Integer-Forcing Source Coding

Or Ordentlich and Uri Erez
Dept. EE-Systems, Tel Aviv University, Israel

***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 develops the source coding dual of 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 lower 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, 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. A special case that received considerable attention is that of distributed lossy compression of jointly Gaussian random 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 [3]. 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 encoding. For the quadratic Gaussian case with $K = 2$, Wagner *et al*. [4] 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 [5], [6] and remains an active area of research. 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 [7] and successive Wyner-Ziv coding [8], or specifically tailored to predefined correlation characteristics of the sources [5]. 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 [7]. Nevertheless, schemes using time-sharing suffer from several 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 essential 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.

An important feature of the proposed approach is that it affords the system designer a flexible trade-off between 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 becomes less significant.

Due to space limitation, technical details are often omitted throughout the paper and may be found in [9].

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

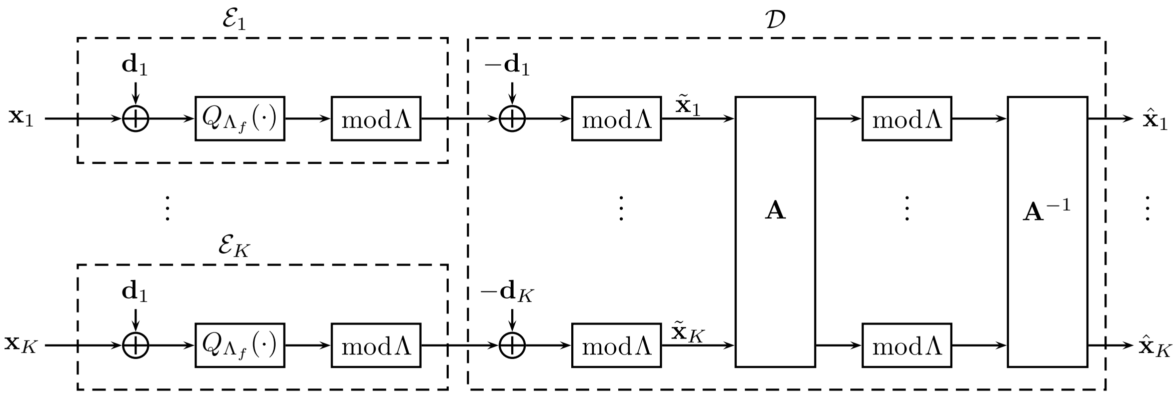

Fig. 1. A schematic overview of the integer-forcing source coding framework with the nested lattice pair $\Lambda \subset \Lambda_f$.

all logarithms are to the base 2.

II. Problem statement and preliminaries

We consider a distributed source coding setting with K encoding terminals and one decoder. Each of the K encoders has access to a vector $\mathbf{x}_k \in \mathbb{R}^{1\times n}$ of n i.i.d. realizations of the random variable x_k, $k = 1,\ldots,K$. The random vector $\mathbf{x} = [x_1 \ \cdots \ x_K]^T$ 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 $\mathcal{E}_k : \mathbb{R}^{1\times n} \to \{1,\ldots,2^{nR_k}\}$ and sends the index to the decoder. The decoder is equipped with K decoding functions $\mathcal{D}_k : \{1,\ldots,2^{nR_1}\} \times \cdots \times \{1,\ldots,2^{nR_K}\} \to \mathbb{R}^{1\times n}$ 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

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

for all $k = 1,\ldots,K$. Let $\mathbf{X} \triangleq [\mathbf{x}_1^T \ \cdots \ \mathbf{x}_K^T]^T \in \mathbb{R}^{K\times n}$. 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 $\mathbf{X}$. 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 $d_1 = \cdots = d_k = d$. 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.* [4] showed that the Berger-Tung approach is optimal. For $K > 2$ it is now known that the Berger-Tung approach does not attain the full rate-distortion region (see e.g. [3]). 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. Under the unbiased requirement (2), this benchmark is given by

$$R^{\text{BT}}_{\text{bench}}(d) \triangleq \frac{1}{2K}\log\left|\mathbf{I} + \frac{1}{d}\mathbf{K_{xx}}\right|. \tag{3}$$

See [9] for further details.

A. Lattice Preliminaries

Several lattice properties that will be useful in the sequel are briefly recalled. 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 $Q_\Lambda(\mathbf{y}) = \mathrm{argmin}_{\mathbf{t}\in\Lambda}\|\mathbf{y}-\mathbf{t}\|$. 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. 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 $[a[\mathbf{y}] \bmod \Lambda] \bmod \Lambda = [a\mathbf{y}] \bmod \Lambda$ for any $a \in \mathbb{Z}$ and $\mathbf{y} \in \mathbb{R}^n$. The second moment of Λ is defined as

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

where $\mathrm{Vol}(\mathcal{V})$ 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 $\mathrm{Vol}(\mathcal{V})$.

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

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

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 $\sigma^2_{\mathbf{Z}}\triangleq\mathbb{E}\|\mathbf{z}\|^2/n$, is *semi norm-ergodic* if for any $\epsilon>0$, $\delta>0$ and n large enough

$$\Pr\left(\|\mathbf{z}\|>\sqrt{(1+\delta)n\sigma^2_{\mathbf{z}}}\right)\leq\epsilon. \tag{4}$$

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

The next Lemma restates Corollary 2 from [10] 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 $\mathbb{E}\|\mathbf{z}\|^2/n<(1-\delta)r^2_{\text{eff}}(\Lambda)/n$

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

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 [10], and the existence of lattice pairs with slightly more demanding "goodness" requirements was shown in [3], [11]. A nested lattice code $\mathcal{C}=\Lambda_f\cap\mathcal{V}$ with rate

$$R=\frac{1}{n}\log\left(\frac{\text{Vol}(\Lambda)}{\text{Vol}(\Lambda_f)}\right)=\frac{1}{2}\log\left(\frac{r^2_{\text{eff}}(\Lambda)}{r^2_{\text{eff}}(\Lambda_f)}\right) \tag{5}$$

is associated with the nested lattice pair.

III. 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 $\Lambda\subset\Lambda_f$. 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

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

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

$$\begin{aligned}\tilde{\mathbf{x}}_k&=\left[\left[Q_{\Lambda_f}(\mathbf{x}_k+\mathbf{d}_k)\right]\bmod\Lambda-\mathbf{d}_k\right]\bmod\Lambda\\&=\left[\mathbf{x}_k+\left[Q_{\Lambda_f}(\mathbf{x}_k+\mathbf{d}_k)\right]\bmod\Lambda-(\mathbf{x}_k+\mathbf{d}_k)\right]\bmod\Lambda\\&\overset{(i.d.)}{=}\left[\mathbf{x}_k+\mathbf{d}_k\right]\bmod\Lambda,\end{aligned} \tag{6}$$

where $\overset{(i.d.)}{=}$ denotes equality in distribution, and (6) is justified by the Crypto Lemma [12, Lemma 1]. If the coarse lattice Λ is chosen such that its effective radius is large enough, the modulo operation in (6) 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 $r^2_{\text{eff}}(\Lambda)$, and we would therefore prefer to choose it as small as possible. In IF source coding $r^2_{\text{eff}}(\Lambda)$ may be chosen such that $\mathbf{x}_k+\mathbf{d}_k$ cannot be recovered from $\tilde{\mathbf{x}}_k$ alone, but can be recovered from $\{\tilde{\mathbf{x}}_k\}_{k=1}^K$.

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, $r^2_{\text{eff}}(\Lambda)$ 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

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

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

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

then for n large enough

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

Moreover, if this holds for all $k=1,\ldots,K$, then for n large enough

$$\widehat{\mathbf{AX}} \overset{(w.h.p.)}{=} \mathbf{A}(\mathbf{X}+\mathbf{D}). \tag{7}$$

For (7) to hold, it suffices to set

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

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

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

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

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

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

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

It can be shown [9] that the problem of finding the optimal matrix $\mathbf{A}$ is equivalent to finding the K shortest linearly independent vectors of a lattice induced by the matrix $\mathbf{I}+\frac{1}{d}\mathbf{K}_{\mathbf{xx}}$. Although this problem is NP-hard in general, its solution can be efficiently approximated using the LLL algorithm [13], whose running time is polynomial.

IV. 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 section 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 $\left[Q_{\Lambda_f}(x_k+d_k)\right] \bmod \Lambda$ to the decoder. Note that for a $1D$ lattice, the quantization operation reduces to a simple slicer. The decoder first subtracts back the dither and applies a $\bmod \Lambda$ opertaion to obtain $\tilde{x}_k \overset{(i.d.)}{=} [x_k+d_k] \bmod \Lambda$ and then chooses some full-rank matrix $\mathbf{A}\in\mathbb{Z}^{K\times K}$ and computes

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

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 $\widehat{\mathbf{Ax}}\neq\mathbf{A}(\mathbf{x}+\mathbf{d})$ 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 R. In [9] this probability is bounded as

$$\Pr\left(\widehat{\mathbf{Ax}}\neq\mathbf{A}(\mathbf{x}+\mathbf{d})\right)\leq 2K\exp\left\{-\frac{3}{2}2^{2(R-R_{\text{IF}}(\mathbf{A},d))}\right\},$$

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

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

Since d_k is statistically independent of $\mathbf{x}$ and $\mathbb{E}(d_k^2)=d$ for all $k=1,\ldots,K$, we see that provided that $\widehat{\mathbf{Ax}}=\mathbf{A}(\mathbf{x}+\mathbf{d})$ the one-shot version of IF source coding produces conditionally unbiased estimates of x_k with distortion d. The probability that $\widehat{\mathbf{Ax}}=\mathbf{A}(\mathbf{x}+\mathbf{d})$ 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 $\Pr\left(\widehat{\mathbf{Ax}}\neq\mathbf{A}(\mathbf{x}+\mathbf{d})\right)\leq 3\cdot 10^{-10}$. 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, $k=1,\ldots,K$ with probability greater than $1-2K\exp\{-\frac{3}{2}2^{2\Delta}\}$.

The next lemma, which is proved in [9], shows that the encoders in the $1D$ version of IF source coding can first reduce their observation modulo Λ and then quantize to the fine lattice Λ_f.

Lemma 2: 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

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

The advantage of 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 analog-to-digital converter (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 $R_{\text{IF}}(d)$. Such a *modulo-ADC* architecture can be useful for sampling spatially correlated signals. For instance, consider the Gaussian MIMO channel $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{z}$, 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. Replacing these ADCs with modulo-ADCs reduces to performing $1D$ integer-forcing source encoding. This allows the encoders 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. It should be noted that the modulo-ADC architecture was proposed in [14] in the context of quantized compute-and-forward. However, its advantages in the context of source coding were not addressed there.

V. Numerical Example

Consider the problem of distributively compressing a K-dimensional Gaussian source $\mathbf{x}$ with zero mean and covariance matrix $\mathbf{K}_{\mathbf{xx}} = \mathsf{SNR}\mathbf{H}\mathbf{H}^T + \mathbf{I}$ for some $\mathsf{SNR} > 0$ and some matrix $\mathbf{H} \in \mathbb{R}^{K \times M}$. This choice of covariance matrix corresponds to, e.g., the joint distribution of the signals observed by K relays in a two-hop Gaussian network with M users and K relays, where it is assumed that each relay observes a noisy linear combination of the signals transmitted by all users and 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.

We compare $R_{\text{IF}}(d)$, the minimal required symmetric compression rate of IF source coding, with that of a naive scheme that compresses each source using standard rate-distortion theory without exploiting the correlations between the sources, as well as with the symmetric rate for a successive Wyner-Ziv compression scheme. In the latter scheme, the kth relay compresses its observation assuming the decoder already possesses the compressed signals of relays $1, \ldots, k-1$. We plot the averages of the minimal required compression rates for the three schemes, i.e. the ergodic rate-distortion functions of the three schemes, along with the ergodic Berger-Tung benchmark rate-distortion function, under the assumption that the entries of $\mathbf{H}$ are i.i.d. standard normal random variables. Figure 2 depicts these rates for $\mathbf{H} \in \mathbb{R}^{8 \times 2}$ and $\mathsf{SNR} = 20$dB as a function of d. This choice of dimensions for $\mathbf{H}$ models a network with 2 transmitters and 8 relays. Such a choice of dimensions 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, as well as successive Wyner-Ziv compression. Nevertheless, as seen from Figure 2, the gap between the performance of the Berger-Tung benchmark and IF source coding is quite small.

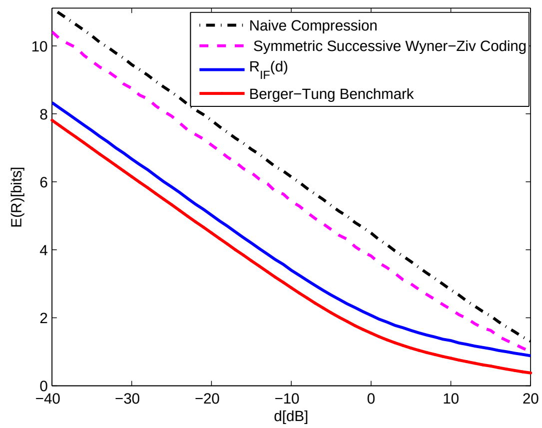

Fig. 2. $\mathbf{H} \in \mathbb{R}^{8 \times 2}$ with i.i.d. $\mathcal{N}(0,1)$ entries

References

[1] S.-Y. Tung, "Multiterminal source coding," Ph.D. dissertation, 1978.

[2] T. Berger, "Multiterminal source coding," July 1977.

[3] D. Krithivasan and S. S. Pradhan, "Lattices for distributed source coding: Jointly Gaussian sources and reconstruction of a linear function," *IEEE Trans. Inf. Theory*, Dec. 2009.

[4] A. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, May 2008.

[5] S. Pradhan and K. Ramchandran, "Distributed source coding: symmetric rates and applications to sensor networks," in *DCC*, 2000.

[6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, Mar. 2003.

[7] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, June 2002.

[8] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, Aug. 2004.

[9] O. Ordentlich and U. Erez, "Integer-forcing source coding," *IEEE Trans. Inf. Theory*, Submitted Aug. 2013, http://arxiv.org/abs/1308.6552.

[10] ——, "A simple proof for the existence of "good" pairs of nested lattices," in *IEEEI*, 2012.

[11] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, Oct. 2005.

[12] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathsf{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, Oct. 2004.

[13] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, 1982.

[14] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, Sep. 2013.