \left(\|\mathbf{y} - \hat{\mathbf{y}}\|^2 \right) = \frac{1}{n} \mathbb{E}(\|\mathbf{d}\|^2) = \sigma^2(\Lambda).$$

Of course, dithered lattice quantization, as described above, requires an infinite rate as there is an infinite number of points in Λ. This can be handled using an entropy coded dithered quantizer (ECDQ) [22]–[24], or a nested lattice codebook [13]. In this work we take the latter approach.

The following definitions characterize the lattice "goodness" properties needed in this paper.

Definition 1 (Goodness for MSE quantization): A lattice Λ, or more precisely, a sequence of lattices with growing

dimension n, is said to be good for MSE quantization if[3]

$$\lim_{n \to \infty} \sigma^2(\Lambda) = \lim_{n \to \infty} \frac{r_{\text{eff}}^2(\Lambda)}{n}.$$

Definition 2 (Semi-norm ergodic noise): We say that a random noise vector $\mathbf{z}$, or more precisely, a sequence of random noise vectors with growing dimension n, with (finite) effective variance $\sigma_{\mathbf{z}}^2 \triangleq \mathbb{E}\|\mathbf{z}\|^2/n$, is *semi norm-ergodic* if for any $\epsilon > 0$, $\delta > 0$ and n large enough

$$\Pr\left(\|\mathbf{z}\| > \sqrt{(1+\delta)n\sigma_{\mathbf{z}}^2} \right) \le \epsilon. \tag{9}$$

Note that by the law of large numbers, any i.i.d. noise is semi norm-ergodic.

The next Lemma restates Corollary 2 from [25] to fit our purposes.

Lemma 1: Let $\mathbf{d}_1, \cdots, \mathbf{d}_K$ be statistically independent random dither vectors, each uniformly distributed over the Voronoi region $\mathcal{V}$ of a lattice Λ that is good for MSE quantization. Let $\mathbf{z}$ be an i.i.d. random vector statistically independent of $\{\mathbf{d}_1, \cdots, \mathbf{d}_K\}$. Any deterministic linear combination of $\mathbf{d}_1, \cdots, \mathbf{d}_K, \mathbf{z}$ is semi norm-ergodic.

Definition 3 (Goodness for channel coding): A lattice Λ, or more precisely, a sequence of lattices with growing dimension n, is said to be good for channel coding if for any $0 < \delta < 1$ and any n-dimensional semi norm-ergodic vector $\mathbf{z}$ with zero mean and effective variance $\mathbb{E}\|\mathbf{z}\|^2/n < (1 - \delta)r_{\text{eff}}^2(\Lambda)/n$

$$\lim_{n \to \infty} \Pr\left(\mathbf{z} \notin \mathcal{V} \right) = 0.$$

A lattice Λ is said to be nested in Λ_f if $\Lambda \subseteq \Lambda_f$. The coding scheme presented in this paper utilizes a pair of nested lattices such that the fine lattice Λ_f is good for MSE quantization and the coarse lattice Λ is good for channel coding. An ensemble for drawing pairs of nested lattices that satisfy these goodness properties is described in [25],[4] and the existence of lattice pairs with slightly more demanding "goodness" requirements was shown in [5], [26]. A nested lattice code $\mathcal{C} = \Lambda_f \cap \mathcal{V}$ with rate

$$R = \frac{1}{n} \log\left(\frac{\text{Vol}(\Lambda)}{\text{Vol}(\Lambda_f)} \right) = \frac{1}{2} \log\left(\frac{r_{\text{eff}}^2(\Lambda)}{r_{\text{eff}}^2(\Lambda_f)} \right) \tag{10}$$

is associated with the nested lattice pair.

Before describing the integer-forcing source coding scheme, let us illustrate how the codebook $\mathcal{C}$ described above can be used for compressing n samples of a single memoryless Gaussian source $Y \sim \mathcal{N}(0, P)$ with distortion d. Assume that

[3]Note that our condition for MSE goodness is equivalent to the more commonly used condition $\sigma^2(\Lambda)/\text{Vol}(\mathcal{V})^{2/n} \to 1/(2\pi e)$ since the volume of a unit n-dimensional ball grows like $(2\pi e/n)^{n/2}$.

[4]In [25] the definition of goodness for channel coding was weaker than that needed here. In particular, only the existence of lattices that achieve a vanishing error probability under *coset* nearest neighbor decoding in the present of semi-norm ergodic noise was proved. However, a more careful inspection of the derivation in [25] reveals that the probability of decoding an erroneous point in the correct coset also vanishes with the dimension n for the choice of lattice parameters made in [25]. Thus, the existence of pairs of nested lattices such that both fine and coarse lattices are good for MSE quantization and channel coding follows.

the fine lattice Λ_f, which is good for MSE quantization, has second moment $\sigma^2(\Lambda_f) = d$. This implies that $r_{\text{eff}}^2(\Lambda_f)/n \to d$. The coarse lattice Λ, which is good for AWGN channel coding, has effective radius $r_{\text{eff}}^2(\Lambda) = n(P + d + \epsilon)$, for some arbitrarily small $\epsilon > 0$. A dither $\mathbf{d}$ uniformly distributed over $\mathcal{V}_f$ is known to both the encoder and the decoder. The encoder computes

$$[Q_{\Lambda_f}(\mathbf{y} + \mathbf{d})] \mod \Lambda \in \mathcal{C},$$

and sends its index to the decoder. The decoder computes

$$\hat{\mathbf{y}} = \left[[Q_{\Lambda_f}(\mathbf{y} + \mathbf{d})] \mod \Lambda - \mathbf{d} \right] \mod \Lambda$$
$$\overset{(i.d.)}{=} [\mathbf{y} + \mathbf{d}] \mod \Lambda$$
$$\overset{(w.h.p.)}{=} \mathbf{y} + \mathbf{d} \tag{11}$$

where $\overset{(i.d.)}{=}$ stands for equality in distribution and $\overset{(w.h.p.)}{=}$ for equality with high probability. The equality (11) follows from the fact that the random vector $\mathbf{y} + \mathbf{d}$ is semi-norm ergodic due to Lemma 1 and has effective variance $\mathbb{E}(\|\mathbf{y} + \mathbf{d}\|^2)/n = P + d$. Since Λ is good for channel coding and $\mathbb{E}\|\mathbf{y} + \mathbf{d}\|^2/n < r_{\text{eff}}^2(\Lambda)/n$, the probability that $Q_\Lambda(\mathbf{y} + \mathbf{d}) \neq \mathbf{0}$ vanishes, and hence, $[\mathbf{y} + \mathbf{d}] \mod \Lambda \overset{(w.h.p.)}{=} \mathbf{y} + \mathbf{d}$. Thus, with high probability

$$\frac{1}{n}\mathbb{E}(\|\mathbf{y} - \hat{\mathbf{y}}\|^2) = \frac{1}{n}\mathbb{E}(\|\mathbf{d}\|^2) = d,$$

as desired. The required rate for achieving this distortion is

$$\begin{aligned} R(d) &= \frac{1}{2} \log\left(\frac{r_{\text{eff}}^2(\Lambda)}{r_{\text{eff}}^2(\Lambda_f)} \right) \\ &= \frac{1}{2} \log\left(\frac{n(P + d + \epsilon)}{nd} \right) \\ &= \frac{1}{2} \log\left(1 + \frac{P + \epsilon}{d} \right) \end{aligned} \tag{12}$$

where the additional $+1$ inside the logarithm, w.r.t. the standard Gaussian rate-distortion function, is a consequence of our requirement that the reconstruction $\hat{\mathbf{y}}$ forms a conditionally unbiased estimate of $\mathbf{y}$. In fact, we can eliminate this term by performing an additional Wiener estimation step on $\hat{\mathbf{y}}$, at the expense of introducing bias [24].

IV. INTEGER-FORCING SOURCE CODING

In the IF distributed source coding scheme all encoders use the same nested lattice codebook $\mathcal{C} = \Lambda_f \cap \mathcal{V}$, constructed from the nested lattice pair $\Lambda \subset \Lambda_f$, with rate

$$R = \frac{1}{2} \log\left(\frac{r_{\text{eff}}^2(\Lambda)}{r_{\text{eff}}^2(\Lambda_f)} \right).$$

As in the previous section, the fine lattice Λ_f is good for MSE quantization with $\sigma^2(\Lambda_f) = d$ whereas the coarse lattice Λ is good for channel coding. All encoders employ a similar encoding operation. The kth encoder uses a dither $\mathbf{d}_k$, statistically independent of everything else and uniformly distributed over $\mathcal{V}_f$, and employs dithered quantization of $\mathbf{x}_k$ onto Λ_f. Then, it reduces the obtained lattice point modulo the coarse lattice Λ and sends nR bits describing the index of

the resulting point to the decoder. Specifically, the kth encoder conveys the index corresponding to the point

$$\left[Q_{\Lambda_f}(\mathbf{x}_k + \mathbf{d}_k)\right] \bmod \Lambda$$

to the decoder.

The decoder first subtracts back the dithers from each of the reconstructed signals and reduces the results modulo Λ, giving rise to

$$
\begin{aligned}
\tilde{\mathbf{x}}_k &= \left[\left[Q_{\Lambda_f}(\mathbf{x}_k + \mathbf{d}_k)\right] \bmod \Lambda - \mathbf{d}_k\right] \bmod \Lambda \\
&= \left[\mathbf{x}_k + \left[Q_{\Lambda_f}(\mathbf{x}_k + \mathbf{d}_k)\right] \bmod \Lambda - (\mathbf{x}_k + \mathbf{d}_k)\right] \bmod \Lambda \\
&\overset{(i.d.)}{=} \left[\mathbf{x}_k + \mathbf{d}_k\right] \bmod \Lambda
\end{aligned}
\tag{13}
$$

If the coarse lattice Λ is chosen such that its effective radius is large enough, the modulo operation in (13) would have no effect on $\mathbf{x}_k + \mathbf{d}_k$, and the decoder would have estimates of each $\mathbf{x}_k$ with average MSE of d, as desired. However, the encoding rate grows with $r_{\text{eff}}^2(\Lambda)$, and we would therefore prefer to choose it as small as possible.

The key idea behind IF source coding is that if the elements of $\mathbf{x}$ are correlated, then linear combinations of $\{\mathbf{x}_k + \mathbf{d}_k\}_{k=1}^K$ with integer-valued coefficients may have smaller effective variances than the original signals. The IF decoder therefore first estimates K integer linear combinations of $\{\mathbf{x}_k + \mathbf{d}_k\}_{k=1}^K$, and then uses these estimates for estimating the desired signals. Using this approach, $r_{\text{eff}}^2(\Lambda)$ should only be greater than the largest effective variance among the K linear combinations. When the entries of $\mathbf{x}$ are sufficiently correlated, and the integer-valued coefficients are chosen appropriately, this may significantly reduce the required encoding rate.

Let $\mathbf{X} = [\mathbf{x}_1^T \ \cdots \ \mathbf{x}_K^T]^T$, $\mathbf{D} = [\mathbf{d}_1^T \ \cdots \ \mathbf{d}_K^T]^T$ and $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1^T \ \cdots \ \tilde{\mathbf{x}}_K^T]^T$. Using this notation, the decoder has access to

$$\tilde{\mathbf{X}} = [\mathbf{X} + \mathbf{D}] \bmod \Lambda,$$

where the notation $\bmod \Lambda$ is to be understood as reducing *each row* of the obtained matrix modulo the coarse lattice. The decoder chooses a full-rank integer-valued matrix $\mathbf{A} \in \mathbb{Z}^{K \times K}$ and computes

$$
\begin{aligned}
\widehat{\mathbf{AX}} &\triangleq \left[\mathbf{A}\tilde{\mathbf{X}}\right] \bmod \Lambda \\
&= \left[\mathbf{A}\left[\mathbf{X} + \mathbf{D}\right] \bmod \Lambda\right] \bmod \Lambda \\
&= \left[\mathbf{A}(\mathbf{X} + \mathbf{D})\right] \bmod \Lambda
\end{aligned}
\tag{14}
$$

where (14) follows from the modulo property (8).

Let $\mathbf{a}_k^T$ be the kth row of the matrix $\mathbf{A}$. The random vector $\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})$ satisfies the conditions of Lemma 1 as $\mathbf{a}_k^T\mathbf{X}$ is an i.i.d. Gaussian vector and each of the statistically independent dithers $\mathbf{d}_1, \ldots, \mathbf{d}_K$ is uniformly distributed over the Voronoi region of a lattice that is good for MSE quantization. Therefore, $\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})$ is semi-norm ergodic. It follows from the goodness of Λ for channel coding that if

$$\frac{\mathbb{E}\left(\|\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})\|^2\right)}{n} < \frac{r_{\text{eff}}^2(\Lambda)}{n}$$

then for n large enough

$$\left[\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})\right] \bmod \Lambda \overset{(w.h.p.)}{=} \mathbf{a}_k^T(\mathbf{X} + \mathbf{D}).$$

Moreover, if this holds for all $k = 1, \ldots, K$, i.e., if

$$\max_{k=1,\ldots,K} \frac{\mathbb{E}\left(\|\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})\|^2\right)}{n} < \frac{r_{\text{eff}}^2(\Lambda)}{n}$$

then for n large enough

$$\widehat{\mathbf{AX}} \overset{(w.h.p.)}{=} \mathbf{A}(\mathbf{X} + \mathbf{D}). \tag{15}$$

Noting that

$$\frac{\mathbb{E}\left(\|\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})\|^2\right)}{n} = \mathbf{a}_k^T(\mathbf{K_{xx}} + d\mathbf{I})\mathbf{a}_k,$$

this implies that for (15) to hold, it suffices to set

$$\frac{r_{\text{eff}}^2(\Lambda)}{n} = \max_{k=1,\ldots,K} \mathbf{a}_k^T(\mathbf{K_{xx}} + d\mathbf{I})\mathbf{a}_k + \epsilon$$

for some arbitrarily small $\epsilon > 0$, which corresponds to a rate of

$$R = \frac{1}{2}\log\left(\frac{\max_{k=1,\ldots,K} \mathbf{a}_k^T(\mathbf{K_{xx}} + d\mathbf{I})\mathbf{a}_k + \epsilon}{d}\right).$$

The decoder proceeds by computing

$$\hat{\mathbf{X}} = \mathbf{A}^{-1}\widehat{\mathbf{AX}} \overset{(w.h.p.)}{=} \mathbf{X} + \mathbf{D},$$

which is (w.h.p.) a conditionally unbiased estimate of $\mathbf{X}$ with average MSE distortion d per component. The next theorem summarizes the performance of IF source coding.

Theorem 1 (Performance of IF source coding): For any distortion $d > 0$ and any choice of $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_K]^T \in \mathbb{Z}^{K \times K}$, there exists a (sequence of) nested lattice pair(s) $\Lambda \subset \Lambda_f$ such that IF source coding can achieve any rate satisfying

$$R > R_{\text{IF}}(\mathbf{A}, d) \triangleq \frac{1}{2}\log\left(\max_{k=1,\ldots,K} \mathbf{a}_k^T\left(\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}}\right)\mathbf{a}_k\right).$$

For the optimal choice of $\mathbf{A}$, IF source coding can achieve any rate satisfying

$$R > R_{\text{IF}}(d) \triangleq \frac{1}{2}\log\left(\min_{\substack{\mathbf{A} \in \mathbb{Z}^{K \times K} \\ \det(\mathbf{A}) \neq 0}} \max_{k=1,\ldots,K} \mathbf{a}_k^T\left(\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}}\right)\mathbf{a}_k\right).$$

The matrix $\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}}$ is symmetric and positive definite, and therefore it admits a Cholesky decomposition

$$\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}} = \mathbf{F}\mathbf{F}^T, \tag{16}$$

where $\mathbf{F}$ is a lower triangular matrix with strictly positive entries. With this notation,

$$R_{\text{IF}}(d) = \frac{1}{2}\log\left(\min_{\substack{\mathbf{A} \in \mathbb{Z}^{K \times K} \\ \det(\mathbf{A}) \neq 0}} \max_{k=1,\ldots,K} \|\mathbf{F}\,\mathbf{a}_k\|^2\right). \tag{17}$$

Denote by $\Lambda(\mathbf{F}^T)$ the K dimensional lattice spanned by the matrix $\mathbf{F}^T$, i.e.,

$$\Lambda(\mathbf{F}^T) \triangleq \left\{\mathbf{F}^T\mathbf{a} \ : \ \mathbf{a} \in \mathbb{Z}^K\right\}.$$

It follows that the problem of finding the optimal matrix $\mathbf{A}$ is equivalent to finding the K shortest linearly independent vectors of $\Lambda(\mathbf{F}^T)$. Although this problem is NP-hard in general, its solution can be efficiently approximated using the LLL algorithm [27], whose running time is polynomial.

Moreover, we can express the rate-distortion function achieved by IF source coding using the *successive minima* of the lattice $\Lambda(\mathbf{F}^T)$.

Definition 4 (Successive minima): Let $\Lambda(\mathbf{G})$ be the lattice spanned by the full-rank matrix $\mathbf{G} \in \mathbb{R}^{K \times K}$. For $k = 1, \ldots, K$, we define the kth successive minimum as

$$\lambda_k(\mathbf{G}) \triangleq \inf \left\{ r \; : \; \dim \left(\text{span} \left(\Lambda(\mathbf{G}) \bigcap \mathcal{B}(\mathbf{0}, r) \right) \right) \geq k \right\}$$

where $\mathcal{B}(\mathbf{0}, r) = \left\{ \mathbf{x} \in \mathbb{R}^K \; : \; \|\mathbf{x}\| \leq r \right\}$ is the closed ball of radius r around $\mathbf{0}$. In words, the kth successive minimum of a lattice is the minimal radius of a ball centered around $\mathbf{0}$ that contains k linearly independent lattice points.

Using Definition 4 and (17), the IF rate-distortion function is given by

$$R_{\text{IF}}(d) = \frac{1}{2} \log \left(\lambda_K^2(\mathbf{F}^T) \right), \qquad (18)$$

where the dependence of the r.h.s. on d is through the matrix $\mathbf{F}$ defined in (16).

Next, we show in Lemma 2 that the performance of IF source coding, in the symmetric setting considered, is inferior to the Berger-Tung benchmark, i.e., $R_{\text{IF}}(d) \geq R_{\text{bench}}^{\text{BT}}(d)$. We will need the simple following proposition.

Proposition 1: For a lattice spanned by some full rank matrix $\mathbf{G} \in \mathbb{R}^{K \times K}$

$$|\mathbf{G}| \leq \prod_{k=1}^{K} \lambda_k(\mathbf{G})$$

Proof: Let $\mathbf{a}_1, \ldots, \mathbf{a}_K \in \mathbb{Z}^K$ be K linearly independent vectors such that $\lambda_k(\mathbf{G}) = \|\mathbf{G}\mathbf{a}_k\|$ for all $k = 1, \ldots, K$, and let $\mathbf{A} = [\mathbf{a}_1 \; \cdots \; \mathbf{a}_K]$. Since all entries of $\mathbf{A}$ are integer-valued we must have $|\mathbf{A}| \geq 1$, and therefore

$$|\mathbf{G}| \leq |\mathbf{G}| \, |\mathbf{A}| = |\mathbf{G}\mathbf{A}|$$
$$= |[\mathbf{G}\mathbf{a}_1 \; \cdots \; \mathbf{G}\mathbf{a}_K]| \leq \prod_{k=1}^{K} \|\mathbf{G}\mathbf{a}_k\|$$
$$= \prod_{k=1}^{K} \lambda_k(\mathbf{G}).$$

$\blacksquare$

Lemma 2: For any $d > 0$ and for any choice of full-rank $\mathbf{A} \in \mathbb{Z}^{K \times K}$ we have

$$\frac{1}{2} \log \left(\max_{k=1,\ldots,K} \mathbf{a}_k^T (\mathbf{I} + \frac{1}{d}\mathbf{K}_{\mathbf{xx}}) \mathbf{a}_k \right) \geq \frac{1}{2K} \log \left| \mathbf{I} + \frac{1}{d}\mathbf{K}_{\mathbf{xx}} \right|,$$

and therefore, in the considered symmetric setting, the rate-distortion function $R_{\text{IF}}(d)$ of IF source coding is never smaller than the benchmark $R_{\text{bench}}^{\text{BT}}(d)$.

Proof: Let $\mathbf{F}$ be as defined in (16). For the optimal choice of $\mathbf{A}$ and for any $d > 0$ we have

$$\frac{1}{2} \log \left(\max_{k=1,\ldots,K} \mathbf{a}_k^T (\mathbf{I} + \frac{1}{d}\mathbf{K}_{\mathbf{xx}}) \mathbf{a}_k \right) = \frac{1}{2} \log \left(\lambda_K^2(\mathbf{F}^T) \right)$$

$$\geq \frac{1}{2} \frac{1}{K} \sum_{k=1}^{K} \log \left(\lambda_k^2(\mathbf{F}^T) \right) \qquad (19)$$

$$= \frac{1}{2K} \log \left(\prod_{k=1}^{K} \lambda_k^2(\mathbf{F}^T) \right)$$

$$\geq \frac{1}{2K} \log \left(|\mathbf{F}|^2 \right) \qquad (20)$$

$$= \frac{1}{2K} \log \left| \mathbf{I} + \frac{1}{d}\mathbf{K}_{\mathbf{xx}} \right|, \qquad (21)$$

where (19) follows from the monotonicity of $\lambda_k(\mathbf{F}^T)$ in k along with the monotonicity of the logarithm function, (20) follows from Proposition 1 and (21) follows from (16). $\blacksquare$

As discussed in Section II, in an asymmetric problem setting, structured binning may result in a better rate-distortion region than the one obtained by Berger-Tung compression. Lemma 2 shows that under the symmetric setup, at least with IF source coding, this may not be the case. Nevertheless, the complexity reduction obtained by using IF source coding rather than Berger-Tung compression makes it an attractive candidate for practical implementation. Moreover, as we shall see in Section VI, a one-shot version of IF source coding can be easily derived and analyzed. Although one-shot versions of Berger-Tung compression were also considered in [28] and an inner bound was derived, it is unclear how to interpret this inner bound for the problem at hand.

Remark 1: The crucial element in the IF source coding scheme is that all encoders reduce their quantized signals modulo the same coarse lattice. The modulo reduction plays the role of binning. Theoretically, each encoder can first reduce its observation $\bmod \Lambda$ and only then quantize it using a quantizer designed for the modulo reduced source [24]. No nesting is required between the quantizer and the coarse lattice. This results in the decoder receiving the signals $\tilde{\mathbf{x}}_k = [\mathbf{x}_k] \bmod \Lambda + \mathbf{d}_k$, $k = 1, \ldots, K$, where $\mathbf{d}_k$ is quantization noise. The decoder can proceed to compute $\widehat{\mathbf{A}\mathbf{X}}$ as described above. The difficulty with such an implementation is that the quantizer needs to be matched to the modulo reduced source, which requires some sort of (high-dimensional) entropy coding. As we shall see in Section VI-A, in the $1D$ version of IF source coding, where the coarse lattice as well as the quantizer are scaled integer lattices, the modulo reduction can precede quantization without increasing complexity.

Remark 2: Another implementation issue to consider is the goodness requirements on Λ. When Λ is used for modulation over the AWGN channel, it suffices to require that Λ is good under *coset* nearest neighbor decoding. This means that Λ is split into cosets, usually using a coarse lattice nested inside it, and the decoder only needs to choose the coset the transmitted point belongs to. As a result, when coding for the AWGN channel is considered, a construction A lattice [29], [30] with a linear codebook of small prime cardinality p suffices to achieve a vanishing error probability. In such a construction,

the minimum distance is limited by p, and the error probability for decoding the actual point transmitted, rather than the coset, cannot vanish with the dimension. However, all pairs of points with nonincreasing (as a function of n, the code dimension) Euclidean distance belong to the same coset, and therefore such a lattice is still good for coset nearest neighbor decoding.

In IF source coding, the decoder needs to decode the *actual* lattice point of Λ closest to $\mathbf{a}_k^T(\mathbf{X} + \mathbf{D})$, rather than just its coset. Therefore, construction A lattices obtained from a linear codebook with small p do not suffice in order to achieve a vanishing error probability. However, one can still achieve a very small error probability, though not vanishing with the dimension, using standard Construction A lattices with moderate values of p. See Section VI for further discussion of implementation issues.

V. EXAMPLES AND APPLICATIONS

This section provides several examples that demonstrate the performance of IF source coding, along with applications and communication scenarios where IF source coding is advantageous. The section consists of three parts. First we compare the performance of IF source coding to that of a naive distributed compression scheme that ignores the correlation between the sources and to the Berger-Tung benchmark. Then, we use IF source coding as a building block in a Gaussian layered relay network, and demonstrate its advantages compared to other known low complexity schemes. Finally, we show how the idea behind IF source coding can be extended to form a signal-to-noise ratio (SNR) independent joint source channel coding scheme, whose distortion decreases as the SNR improves.

A. Examples

In this subsection we evaluate the minimal symmetric rate needed in order to achieve a conditionally unbiased average MSE of d for two schemes:

1) IF source coding - this rate is given in Theorem 1.
2) Compressing each source using standard rate-distortion theory without exploiting the correlations between the sources - this rate is given by

$$R_{\text{naive}}(d) = \max_{k=1,\ldots,K} \frac{1}{2} \log \left(1 + \frac{\mathbf{K}_{\mathbf{xx}}(k,k)}{d} \right), \quad (22)$$

and is identical to the rate obtained using IF source coding with the choice $\mathbf{A} = \mathbf{I}$.

We also compare these rates to the Beger-Tung benchmark $R_{\text{bench}}^{\text{BT}}(d)$ (6).

Example 1 (Integer decomposable covariance matrix): As a first example, consider the case where $\mathbf{x}$ is a Gaussian source with zero mean and covariance matrix $\mathbf{K}_{\mathbf{xx}} = \mathbf{B}^{-1}\mathbf{B}^{-T}$ for some full-rank integer matrix $\mathbf{B} \in \mathbb{Z}^{K \times K}$ with determinant $|\mathbf{B}| = 1$.

The Berger-Tung benchmark symmetric rate-distortion function is given by

$$\begin{aligned} R_{\text{bench}}^{\text{BT}}(d) &= \frac{1}{2K} \log \left| \mathbf{I} + \frac{1}{d}\mathbf{B}^{-1}\mathbf{B}^{-T} \right| \\ &= \frac{1}{2K} \left(\log |\mathbf{B}|^{-2} + \log \left| \mathbf{B}\mathbf{B}^T + \frac{1}{d}\mathbf{I} \right| \right) \\ &= \frac{1}{2K} \log \left| \mathbf{B}\mathbf{B}^T + \frac{1}{d}\mathbf{I} \right|. \end{aligned}$$

It can be seen that $R_{\text{bench}}^{\text{BT}}(d) \to -1/2 \log(d)$ as $d \to 0$.

For IF source coding, one can choose $\mathbf{A} = \mathbf{B}$. This choice gives

$$\begin{aligned} R_{\text{IF}}(\mathbf{B}, d) &= \frac{1}{2} \log \left(\max_{k=1,\ldots,K} \mathbf{b}_k^T \left(\mathbf{I} + \frac{1}{d}\mathbf{K}_{\mathbf{xx}} \right) \mathbf{b}_k \right) \\ &= \frac{1}{2} \log \left(\max_{k=1,\ldots,K} \|\mathbf{b}_k\|^2 + \frac{1}{d} \right) \end{aligned}$$

It is easy to see that $R_{\text{IF}}(\mathbf{B}, d) \to -1/2 \log(d)$ as $d \to 0$, just as the benchmark rate-distortion function, and therefore, according to Lemma 2, the choice $\mathbf{A} = \mathbf{B}$ is optimal at high resolution.

The naive approach that compresses each source without exploiting the existing correlations fails to achieve the benchmark rate-distortion function. In fact, it can only achieve

$$R_{\text{naive}}(d) = \frac{1}{2} \log \left(1 + \frac{\max_{k=1,\ldots,K} \|\tilde{\mathbf{b}}_k\|^2}{d} \right), \quad (23)$$

where $\tilde{\mathbf{b}}_k^T$ is the kth row of $\mathbf{B}^{-1}$. All entries of $\tilde{\mathbf{b}}_k$ are integer-valued since the matrix $\mathbf{B}$ is integer-valued with determinant 1. Therefore $\|\tilde{\mathbf{b}}_k\|^2 \geq 1$ for all $k = 1, \ldots, K$. The obtained compression rate approaches $\frac{1}{2} \log(\max \|\tilde{\mathbf{b}}_k\|^2) - 1/2 \log(d)$ as $d \to 0$. Thus, at high resolution, IF source coding requires $\frac{1}{2} \log(\max \|\tilde{\mathbf{b}}_k\|^2)$ bits less than the naive approach in order to achieve the same distortion. This improvement can be made unbounded by choosing $\mathbf{B}$ appropriately.

Example 2 (Compressing observations of correlated relays): Consider the problem of distributively compressing a K-dimensional Gaussian source $\mathbf{x}$ with zero mean and covariance matrix $\mathbf{K}_{\mathbf{xx}} = \text{SNR}\mathbf{H}\mathbf{H}^T + \mathbf{I}$ for some $\text{SNR} > 0$ and some matrix $\mathbf{H} \in \mathbb{R}^{K \times K}$. This choice of covariance matrix corresponds to the joint distribution of the signals observed by K relays in the Gaussian network depicted in Figure 3, where it is assumed that each of the K transmitters uses a random i.i.d. Gaussian codebook such that each of the signals $\mathbf{s}_1, \ldots, \mathbf{s}_K$ behaves statistically as white Gaussian noise. This network will be studied in more detail in the next subsection.

We plot the averages of the minimal required compression rates for the two schemes, i.e. the ergodic rate-distortion functions of the two schemes, along with the ergodic benchmark rate-distortion function, under the assumption that the entries of $\mathbf{H}$ are i.i.d. standard normal random variables. Figure 4a depicts these rates for $K = 4$ and $\text{SNR} = 20\text{dB}$ as a function of d. It is seen that at moderate to high resolution (small to moderate values of d) IF source coding closes about half of the gap between the naive compression scheme and the benchmark which corresponds to the Berger-Tung compression scheme.

One can argue that in the considered scenario the gap between the performance of the naive scheme and the benchmark
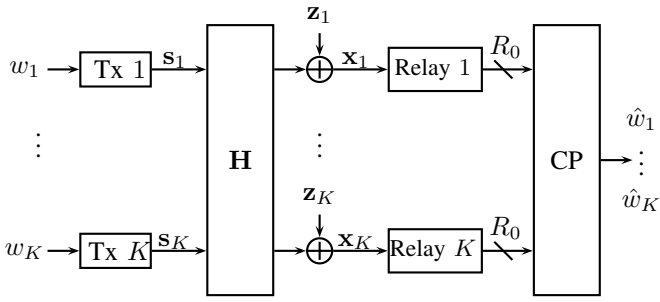
Fig. 3. A Gaussian network with K users and K relays. Each relay sees one output of the channel $\mathbf{x} = \mathbf{Hs} + \mathbf{z}$ and has a clean bit-pipe of R_0 bits/channel use to the central processor (CP). The CP tries to estimate the messages transmitted by the K users.

is quite small, and therefore it is not clear if IF source coding only slightly improves over the naive scheme, or closely follows the performance of the Berger-Tung benchmark. To illustrate that the latter is true, in Figure 4b we consider a similar scenario where now $\mathbf{H} \in \mathbb{R}^{8 \times 2}$ with i.i.d. $\mathcal{N}(0,1)$ entries. This models a network with 2 transmitters and 8 relays. This choice of distribution tends to induce more correlation between the entries of $\mathbf{x}$, which enlarges the performance gap between Berger-Tung's compression and the naive compression approach. Nevertheless, as seen from Figure 4b, the gap between the performance of the Berger-Tung benchmark and IF source coding remains approximately the same.

B. Layered Gaussian relay network

In this subsection we consider the Gaussian network from Figure 3, and show that for a wide regime of parameters using IF source coding as a building block improves upon other competing low-complexity coding schemes.

The Gaussian network we consider consists of K non-cooperating transmitters, each with message w_k and rate R_k. A central processor (CP) is interested in decoding all K messages. However, it does not have a direct access to the signals transmitted by the K transmitters. Instead, there are K relays, each of which observes a noisy linear combination of the transmitted signals. Each relay has a clean bit-pipe of rate R_0 bits/channel use connecting it to the CP which it uses for helping the CP decode all messages.

Let $\mathbf{s}_k \in \mathbb{R}^{1 \times n}$ be the signal transmitted by the kth transmitter during n consecutive channel uses. We assume all transmitters are subject to the same power-constraint such that $\mathbb{E}\|\mathbf{s}_k\|^2 \leq n\mathsf{SNR}$ for all $k = 1, \ldots, K$. Let $\mathbf{x}_k \in \mathbb{R}^{1 \times n}$ be the signal received by the kth relay during n consecutive channel uses, and let $\mathbf{S} = [\mathbf{s}_1^T \cdots \mathbf{s}_K^T]^T$ and $\mathbf{X} = [\mathbf{x}_1^T \cdots \mathbf{x}_K^T]^T$. The signals are related by

$$\mathbf{X} = \mathbf{HS} + \mathbf{Z}, \tag{24}$$

where $\mathbf{H} \in \mathbb{R}^{K \times K}$ is the channel matrix between the K transmitters and the K relays and the entries of $\mathbf{Z} \in \mathbb{R}^{K \times n}$ are i.i.d. $\mathcal{N}(0,1)$. We are interested in the maximal achievable sum-rate $R_{\text{sum}} = \sum_{k=1}^{K} R_k$.

Clearly, R_{sum} cannot exceed the MIMO capacity[5] corresponding to the channel (24) between the transmitters and relays, and it also cannot exceed KR_0 because even if each relay could decode all messages, the K relays cannot convey more than KR_0 bits/channels use to the CP through the bit-pipes. Thus, we have

$$R_{\text{sum}} \leq R_{\text{MIMO}} \triangleq \min\left(\frac{1}{2}\log|\mathbf{I} + \mathsf{SNR}\mathbf{HH}^T|, KR_0\right). \tag{25}$$

An inner bound for R_{sum} can be attained by the following scheme. Each relay can compress its observation $\mathbf{x}_k$ with rate R_0 and send the compression index to the CP. The CP obtains K estimates $\hat{\mathbf{x}}_k = \mathbf{x}_k + \mathbf{d}_k$ of the relays' observations, where $\mathbf{d}_k \in \mathbb{R}^{1 \times n}$ is the quantization error, and can use these estimates in order to decode the desired messages. Specifically, using this approach the CP decodes the messages from

$$\hat{\mathbf{X}} = \mathbf{HS} + \mathbf{Z} + \mathbf{D}, \tag{26}$$

where $\mathbf{D} = [\mathbf{d}_1^T \cdots \mathbf{d}_K^T]^T$. If the quantization errors are statistically independent of everything else, as in IF source coding, $\mathbf{D}$ can be treated as another additive noise. Let

$$d(R_0) = \max_{k=1,\ldots,K} \frac{1}{n}\mathbb{E}(\|\mathbf{d}_k\|^2).$$

Assuming that all transmitters use i.i.d. Gaussian codebooks, it follows from the entropy power inequality [31, Problem 9.21] that the CP can decode all messages $w_1, \ldots, w_K$ from the channel (26) if

$$R_{\text{sum}} \leq \frac{1}{2}\log\left|\mathbf{I} + \frac{\mathsf{SNR}}{1 + d(R_0)}\mathbf{HH}^T\right| \tag{27}$$

Clearly, the degradation of this scheme w.r.t. the MIMO capacity depends on the value of $d(R_0)$. Improving the compression scheme decreases $d(R_0)$ which in turn increases R_{sum}. One can use the conditionally unbiased version of Berger-Tung in order to obtain a small $d(R_0)$. However, this solution requires joint typicality decoding at the CP which is difficult to implement. Alternatively, IF source coding can be employed, which considerably reduces the implementation complexity at the price of slightly increasing $d(R_0)$. The relays can also employ naive conditionally unbiased compression, which is also a low-complexity scheme. This reduces to performing IF source coding with the choice $\mathbf{A} = \mathbf{I}$ which is often suboptimal. The latter approach is often termed compress-and-forward in the literature [19].

Alternatively, instead of compressing their noisy observations, the relays can attempt to decode the transmitted messages, or a function of the transmitted messages. In the decode-and-forward scheme [32] each relay decodes one of the messages and forwards this message to the CP. The compute-and-forward scheme [19] generalizes decode-and-forward and allows each relay to decode a linear combination of the messages, which is forwarded to the CP. Since decode-and-forward is a special case of compute-and-forward, its performance is never better.

[5]Here, by capacity we mean the mutual information corresponding to a white input, as the transmitters are non-cooperating.

(a) $\mathbf{H} \in \mathbb{R}^{4 \times 4}$ with i.i.d. $\mathcal{N}(0,1)$ entries

(b) $\mathbf{H} \in \mathbb{R}^{8 \times 2}$ with i.i.d. $\mathcal{N}(0,1)$ entries
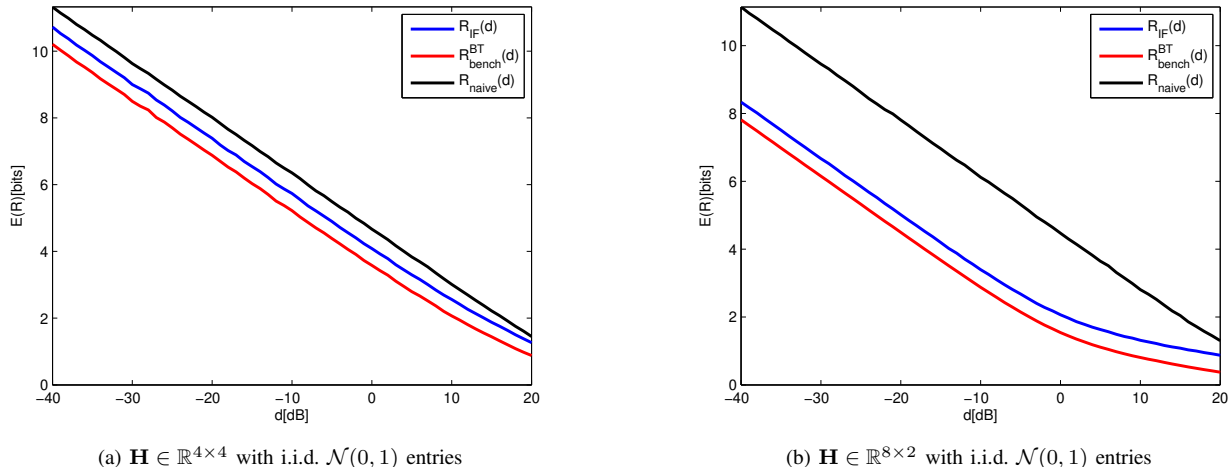
Fig. 4. Comparison between the ergodic rates for the various compression schemes from Example 2.

In Figure 5a we plot the ergodic rates achieved using IF source coding, compress-and-forward and compute-and-forward, over the Gaussian network from Figure 3 For $R_0 = 2$ and $K = 4$, where the entries of $\mathbf{H}$ are assumed i.i.d. $\mathcal{N}(0,1)$. Figure 5b depicts the same ergodic rates for $R_0 = 3$.

The figures demonstrate that while compute-and-forward outperforms both compression-based schemes when R_0 is the system's bottleneck, for relatively large R_0 (w.r.t. the $1/K$ times the MIMO capacity) compression is preferable over decoding. The gains of IF source coding over naive compression are evident.

One can further improve performance using a quantize-map-and-forward like scheme [33], [34] where each relay quantizes its observation, bins it, and sends the bin index to the CP. The difference between such schemes and the compression based schemes described above is that in quantize-map-and-forward the CP decodes the messages from the bin indices themselves without "decompressing" the relays' observations. Such an approach improves upon compression based schemes. However, to date it lacks a signal processing based architecture allowing to reduce the problem to multiple instances of a point-to-point problem, as is the case for IF source coding. We note however that progress in the direction of developing a low-complexity architecture for quantize-map-and-forward has been made in [35].

C. Distributed joint source-channel coding

In this subsection we consider the setup depicted in Figure 6. In this setup, there are K distributed encoders, each with access to the vector $\mathbf{x}_k$ that contains n i.i.d. samples of the random variable x_k. We assume that the random vector $\mathbf{x} = [x_1 \ \cdots \ x_K]^T$ is a Gaussian vector with zero mean and covariance matrix $\mathbf{K_{xx}}$. Each encoder is equipped with an encoding function $\mathcal{E}_k : \mathbb{R}^n \to \mathbb{R}^n$, such that the signal it transmits to the decoder is $\mathbf{s}_k = \mathcal{E}_k(\mathbf{x}_k)$. All encoders are subject to the same power constraint $\mathbb{E}(\|\mathbf{s}_k\|^2) = nP$. The decoder observes the transmitted signals through K parallel AWGN channels

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{z}_k, \quad k = 1, \ldots, K$$

where the entries of $\mathbf{z}_1, \ldots, \mathbf{z}_K$ are i.i.d. Gaussian random variables with zero mean and variance N. The decoder has K functions $\mathcal{D}_k : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^n$ that it uses in order to form estimates $\hat{\mathbf{x}}_k = \mathcal{D}_k(\mathbf{y}_1, \ldots, \mathbf{y}_K)$ for each source.

Let $\mathsf{SNR} \triangleq P/N$. An SNR-distortion vector $(\mathsf{SNR}, d_1, \ldots, d_K)$ is achievable if there exist encoding functions $\mathcal{E}_1, \ldots, \mathcal{E}_K$ and decoding functions $\mathcal{D}_1, \ldots, \mathcal{D}_K$ such that

$$\frac{1}{n}\mathbb{E}\left(\|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^2\right) \le d_k, \tag{28}$$

for all $k = 1, \ldots, K$. A conditionally *unbiased* SNR-distortion vector $(\mathsf{SNR}, d_1, \ldots, d_K)$ is achievable if in addition to (28), the condition

$$\mathbb{E}(\hat{\mathbf{x}}_k | \mathbf{X}) = \mathbf{x}_k, \quad k = 1, \ldots, K \tag{29}$$

is satisfied. As before, we restrict attention to conditionally unbiased estimates, and focus on the maximal distortion among the K vectors, i.e., $d = \max_{k=1,\ldots,K} d_k$.

An obvious approach for the considered problem is *separation* of source coding and channel coding. This corresponds to using AWGN capacity achieving codebooks for transforming the K AWGN channels into K bit-pipes each with capacity $C = 1/2 \log(1 + \mathsf{SNR})$ bits/channel use, and then using distributed source coding with rate C bits/sample at each encoder in order to describe the sources to the decoder. The main drawback of this approach is that it must be designed for specific values of SNR and required distortions $d_1, \ldots, d_K$. The predefined SNR acts as a threshold. If the actual SNR experienced by the communication system turns out to be higher than this threshold, the expected distortions would be $d_1, \ldots, d_k$, but would not improve when the actual SNR is improved.

Taking $K = 1$ in our setup reduces it to a point-to-point problem of Gaussian source transmission over an AWGN channel. It is well known [36] that analog transmission of the
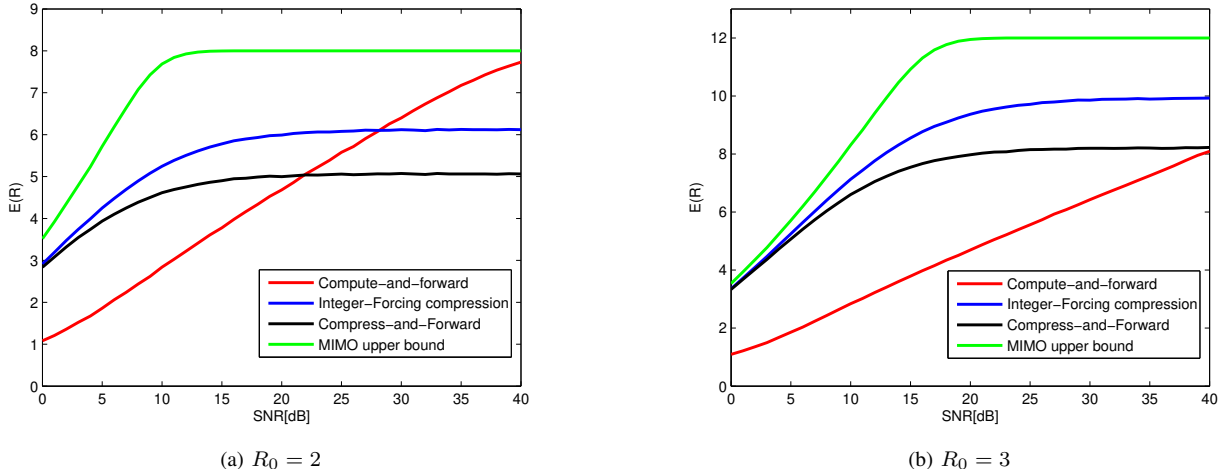
(a) $R_0 = 2$

(b) $R_0 = 3$

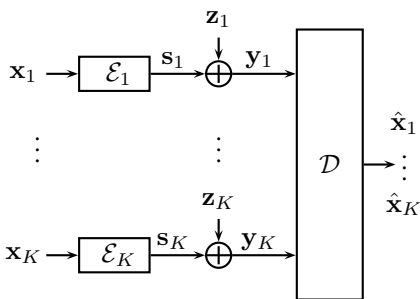Fig. 5.   Ergodic rates over the network from Figure 3 for $K = 4$

Fig. 6.   A distributed joint source-channel coding setting. Each encoder wishes to describe its observation $\mathbf{x}_k$ to the decoder through an AWGN channel, with minimal average MSE distortion. The sources are correlated and the encoders are distributed.

source with appropriate scaling at the encoder and decoder achieves the optimal performance. Moreover, the transmitter's operation does not depend on the noise's variance at the receiver. As a result, if the noise variance turns out to be smaller than expected, the decoder can improve the quality of its estimate for the source. This desirable phenomena was extended to the Wyner-Ziv/dirty-paper setting in [37]. Here, we use the idea of IF source coding for constructing a joint source-channel coding scheme for our setup with an arbitrary number of users. The encoders' operation in the proposed scheme is independent of the noise variance, and the obtained expected distortion at the decoder decreases with N, provided that N is below some predefined threshold.

The proposed coding approach utilizes a single lattice Λ with $\sigma^2(\Lambda) = P$, that is good for channel coding and for MSE quantization. In particular, its goodness for MSE quantization implies that $r_{\text{eff}}^2(\Lambda)/n \approx P$. The coding scheme is designed assuming that the AWGN variance is not greater than some nominal value N^{nom}. However, when $N < N^{\text{nom}}$, the obtained distortion decreases as N decreases.

Each encoder scales its observation by some $\beta > 0$ to be

defined shortly,[6] adds a dither $\mathbf{d}_k$ uniformly distributed over $\mathcal{V}$, and reduces the result $\bmod \, \Lambda$ such that the transmitted signals are

$$\mathbf{s}_k = [\beta \mathbf{x}_k + \mathbf{d}_k] \bmod \Lambda, \quad k = 1, \ldots, K.$$

Note that the power constraints are satisfied as $\mathbf{s}_k$ is uniformly distributed over $\mathcal{V}$ and therefore its second moment equals $\sigma^2(\Lambda)$. The decoder first performs MMSE estimation of each $\mathbf{s}_k$, by scaling each $\mathbf{y}_k$ by $\alpha = \sqrt{P/(P+N)}$, subtracting back the dither and reducing $\bmod \, \Lambda$. This gives

$$
\begin{aligned}
\tilde{\mathbf{y}}_k &= [\alpha \mathbf{y}_k - \mathbf{d}_k] \bmod \Lambda \\
&= [\mathbf{s}_k + (\alpha - 1)\mathbf{s}_k + \alpha \mathbf{z}_k - \mathbf{d}_k] \bmod \Lambda \\
&= [\beta \mathbf{x}_k + \mathbf{z}_{\text{eff},k}] \bmod \Lambda,
\end{aligned}
$$

where

$$\mathbf{z}_{\text{eff},k} \triangleq (\alpha - 1)\mathbf{s}_k + \alpha \mathbf{z}_k.$$

The noise $\mathbf{z}_{\text{eff},k}$ is statistically independent of $\mathbf{x}_k$, and has effective variance of

$$\frac{1}{n}\mathbb{E}(\|\mathbf{z}_{\text{eff},k}\|^2) = \frac{NP}{N+P}, \quad k = 1, \ldots, K.$$

Moreover, it is a linear combination of a dither uniformly distributed over the Voronoi region of a lattice that is good for MSE quantization and an AWGN, and therefore, by Lemma 1, it is semi-norm ergodic.

As before, let $\mathbf{X} = [\mathbf{x}_1^T \, \cdots \, \mathbf{x}_K^T]^T$, and define $\tilde{\mathbf{Y}}$ and $\mathbf{Z}_{\text{eff}}$ in a similar manner. The decoder chooses a full-rank matrix $\mathbf{A} \in \mathbb{Z}^{K \times K}$ and computes

$$
\begin{aligned}
\widehat{\beta \mathbf{A} \mathbf{X}} &\triangleq \left[\mathbf{A} \tilde{\mathbf{Y}} \right] \bmod \Lambda \\
&= [\mathbf{A}([\beta \mathbf{X} + \mathbf{Z}_{\text{eff}}] \bmod \Lambda)] \bmod \Lambda \\
&= [\mathbf{A}(\beta \mathbf{X} + \mathbf{Z}_{\text{eff}})] \bmod \Lambda.
\end{aligned}
$$

[6]In general, performance can be improved by letting each encoder use a different β_k. We disregard this possibility for simplicity of exposition.

Let $\mathbf{a}_k^T$ be the kth row of $\mathbf{A}$. The random vector $\mathbf{a}_k^T(\beta\mathbf{X}+\mathbf{Z}_{\text{eff}})$ is semi-norm ergodic with zero mean and effective variance

$$\sigma_k^2 \triangleq \frac{1}{n}\mathbb{E}(\|\mathbf{a}_k^T(\beta\mathbf{X}+\mathbf{Z}_{\text{eff}})\|^2)$$
$$= \mathbf{a}_k^T\left(\beta^2\mathbf{K}_{\mathbf{xx}} + \frac{NP}{N+P}\mathbf{I}\right)\mathbf{a}_k.$$

Since Λ is good for channel coding, if $\sigma_k^2 < P$ for all $k = 1,\ldots,K$, then

$$\widehat{\beta\mathbf{A}\mathbf{X}} \overset{(w.h.p.)}{=} \mathbf{A}(\beta\mathbf{X}+\mathbf{Z}_{\text{eff}}), \qquad (30)$$

and the decoder can further compute

$$\hat{\mathbf{X}} = \frac{1}{\beta}A^{-1}\widehat{\beta\mathbf{A}\mathbf{X}}$$
$$\overset{(w.h.p.)}{=} \mathbf{X} + \frac{1}{\beta}\mathbf{Z}_{\text{eff}},$$

which are unbiased estimates of each $\mathbf{x}_k$ with average MSE distortion of $d_{\text{IF}} = NP/\beta^2(N+P)$.

The remaining question is how to choose β such that (30) indeed holds. Recall that β is chosen by the encoders that only know that $N < N^{\text{nom}}$, rather than the exact value of N. Therefore, the encoders should choose β as

$$\beta_{\text{opt}}(P, N^{\text{nom}}, \mathbf{K}_{\mathbf{xx}}) \triangleq$$

$$\max_{\beta>0} s.t. \min_{\substack{\mathbf{A}\in\mathbb{Z}^{K\times K} \\ \det(\mathbf{A})\neq 0}} \max_{k=1,\ldots,K} \mathbf{a}_k^T(\beta^2\mathbf{K}_{\mathbf{xx}} + \frac{N^{\text{nom}}P}{N^{\text{nom}}+P}\mathbf{I})\mathbf{a}_k = P$$

and the symmetric distortion obtained by the proposed scheme is

$$d_{\text{IF}} = \frac{N}{\beta_{\text{opt}}^2(P, N^{\text{nom}}, \mathbf{K}_{\mathbf{xx}})}$$

which decreases as N decreases, as desired.

A naive joint source-channel coding schemes that ignores the correlations between the entries of $\mathbf{x}$ would be transmitting each x_k in an analog Goblick-like scheme. The distortion achieved by such a scheme would be[7]

$$d_{\text{naive}} = \frac{N}{P}\max_{k=1,\ldots,K}\mathbf{K}_{\mathbf{xx}}(k,k).$$

It can be easily verified that the same distortion is achieved if one constrains $\mathbf{A} = \mathbf{I}$ in the scheme proposed here. Therefore, the proposed IF based joint source-channel coding scheme strictly improves upon the naive one.

It is also worth mentioning that the proposed scheme easily generalizes to a dirty paper scenario, where the output of each AWGN channel is further corrupted by an arbitrary interference $\mathbf{v}_k$ known to encoder k but not to the decoder, i.e., $\mathbf{y}_k = \mathbf{s}_k + \mathbf{v}_k + \mathbf{z}_k$. In the proposed scheme, the encoders can transmit $\mathbf{s}_k = [\beta\mathbf{x}_k - \mathbf{v}_k + \mathbf{d}_k] \mod \Lambda$ and the decoder remains the same.

[7]Taking into account the constraint that the estimate for each x_k must be conditionally unbiased.

VI. ONE-SHOT INTEGER-FORCING SOURCE CODING

One of the advantages of IF source coding is that its complexity and performance can be traded-off, by choosing nested lattice codes that can be easily implemented, but are less effective as channel codes and MSE quantizers.

In the previous sections we have considered the extreme case of high-dimensional pairs of nested lattices where the fine lattice is good for MSE quantization and the coarse lattice is good for channel coding. In this section we consider the other extreme, where both lattices are scaled versions of the integer lattice $\mathbb{Z}$. With this choice of nested lattice pair, IF source coding becomes extremely easy to implement. Moreover, this one-shot version of IF source coding does not induce any latency and does not assume the existence of an unlimited number of i.i.d. samples to be compressed.

Let $\Lambda_f = \sqrt{12d}\mathbb{Z}$ and $\Lambda = 2^R\sqrt{12d}\mathbb{Z}$. If 2^R is a positive integer then $\Lambda \subseteq \Lambda_f$, and the codebook $\mathcal{C} = \Lambda_f \cap \mathcal{V}$ with rate R is a valid codebook for IF source coding. Let d_k be a random dither uniformly distributed over $\mathcal{V}_f$, known to both the kth encoder and the decoder. The kth encoder conveys the index corresponding to the point

$$[Q_{\Lambda_f}(x_k + d_k)] \mod \Lambda$$

to the decoder. Note that for a 1D lattice, the quantization operation reduces to a simple slicer. Thus all operations are easy to implement.

The decoder first subtracts back the dither and reduces mod Λ to obtain

$$\tilde{x}_k \overset{(i.d.)}{=} [x_k + d_k] \mod \Lambda,$$

and then chooses some full-rank matrix $\mathbf{A} \in \mathbb{Z}^{K\times K}$ and computes

$$\widehat{\mathbf{Ax}} \triangleq [\mathbf{A}\tilde{\mathbf{x}}] \mod \Lambda = [\mathbf{A}(\mathbf{x}+\mathbf{d})] \mod \Lambda, \qquad (31)$$

where $\mathbf{d} = [d_1 \cdots d_K]^T$. In contrast to the case of a high-dimensional nested lattice codebook, where the probability that $\widehat{\mathbf{Ax}} \neq \mathbf{Ax}$ could be made as low as desired if $r_{\text{eff}}^2(\Lambda)$ is large enough, here this probability is finite for any finite value of $2^R\sqrt{12d}$. In particular, let $\mathbf{a}_k^T$ be the kth row of $\mathbf{A}$ and define the random variable

$$w_k \triangleq \mathbf{a}_k^T(\mathbf{x}+\mathbf{d})$$

with zero mean and variance

$$\sigma_{w,k}^2 = \mathbf{a}_k^T(\mathbf{K}_{\mathbf{xx}} + d\mathbf{I})\mathbf{a}_k.$$

We have

$$\Pr\left(\widehat{\mathbf{Ax}} \neq \mathbf{Ax}\right) = \Pr\left(\bigcup_{k=1}^{K}[w_k]\mod\Lambda \neq w_k\right)$$
$$= \Pr\left(\bigcup_{k=1}^{K}Q_\Lambda(w_k) \neq 0\right)$$
$$= \Pr\left(\bigcup_{k=1}^{K}|w_k| \geq \frac{1}{2}2^R\sqrt{12d}\right)$$
$$\leq \sum_{k=1}^{K}\Pr\left(|w_k| \geq 2^R\sqrt{3d}\right), \qquad (32)$$

where the last inequality follows from the union bound. Next, we apply the following Lemma from [38], [39]

Lemma 3: [39, Lemma 3] Consider the random variable

$$z_{\text{eff}} = \sum_{\ell=1}^{L} \alpha_\ell z_\ell + \sum_{k=1}^{K} \beta_k d_k$$

where $\{z_\ell\}_{\ell=1}^{L}$ are i.i.d. Gaussian random variables with zero mean and some variance σ_z^2 and $\{d_k\}_{k=1}^{K}$ are i.i.d. random variables, statistically independent of $\{z_\ell\}_{\ell=1}^{L}$, uniformly distributed over the interval $[-\rho/2, \rho/2]$ for some $\rho > 0$. Let $\sigma_{\text{eff}}^2 \triangleq \mathbb{E}(z_{\text{eff}}^2)$. Then

$$\Pr(z_{\text{eff}} > \tau) = \Pr(z_{\text{eff}} < -\tau) \le \exp\left\{-\frac{\tau^2}{2\sigma_{\text{eff}}^2}\right\}.$$

One can easily verify that w_k satisfies the conditions of Lemma 3 as $\mathbf{a}_k^T \mathbf{x}$ is a Gaussian random variable statistically independent of the dither vector $\mathbf{d}$. Therefore, we can further bound (32) as

$$\Pr\left(\widehat{\mathbf{Ax}} \ne \mathbf{Ax}\right) \le \sum_{k=1}^{K} 2\exp\left\{-\frac{2^{2R}3d}{2\mathbf{a}_k^T\left(\mathbf{K_{xx}} + d\mathbf{I}\right)\mathbf{a}_k}\right\}$$

$$\le 2K\exp\left\{-\frac{3}{2}2^{2\left(R - \frac{1}{2}\log(\max_{k=1,\dots,K} \mathbf{a}_k^T\left(\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}}\right)\mathbf{a}_k)\right)}\right\}$$

$$= 2K\exp\left\{-\frac{3}{2}2^{2(R - R_{\text{IF}}(\mathbf{A},d))}\right\}, \tag{33}$$

where $R_{\text{IF}}(\mathbf{A}, d)$ is the minimum required rate for a IF source coding when a good nested lattice pair is used, as defined in Theorem 1. The decoder proceeds by computing

$$\hat{\mathbf{x}} = \mathbf{A}^{-1}\widehat{\mathbf{Ax}} = \mathbf{x} + \mathbf{d} + \mathbf{A}^{-1}\left(\widehat{\mathbf{Ax}} - \mathbf{Ax}\right). \tag{34}$$

Since d_k is statistically independent of $\mathbf{x}$ and $\mathbb{E}(d_k^2) = d$ for all $k = 1, \dots, K$, we see that provided that $\widehat{\mathbf{Ax}} = \mathbf{Ax}$ the one-shot version of IF source coding produces conditionally unbiased estimates of x_k with distortion d. The probability that $\widehat{\mathbf{Ax}} = \mathbf{Ax}$ can be controlled by increasing $R - R_{\text{IF}}(\mathbf{A}, d)$ which is the coding overhead w.r.t. to IF source coding with an optimal nested lattice pair. For instance, if $K = 4$, taking $R = R_{\text{IF}}(\mathbf{A}, d) + 2$ results in $\Pr\left(\widehat{\mathbf{Ax}} \ne \mathbf{Ax}\right) \le 3 \cdot 10^{-10}$. The next theorem summarizes the discussion above.

Theorem 2 (One-shot IF source coding): Let $R_{\text{IF}}(d)$ be as defined in Theorem 1 and set $R = R_{\text{IF}}(d) + \Delta$ for some $\Delta > 0$. If 2^R is a positive integer, the one-shot version of IF source coding with lattices $\Lambda_f = \sqrt{12d}\mathbb{Z}$ and $\Lambda = 2^R\sqrt{12d}\mathbb{Z}$ produces conditionally unbiased estimates with average MSE distortion d for each x_k, $k = 1, \dots, K$ with probability greater than $1 - 2K\exp\{-\frac{3}{2}2^{2\Delta}\}$.

A. Modulo Analog-to-Digital Converters

Theorem 2 shows that a simple implementation of IF source coding with $1D$ lattices only requires a small rate overhead w.r.t. to the asymptotic performance of IF source coding. The simplicity of the one-shot IF source coding scheme suggests that this framework may be useful for designing Analog-to-Digital converters (ADCs) that can exploit correlations in a distributed manner. To illustrate the problem, consider the Gaussian MIMO channel $\mathbf{x} = \mathbf{Hs} + \mathbf{z}$, where $\mathbf{H} \in \mathbb{R}^{K \times M}$ is the channel matrix, $\mathbf{z} \in \mathbb{R}^{K \times 1}$ is a vector of AWGN and $\mathbf{s}$ are the M inputs to channel, which are assumed to be i.i.d. normally distributed. The front-end of the MIMO receiver consists of K ADCs, one for the output of each receive antenna. Today, each of these ADCs is designed w.r.t. the *marginal* distribution of each output, ignoring the fact that the K ADCs sample correlated signals. Often, the variance of each output is quite large although the *conditional* variance when all other samples are given is small. Thus, exploiting the spatial correlation may significantly reduce the distortion created by the ADCs. However, the ADCs are expected to work at very high rates, which precludes cooperation between their operations. We show that a variant of the one-shot IF source coding scheme allows the ADCs to exploit the spatial correlations with no cooperation and with roughly the same encoding complexity as a standard ADC, and only a small increase in the decoding complexity.

The one-shot version of IF source coding described above requires each encoder to first quantize its observation using a scaled integer lattice, and then reduce the result modulo the coarse lattice, which is also a scaled version of $\mathbb{Z}$. This can be implemented by applying an ADC as the quantizer followed by a digital modulo reduction. However, the power consumption and the complexity of an ADC are dictated by the number of bits it produces. Therefore, if the modulo operation can be implemented efficiently in the analog domain, performance can be improved by first applying the modulo reduction, and only then incorporating the ADC. Since the modulo reduced signal is of a smaller support, less bits are required for describing it with the same average distortion level. The next lemma shows that if $\Lambda_f = \sqrt{12d}\mathbb{Z}$ and $\Lambda = 2^R\sqrt{12d}\mathbb{Z}$ the operations Q_{Λ_f} and $\mod \Lambda$ commute, i.e., one can first reduce the signal $\mod\Lambda$ and then quantize to Λ_f, rather than first quantizing and then reducing $\mod\Lambda$.

Lemma 4: Let 2^R be a positive odd integer and define the nested lattices $\Lambda = \sqrt{12d}\mathbb{Z}$ and $\Lambda_f = 2^R\sqrt{12d}\mathbb{Z}$ for some $d > 0$. for any $x \in \mathbb{R}$ we have

$$\left[Q_{\Lambda_f}(x)\right] \mod \Lambda = Q_{\Lambda_f}\left([x] \mod \Lambda\right).$$

Proof: See Appendix A ∎

Lemma 4 implies that the $1D$ version of IF source coding can indeed be implemented by first reducing the source x modulo Λ and only then quantizing it to Λ_f. The advantage in switching the order of the operations is that if the $1D$ modulo reduction, which is equivalent to the "saw-tooth" function, can be efficiently implemented in the analog domain, then the quantizer that follows it can be implemented using an ADC with only R bits/sample. The relation between R, the obtained distortion, and the error probability is characterized in Theorem 2 and depends on $R_{\text{IF}}(d)$. Figure 7 depicts the architecture of the proposed modulo ADC, that can replace the encoders in the one-shot IF source coding scheme.
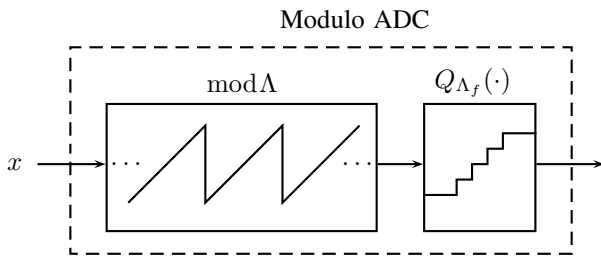
Modulo ADC

Fig. 7. A schematic illustration of the modulo ADC for $2^R = 5$. This component can act as an encoder in the one-shot version of IF source coding.

VII. SUMMARY AND CONCLUSIONS

We have presented and analyzed a new low-complexity framework for distributed lossy compression, which is based on the integer-forcing architecture. This framework allows the system designer to trade performance and complexity by appropriately choosing the nested lattice codebooks that are used. A remarkable feature of the proposed scheme is that it admits a very simple one-shot version, whose performance is not very far from that obtained using IF source coding with asymptotically good nested lattice codes. We have also shown that if one can implement the $1D$ modulo operation with an analog circuit, which corresponds to implementing the "saw-tooth" function, then the IF source coding approach can translate to a novel ADC design, suitable for sampling spatially correlated sources. Such ADCs can potentially be very useful for the front-end of a MIMO receiver, where standard ADC designs are already challenged by the growing transmission rates.

We remark that the IF equalization framework for Gaussian MIMO channels [18] has been extended to an equalization framework for Gaussian intersymbol-interference channels [40]. In a similar manner, the IF source coding framework proposed here, which is suitable for distributed lossy compression of spatially correlated signals, can be extended to an IF compression framework for stationary temporally correlated signals. Nevertheless, such a solution is less attractive as one can always use a sequential Wyner-Ziv like compression scheme for a stationary source. In such a scheme the first samples of the source are compressed without binning/modulo reduction, and the next samples are first binned/modulo reduced and then compressed. The decoder uses the samples that are not binned for recovering the next samples in a sequential manner. This Wyner-Ziv scheme suffers from the intrinsic overhead of having to describe the first samples to the decoder without binning. This overhead can be made negligible by increasing the length of the compression block. For spatially correlated sources a similar Wyner-Ziv like compression scheme will result in asymmetric compression rates, which is a consequence of the lack of "spatial stationarity".

APPENDIX A
PROOF OF LEMMA 4

We begin with two general Lemmas from which Lemma 4 is immediately deduced.

Lemma 5: For any pair of n-dimensional nested lattices $\Lambda \subseteq \Lambda_f$ and any $\mathbf{x} \in \mathbb{R}^n$

$$
\begin{aligned}
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda = & Q_{\Lambda_f}\left([\mathbf{x}] \bmod \Lambda\right) \\
& + Q_\Lambda\left(\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x})\right).
\end{aligned}
$$

Proof:

$$
\begin{aligned}
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda &= Q_{\Lambda_f}(\mathbf{x}) - Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right) \\
&= Q_{\Lambda_f}\left(\mathbf{x} - Q_\Lambda(\mathbf{x}) + Q_\Lambda(\mathbf{x})\right) - Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right) \\
&= Q_{\Lambda_f}\left(\mathbf{x} - Q_\Lambda(\mathbf{x})\right) + Q_{\Lambda_f}\left(Q_\Lambda(\mathbf{x})\right) - Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right) \\
&= Q_{\Lambda_f}\left([\mathbf{x}] \bmod \Lambda\right) + Q_\Lambda(\mathbf{x}) - Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right), \quad (35)
\end{aligned}
$$

where in the last equality we have used the fact that $Q_{\Lambda_f}\left(Q_\Lambda(\mathbf{x})\right) = Q_\Lambda(\mathbf{x})$ since $\Lambda \subseteq \Lambda_f$. We have,

$$
\begin{aligned}
Q_\Lambda(\mathbf{x}) &= Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x}) + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x})\right) \\
&= Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x}) - Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right) + Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right) + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x})\right) \\
&= Q_\Lambda\left(\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x})\right) + Q_\Lambda\left(Q_{\Lambda_f}(\mathbf{x})\right).
\end{aligned}
$$
(36)

Substituting (36) in (35) gives the desired result. ∎

Lemma 6: If the pair of nested lattices $\Lambda \subseteq \Lambda_f$ satisfies the tiling condition $\mathcal{V} = (\Lambda_f \cap \mathcal{V}) + \mathcal{V}_f$ then

$$
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda = Q_{\Lambda_f}\left([\mathbf{x}] \bmod \Lambda\right).
$$

for any $\mathbf{x} \in \mathbb{R}^n$.

Proof: For any $\mathbf{x} \in \mathbb{R}^n$ we have

$$
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda \in (\Lambda_f \cap \mathcal{V}), \text{ and } \mathbf{x} - Q_{\Lambda_f}(\mathbf{x}) \in \mathcal{V}_f.
$$

Therefore

$$
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x}) \in (\Lambda_f \cap \mathcal{V}) + \mathcal{V}_f,
$$

The tiling condition $\mathcal{V} = (\Lambda_f \cap \mathcal{V}) + \mathcal{V}_f$ implies that

$$
\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x}) \in \mathcal{V},
$$

which implies that

$$
Q_\Lambda\left(\left[Q_{\Lambda_f}(\mathbf{x})\right] \bmod \Lambda + \mathbf{x} - Q_{\Lambda_f}(\mathbf{x})\right) = 0.
$$

The result now follows immediately from Lemma 5. ∎

It is easy to verify that if 2^R is a positive odd integer the nested lattices $\Lambda = \sqrt{12d}\mathbb{Z}$ and $\Lambda_f = 2^R\sqrt{12d}\mathbb{Z}$ satisfy the tiling condition $\mathcal{V} = (\Lambda_f \cap \mathcal{V}) + \mathcal{V}_f$, and Lemma 4 immediately follows from Lemma 6

REFERENCES

[1] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, 1978.
[2] T. Berger, "Multiterminal source coding," in *Lectures presented CISM Summer School on the Information Theory Approach to Communications*, July 1977.
[3] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.
[4] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, October 2007.
[5] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5268–5651, December 2009.

[6] A. B. Wagner, "On distributed compression of linear functions," *IEEE Transactions on Information Theory*, vol. 57, no. 1, pp. 79–94, January 2011.

[7] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[8] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 1938–1961, 2008.

[9] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, March 2003.

[10] S. Pradhan and K. Ramchandran, "Distributed source coding: symmetric rates and applications to sensor networks," in *Data Compression Conference, 2000. Proceedings. DCC 2000*, 2000, pp. 363–372.

[11] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *Signal Processing Magazine, IEEE*, vol. 21, no. 5, pp. 80–94, 2004.

[12] H. Karl and A. Willig, *Protocols and architectures for wireless sensor networks*. Wiley. com, 2007.

[13] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[14] R. Zamir and S. Shamai, "Nested linear/lattice codes for Wyner-Ziv encoding," in *Information Theory Workshop, 1998*, 1998, pp. 92–93.

[15] J. van Valburg and R. van de Plassche, "An 8-b 650-MHz folding ADC," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 12, pp. 1662–1666, 1992.

[16] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[17] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, 1999.

[18] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Transactions on Information Theory*, Submitted January 2012, available online: http://arxiv.org/abs/1003.5966.

[19] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[20] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian many-help-one distributed source coding problem," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 564–571, January 2010.

[21] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[22] J. Ziv, "On universal quantization," *IEEE Transactions on Information Theory*, vol. 31, no. 3, pp. 344–347, 1985.

[23] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 428–436, 1992.

[24] R. Zamir, *Lattice coding for signals and networks*, a book in preparation.

[25] O. Ordentlich and U. Erez, "A simple proof for the existence of "good" pairs of nested lattices," in *Proceedings of the 27th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Nov. 2012, pp. 1–12.

[26] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3401–3416, October 2005.

[27] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, no. 4, pp. 515–534, 1982.

[28] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," 2013, see http://arxiv.org/abs/1303.0696.

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

[30] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.

[31] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ: Wiley-Interscience, 2006.

[32] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.

[33] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3457–3478, August 2009.

[34] S. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[35] V. Nagpal, I.-H. Wang, M. Jorgovanovic, D. Tse, and B. Nikolic, "Coding and system design for quantize-map-and-forward relaying," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 8, pp. 1423–1435, 2013.

[36] J. Goblick, T., "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558–567, 1965.

[37] Y. Kochman and R. Zamir, "Joint Wyner-Ziv/dirty-paper coding by modulo-lattice modulation," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 4878–4889, 2009.

[38] C. Feng, D. Silva, and F. Kschischang, "An algebraic approach to physical-layer network coding," *IEEE Transactions on Information Theory*, Submitted July 2011, available online http://arxiv.org/abs/1108.1695.

[39] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Trans. on Information Theory*, Submitted April 2013, see http://www.eng.tau.ac.il/~ordent/publications/IFConstantGap.pdf.

[40] ——, "Cyclic-coded integer-forcing equalization," *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5804–5815, 2012.