Network Compression: Worst-Case Analysis

Himanshu Asnani
Stanford University
Stanford, CA, USA
asnani@stanford.edu

Ilan Shomorony
Cornell University
Ithaca, NY, USA
is256@cornell.edu

A. Salman Avestimehr
Cornell University
Ithaca, NY, USA
avestimehr@ece.cornell.edu

Tsachy Weissman
Stanford University
Stanford, CA, USA
tsachy@stanford.edu

***Abstract*—We consider the problem of communicating a distributed correlated memoryless source over a memoryless network, from source nodes to destination nodes, under quadratic distortion constraints. We show the following two complementary results: (a) for an arbitrary memoryless network, among all distributed memoryless sources with a particular correlation, Gaussian sources are the worst compressible, that is, they admit the smallest set of achievable distortion tuples, and (b) for any arbitrarily distributed memoryless source to be communicated over a memoryless additive noise network, among all noise processes with a fixed correlation, Gaussian noise admits the smallest achievable set of distortion tuples. In each case, given a coding scheme for the corresponding Gaussian problem, we provide a technique for the construction of a new coding scheme that achieves the same distortion at the destination nodes in a non-Gaussian scenario with the same correlation structure.**

I. Introduction

Stochastic modeling of the data source and the communication medium are essential in data compression and data communication problems. However, extracting these descriptions from a practical system is in general difficult and often leads to intractable problems from a theoretical point of view. As a result, Gaussian models for both the data sources and the noise in communication networks prevail.

From a theoretical standpoint, one way of supporting the Gaussian assumption is by establishing that it is worst-case, meaning that, in a given family of distributions, the Gaussian distribution results in the smallest possible capacity or rate-distortion region. For example, it is known that, given a fixed noise variance, the Gaussian noise minimizes the capacity of a memoryless additive noise channel, and that, for a fixed-variance i.i.d. random source, the Gaussian distribution minimizes the rate-distortion region. These assertions can be proved using the fact that, subject to a variance constraint, the Gaussian distribution maximizes the entropy. In addition, a more operational proof of the worst-case noise assertion is provided in [1], where it is shown that Gaussian codebooks and nearest-neighbor decoding achieve the capacity of the corresponding AWGN channel on a non-Gaussian channel. Other worst-case noise characterizations in the literature include [2], where the worst-case additive noise in vector channels is shown to be Gaussian, and [3], where the worst-case noise of an additive noise channel with binary input is characterized. In terms of worst-case source results, one example is in [4], where it is shown that the worst-case source for the two-encoder quadratic source coding problem is jointly Gaussian.

Recently, a new approach was introduced in [5] that allowed to generalize the worst-case noise result from additive noise point-to-point channels to additive noise wireless networks. The framework in [5] can be described in two main steps. First, a DFT-based linear transformation is applied to all transmit and received signals in the network in order to create an effective network where the additive noise terms are "approximately Gaussian". Next, by demonstrating the optimality of coding schemes with finite reading precision (as we later explain in Section III-B) in Gaussian networks, it is proven that the capacity region of the Gaussian network is contained in the capacity region of the effective network asymptotically as the size of the blocks to which we apply the DFT-based transformation increases. This approach was later utilized in [6] to establish that Gaussian sources are worst-case data sources for distributed compression of correlated sources over rate-constrained, noiseless channels, with a quadratic distortion measure (in the context of the quadratic k-encoder source coding problem).

In this work, we pursue the analogue of these worst-case results in joint source-channel coding, by considering the problem of distributed compression of information over an arbitrary network. More precisely, k nodes in the network have access to correlated stochastic sources and wish to transmit them over an N-node network to respective destinations. The performance metric is the mean square error in the destinations' reconstruction of their desired sources. This problem lies at the heart of increasingly many applications concerning distributed compression of information over a network, such as sensor networks, weather monitoring systems, etc.

Since this setup involves the modeling of both the sources and the network, the worst-case characterization takes the form of two related sub-questions:

- **Question 1**: Given an arbitrary memoryless network, for a fixed correlation among its distributed memoryless components, are the jointly Gaussian sources the worst compressible? In other words, do they have the smallest set of achievable distortion tuples?
- **Question 2**: Given an arbitrary memoryless distributed source, for an additive noise network with a given noise correlation, is the Gaussian noise worst-case in the sense of having the smallest set of achievable distortion tuples?

In this paper, we answer both of these questions in the affirmative. This generalizes the result in [6] and extends the result in [5] to the unified framework of distributed source-channel coding. We utilize the aforementioned ideas to propose a universal way of converting a coding scheme designed under the Gaussian assumption into coding schemes that can handle and attain similar performances for non-Gaussian sources or noises. In particular, we start by using the DFT-based linear transformation as a way to make either the sources or the noises approximately Gaussian. Since this operation introduces a statistical dependence between the resulting sources or noises, an interleaving scheme is employed, in order to create blocks of i.i.d. approximately Gaussian sources and noises. Within each of the resulting blocks we then apply the original coding scheme designed under Gaussian models. The main technical novelty of this work is in the machinery required to show that such a scheme, when performed over sufficiently long blocks, can achieve distortions arbitrarily close to those achieved by the original coding scheme designed for Gaussian sources or noises. In particular, we introduce the concepts of coding schemes with *finite encoding precision* and *bounded outputs*. Then, by showing that the original coding scheme can be assumed without loss of generality to satisfy these two properties, as well as *finite reading precision* (as introduced in [5]), we can use standard tools regarding the convergence of random variables, such as the Dominated Convergence Theorem, to bound the distortion attained by the new coding scheme constructed based on the DFT-based linear transformation.

II. Problem Formulation and Main Results

We begin by introducing the notation used throughout this paper. If a random variable X has a probability density function, it is denoted as $f_X(x)$, and if the conditional distribution of X given Y has a conditional probability density function, it is denoted as $f_{X|Y}(x|y)$. X^t is a shorthand for the t-tuple $\{X[0], \cdots, X[t-1]\}$. The notation $[0:k]$ is shorthand for the set of natural numbers $\{0, 1, \cdots, k\}$ and $X[0:k] = \{X[0], X[1], ..., X[k]\}$. For random variables $X_1, X_2, ...$ and X, $X_n \xrightarrow{d} X$ means that X_n converges in distribution to X as $n \to \infty$.

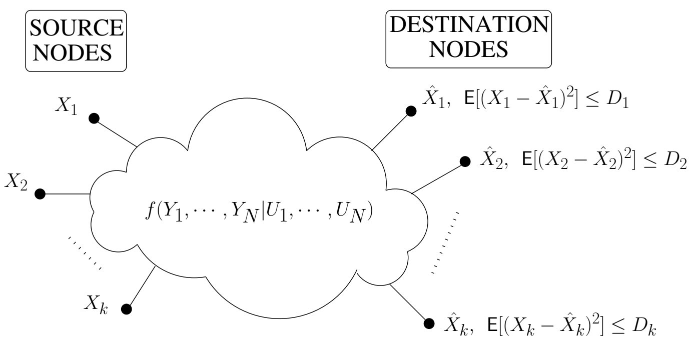

Fig. 1. A (k, N)-memoryless network.

A (k, N)-memoryless network, shown in Fig. 1, is characterized by the conditional density $f_{Y_1,\cdots,Y_N|U_1,\cdots,U_N}$, which relates the real valued network inputs $(U_1, \cdots, U_N)$ to real valued network outputs $(Y_1, \cdots, Y_N)$. The set of source nodes is denoted as $\mathcal{S} = \{s_1, s_2, \cdots, s_k\} \subseteq [1:N]$, and the set of destination nodes is denoted as $\mathcal{D} = \{d_1, d_2, \cdots, d_k\} \subseteq [1:N]$. We assume without loss of generality that $\mathcal{S} \cap \mathcal{D} = \emptyset$. The remaining nodes are relays $\mathcal{R} = \{r_1, r_2, \cdots, r_{N-2k}\} \subseteq [1:N]$. Source node $s_m \in \mathcal{S}$ has access to the i.i.d. source $X_m[t]$, $t = 0, 1, ...$, which must be communicated to the destination node $d_m \in \mathcal{D}$. The i.i.d. vectors $(X_1[t], ..., X_k[t])$ have a joint distribution with covariance matrix $\mathbf{K}$.

Definition 1. *A coding scheme $\mathcal{C}$ with block length $n \in \mathbf{N}$ for distributed compression over a (k, N) memoryless network consists of:*

1) *Source Encoding Functions: Source node $s_m \in \mathcal{S}$ encodes its source as $U_{s_m}[t] = f_{s_m,t}(\mathbf{X}_m, Y_{s_m}^{t-1})$, $\forall\, t \in [0:n-1]$, where $f_{s_m,t} : \mathbb{R}^n \times \mathbb{R}^{t-1} \to \mathbb{R}$, $m \in [1:k]$, $\forall\, t \in [0:n-1]$ are the source encoding functions.*
2) *Relay Encoding Functions: Relay node $r_p \in \mathcal{R}$ receives the channel outputs from the network and encodes them as $U_{r_p}[t] = f_{r_p,t}(Y_{r_p}^{t-1})$, $\forall\, t \in [0:n-1]$, where $f_{r_p,t} : \mathbb{R}^{t-1} \to \mathbb{R}$, $p \in [1:N-2k]$, $\forall\, t = [0:n-1]$ are the relay encoding functions.*
3) *Destination Encoding Functions: Destination node $d_m \in \mathcal{D}$ receives the channel outputs from the network and encodes them as $U_{d_m}[t] = f_{d_m,t}(Y_{d_m}^{t-1})$, $\forall\, t \in [0:n-1]$, where $f_{d_m,t} : \mathbb{R}^{t-1} \to \mathbb{R}$, $m \in [1:k]$, $\forall\, t \in [0:n-1]$ are the destination encoding functions.*
4) *Destination Decoding Functions: After N time steps, each destination $d_m \in \mathcal{D}$ constructs an estimate of the source as $\hat{\mathbf{X}}_m = g_{d_m}(\mathbf{Y}_{d_m})$, where $g_{d_m} : \mathbb{R}^n \to \mathbb{R}^n$, $m \in [1:k]$, are the destination decoding functions.*

Definition 2. *A distortion tuple $(D_1, D_2, \cdots, D_k)$ is achievable if for some block length n, there exists a coding scheme $\mathcal{C}$, as described above, such that,*

$$\frac{1}{n}\mathsf{E}\Big[\, \| \mathbf{X}_m - \hat{\mathbf{X}}_m \|^2 \,\Big] \le D_m, \ \forall\, m \in [1:k].$$

Definition 3. *The distortion region $\mathcal{D}$ of a (k, N)-memoryless network is the closure of the set of achievable distortion tuples.*

Theorem 1. *For a (k, N) memoryless network, let $\mathcal{D}_{NG}^{source}$ and $\mathcal{D}_{G}^{source}$ stand for the distortion regions for an arbitrary memoryless non-Gaussian source with covariance matrix $\mathbf{K}$ and for a memoryless Gaussian source with the same covariance matrix, respectively. Then we have*

$$\mathcal{D}_{G}^{source} \subseteq \mathcal{D}_{NG}^{source}. \tag{1}$$

Definition 4. *A (k, N)-memoryless network is an additive noise network if the input-output relationship is given by*

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = H \begin{bmatrix} U_1 \\ \vdots \\ U_N \end{bmatrix} + \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}, \tag{2}$$

where H is a real-valued $N \times N$ matrix and $(Z_1, ..., Z_N)$ is a noise vector with joint distribution $\mu_{\mathbf{Z}}$ independent of

$(U_1, ..., U_N)$. *If* $(Z_1, ..., Z_N)$ *is distributed as* $\mathcal{N}(\mathbf{0}, \mathbf{K})$ *for some covariance matrix* $\mathbf{K}$*, then we call the network a* (k, N)*-additive white Gaussian noise (AWGN) network.*

Theorem 2. *For a* (k, N) *memoryless additive noise network, let* $\mathcal{D}_{NG}^{noise}$ *and* $\mathcal{D}_{G}^{noise}$ *stand for the distortion regions for an arbitrary additive noise non-Gaussian distribution with covariance matrix* $\mathbf{K}$ *and for additive Gaussian noise with the same covariance matrix, respectively. Then we have*

$$\mathcal{D}_{G}^{noise} \subseteq \mathcal{D}_{NG}^{noise}. \tag{3}$$

III. Overview of Proofs

In Section III-A we describe the proof of Theorem 1 and, in Section III-B, we describe how the same tools can be used to prove Theorem 2. We refer to [7] for the complete proofs.

A. Proof of Theorem 1

The main idea is to use a coding scheme $\mathcal{C}$ with block length n for Gaussian sources to construct a new coding scheme $\tilde{\mathcal{C}}$ which achieves the same distortion tuple when the sources are non-Gaussian with the same covariance. Notice that we may assume without loss of generality that the sources have zero mean, since otherwise, we can remove the mean at the source nodes and add it back at the destination nodes.

The first step in the construction of this new coding scheme is to use the DFT-based linear transformation introduced in [5] in order to transform blocks of i.i.d. non-Gaussian random variables into "approximately Gaussian" random variables. More specifically, we define the unitary $b \times b$ matrix $\mathbf{Q}$ by setting the entry in the $(i+1)$th row and $(j+1)$th column be

$$Q(i,j) = \begin{cases} 1/\sqrt{b} & \text{if } i = 0 \\ \sqrt{2/b}\cos\left(\frac{2\pi j i}{b}\right) & \text{if } i = 1, ..., \frac{b}{2} - 1 \\ (-1)^j/\sqrt{b} & \text{if } i = \frac{b}{2} \\ \sqrt{2/b}\sin\left(\frac{2\pi j (i - b/2)}{b}\right) & \text{if } i = \frac{b}{2} + 1, ..., b - 1 \end{cases} \tag{4}$$

for $i, j \in \{0, ..., b-1\}$. Applying $\mathbf{Q}$ to a vector $\mathbf{x}$ can be intuitively seen as first taking the DFT of $\mathbf{x}$, then separating the real and imaginary parts of the resulting vector, and renormalizing them so that the resulting transformation is unitary. It is straightforward to check that $\mathbf{Q}$ is a unitary transformation, i.e., that $\|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^b$.

The fact that $\mathbf{Q}$ can make a random vector approximately Gaussian is expressed in the following lemma.

Lemma 1. *Suppose* $\{(X_1[i], ..., X_k[i])\}_{i=0}^{nb-1}$ *is an i.i.d. sequence of length-*k *zero-mean random vectors with covariance matrix* $\mathbf{K}$*, and let* $\mathbf{Q}$ *be the* $b \times b$ *matrix defined in (4) and*

$$\begin{bmatrix} \tilde{X}_1^{(0)}[t] & \cdots & \tilde{X}_k^{(0)}[t] \\ \vdots & \ddots & \vdots \\ \tilde{X}_1^{(b-1)}[t] & \cdots & \tilde{X}_k^{(b-1)}[t] \end{bmatrix} = \mathbf{Q} \begin{bmatrix} X_1[tb] & \cdots & X_k[tb] \\ \vdots & \ddots & \vdots \\ X_1[tb+b-1] & \cdots & X_k[tb+b-1] \end{bmatrix} \tag{5}$$

for $t = 0, 1, ..., n-1$*. Then, for any sequence* ℓ_b *such that, for* $b = 1, 2, ...,$ $\ell_b \in \{0, 1, ..., b-1\}$*, and any* $t \in \{0, 1, ..., n-1\}$*,*

$$\left(X_1^{(\ell_b)}[t], ..., \tilde{X}_k^{(\ell_b)}[t]\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{K}), \text{ as } b \to \infty.$$

To prove this result we use Lindeberg's Central Limit Theorem [8] to show that, for $\alpha_1, ..., \alpha_k \in \mathbb{R}$ and any sequence $\{\ell_b\}$, $\sum_{i=1}^{k} \alpha_i \tilde{X}_i^{(\ell_b)}[t]$ converges in distribution to a Gaussian random variable, and then invoke the Cramér-Wold Theorem.

To construct $\tilde{\mathcal{C}}$, we take n blocks of b source symbols, apply $\mathbf{Q}$ to each of them and then interleave the resulting symbols, obtaining b blocks $\tilde{\mathbf{X}}_m^{(\ell)}$, for $\ell = 0, ..., b-1$, each of length n, for each source s_m, referred to as the effective source symbols (cf. Figure 2). It can be seen that each of the b resulting length-n blocks have i.i.d. effective sources. We then apply the source encoding functions of $\mathcal{C}$, designed to achieve a given distortion tuple $(D_1, ..., D_k)$ with Gaussian sources, to these resulting i.i.d. blocks. The relay and destination encoding functions of $\tilde{\mathcal{C}}$ and $\mathcal{C}$ are the same, and the destination decoding functions of $\tilde{\mathcal{C}}$ are the composition of the destination decoding functions of $\mathcal{C}$ with a transformation that inverts the construction of the effective sources in Fig. 2.

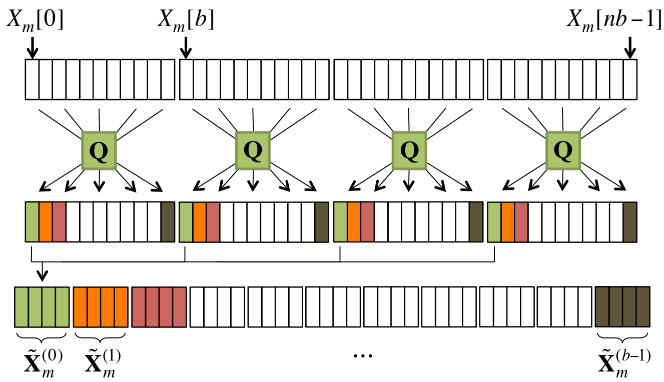

Fig. 2. Construction of the effective source by s_m.

Our main goal is to show that, as $b \to \infty$, the distortion of the resulting coding scheme $\tilde{\mathcal{C}}$ converges to $(D_1, ..., D_k)$. We begin with a lemma that allows us to just concentrate on *bounded output* coding schemes.

Lemma 2. *Suppose* $(X_1[t], ..., X_k[t])$ *has an arbitrary joint distribution with covariance matrix* $\mathbf{K}$ *and a coding scheme* $\mathcal{C}$ *with block length* n *achieves distortion vector* $(D_1, ..., D_k)$*. Then, for any* $\epsilon > 0$*, one can build another coding scheme* $\tilde{\mathcal{C}}$ *of block length* n *with decoding functions* $\tilde{g}_{d_m}$ *such that*

$$\left\|\tilde{g}_{d_j}(y_1, ..., y_n)\right\|_\infty \leq M,$$

for any $(y_1, ..., y_n) \in \mathbb{R}^n$*,* $j = 1, ..., k$ *and a fixed* $M > 0$*, which achieves distortion vector* $(D_1 + \epsilon, ..., D_k + \epsilon)$*.*

Another important property that we need to assume for the original coding scheme designed for a Gaussian model is that of *finite precision*. For a real-valued vector $x^n = (x_1, ..., x_n)$ and a positive integer ρ, we let $\lfloor x^n \rfloor_\rho = 2^{-\rho}\left(\lfloor 2^\rho x_1 \rfloor, ..., \lfloor 2^\rho x_n \rfloor\right)$, and define the following:

Definition 5. *A coding scheme $\mathcal{C}$ of block length n is said to have finite encoding precision $\rho = [\rho_1, \cdots, \rho_k] \in \mathbf{N}^k$ if the encoding function at each source $s_m \in \mathcal{S}$ satisfies*

$$f_{s_m,t}(x_m^n, y^{t-1}) = f_{s_m,t}(\lfloor x_m^n \rfloor_{\rho_m}, y^{t-1}),\ \forall\, m \in [1:k]$$

for any $x_m^n \in \mathbb{R}^n$, any $y^{t-1} \in \mathbb{R}^{t-1}$, and any time t.

Lemma 3. *Suppose the distortion tuple $(D_1, \cdots, D_k)$ is achievable over the (k, N)-memoryless network. Then for any $\epsilon > 0$, there exists a coding scheme with finite encoding precision that achieves distortion tuple $(D_1 + \epsilon, \cdots, D_k + \epsilon)$.*

From the proofs of Lemmas 2 and 3 [7], it can be seen that there exists a single coding scheme that has both bounded outputs and finite encoding precision and achieves distortion tuple $(D_1 + \epsilon, \cdots, D_k + \epsilon)$. Thus, we may assume initially that $\mathcal{C}$ has bounded outputs and finite encoding precision.

The importance of finite encoding precision is expressed in the following lemma.

Lemma 4. *If, for some $\rho \in \mathbb{N}$, $f : \mathbb{R}^a \to \mathbb{R}^b$ satisfies*

$$f(\mathbf{x}) = f(\lfloor \mathbf{x} \rfloor_\rho)$$

for any $\mathbf{x} \in \mathbb{R}^a$, f is locally constant (and thus continuous) almost everywhere.

Lastly we will need a lemma that allows us to view our stochastic network as a collection of deterministic networks, which helps in bounding the resulting distortion.

Lemma 5. *For any two random vectors $\mathbf{Y}$ and $\mathbf{U}$, there exist a (deterministic, measurable) function F and a random vector $\mathbf{Z}$, independent of $\mathbf{U}$, for which the pair $(F(\mathbf{U}, \mathbf{Z}), \mathbf{U})$ has the same distribution as $(\mathbf{Y}, \mathbf{U})$.*

This lemma implies that there exist functions F_{d_m}, for $d_m \in \mathcal{D}$, and a random vector $\mathbf{Z}$, such that, if the length-n source sequences are $\mathbf{x}_1, ..., \mathbf{x}_k$, then the length-$n$ block of received signals at destination d_m is given by $F_{d_m}(\mathbf{x}_1, ..., \mathbf{x}_k, \mathbf{Z})$. Therefore, since $\mathbf{Q}$ is a unitary linear transformation, for each realization $\mathbf{z}$ of $\mathbf{Z}$, the distortion of $\tilde{\mathcal{C}}$ can be written as

$$\frac{1}{b} \sum_{\ell=0}^{b-1} \frac{1}{n} \left\| \tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{z}) \right) \right\|^2. \tag{6}$$

For each $b = 1, 2, ...$, we choose ℓ_b such that the ℓ_bth length-n block has the largest expected distortion, i.e.,

$$\ell_b = \arg \max_{0 \le \ell \le b-1} \mathsf{E} \left\| \tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{z}) \right) \right\|^2.$$

Note that $\left\{ \left(\tilde{X}_1^{(\ell_b)}[i], ..., \tilde{X}_k^{(\ell_b)}[i] \right) \right\}_{i=0}^{n-1}$ is an i.i.d. sequence of length-k random vectors. From Lemma 1, we see that it converges in distribution to a sequence of i.i.d. jointly Gaussian random vectors with covariance matrix $\mathbf{K}$, as $b \to \infty$.

Each of the source encoding functions $f_{s_m,t}$ of the original coding scheme $\mathcal{C}$ is locally constant almost everywhere, since they were assumed to have finite encoding precision, by Lemma 4. In fact, this implies that the mapping

$$\left\{ \tilde{\mathbf{X}}_m^{(\ell_b)} \right\}_{m=1}^{k} \mapsto \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2,$$

for $m = 1, ..., k$, is continuous almost everywhere. Hence,

$$\begin{aligned} &\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \\ &\quad \xrightarrow{d} \left\| \mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{z}) \right) \right\|^2, \end{aligned}$$

as $b \to \infty$, where $\mathbf{X}_m^G = (X_m^G[0], ..., X_m^G[n-1])$, for $m = 1, ..., k$, and $\left\{ \left(X_1^G[i], ..., X_k^G[i] \right) \right\}_{i=0}^{n-1}$ is an i.i.d. sequence such that $(X_1^G[0], ..., X_k^G[0])$ is jointly Gaussian with zero mean and covariance matrix $\mathbf{K}$. Moreover, we have that

$$\begin{aligned} &\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \\ &\quad \le 2 \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 + 2 \left\| g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{z}) \right) \right\|^2 \\ &\quad \le 2 \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 + 2nM^2, \end{aligned} \tag{7}$$

and also that

$$\begin{aligned} \mathsf{E} \left\| \tilde{\mathbf{X}}_m^{(\ell_b)} \right\|^2 &= n\, \mathsf{E} \left(\sum_{j=0}^{b-1} X_m[j]\ Q(\ell_b, j) \right)^2 \\ &= n\mathbf{K}_{m,m} \sum_{j=0}^{b-1} Q^2(\ell_b, j) = n\mathbf{K}_{m,m} < \infty. \end{aligned} \tag{8}$$

Thus, from a variation of the Dominated Convergence Theorem (see Problem 16.4 in [8]), we conclude that, as $b \to \infty$,

$$\begin{aligned} &\mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z}) \right) \right\|^2 \middle| \mathbf{Z} = \mathbf{z} \right] \\ &\quad \to \mathsf{E} \left[\left\| \mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z}) \right) \right\|^2 \middle| \mathbf{Z} = \mathbf{z} \right], \end{aligned}$$

for all $\mathbf{z}$. The Dominated Convergence Theorem can then be used once again yielding

$$\begin{aligned} &\mathsf{E} \left[\mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z}) \right) \right\|^2 \middle| \mathbf{Z} \right] \right] \\ &\quad \to \mathsf{E} \left[\mathsf{E} \left[\left\| \mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z}) \right) \right\|^2 \middle| \mathbf{Z} \right] \right], \end{aligned}$$

which implies

$$\begin{aligned} &\mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z}) \right) \right\|^2 \right] \\ &\quad \to \mathsf{E} \left[\left\| \mathbf{X}_m^G - g_{d_m}\left(F_{d_m}(\mathbf{X}_1^G, ..., \mathbf{X}_k^G, \mathbf{Z}) \right) \right\|^2 \right] \le D_m. \end{aligned}$$

Therefore, we can choose b sufficiently large so that

$$\begin{aligned} &\frac{1}{n} \mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z}) \right) \right\|^2 \right] \\ &\quad \le D_m + \epsilon. \end{aligned}$$

The expected distortion of code $\tilde{\mathcal{C}}$ (with block length nb) thus satisfies, for $m = 1, ..., k$,

$$\begin{aligned} &\frac{1}{nb} \sum_{\ell=0}^{b-1} \mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell)}, ..., \tilde{\mathbf{X}}_k^{(\ell)}, \mathbf{Z}) \right) \right\|^2 \right] \\ &\quad \le \frac{1}{n} \mathsf{E} \left[\left\| \tilde{\mathbf{X}}_m^{(\ell_b)} - g_{d_m}\left(F_{d_m}(\tilde{\mathbf{X}}_1^{(\ell_b)}, ..., \tilde{\mathbf{X}}_k^{(\ell_b)}, \mathbf{Z}) \right) \right\|^2 \right] \\ &\quad \le D_m + \epsilon. \end{aligned}$$

B. Proof of Theorem 2

We start with a coding scheme $\mathcal{C}$ designed for an AWGN network. We then define the notion of *finite reading precision*.

Definition 6. *A coding scheme $\mathcal{C}$ of block length n is said to have finite reading precision $\rho = [\rho_1, \cdots, \rho_N] \in \mathbf{N}^N$ if the encoding function at each source $s_m \in \mathcal{S}$ satisfies*

$$f_{s_m,t}(x_m^n, y^{t-1}) = f_{s_m,t}(x_m^n, \lfloor y^{t-1} \rfloor_{\rho_{s_m}}),$$

and the encoding functions at each node $i \in \mathcal{R} \cup \mathcal{D}$ satisfies

$$f_{i,t}(y^{t-1}) = f_{i,t}(\lfloor y^{t-1} \rfloor_{\rho_i}),$$

for any $x_m^n \in \mathbb{R}^n$, any $y^{t-1} \in \mathbb{R}^{t-1}$, and any time t.

Lemma 6. *Suppose the distortion tuple $(D_1, \cdots, D_k)$ is achievable over the (k, N)-AWGN network. Then for any $\epsilon > 0$, there exists a coding scheme with finite reading precision that achieves distortion tuple $(D_1 + \epsilon, \cdots, D_k + \epsilon)$.*

Thus, w.l.o.g. we assume that $\mathcal{C}$ has both bounded outputs and finite reading precision. Moreover, we may assume that the additive noises have zero mean. As in the proof of Theorem 1, we use the linear transformation $\mathbf{Q}$ together with an interleaving procedure in order to create b blocks of n i.i.d. network uses. This time, each node i takes b blocks of n *effective* transmit signals, and interleaves them, obtaining n blocks of length b to which $\mathbf{Q}^{-1}$ is applied. In addition, each node i takes n blocks of b received signals, applies $\mathbf{Q}$ to each one, and interleaves the result, obtaining b blocks of n effective received signals $\tilde{\mathbf{Y}}_i^{(\ell)}$, for $\ell = 0, ..., b-1$. The resulting effective (additive) noises at i are referred to as $\tilde{\mathbf{Z}}_i^{(\ell)}$, for $\ell = 0, ..., b-1$. As in (6), we write the distortion of $\mathcal{C}$ as

$$\frac{1}{b} \sum_{\ell=0}^{b-1} \frac{1}{n} \left\| \mathbf{X}_m^{(\ell)} - g_{d_m} \left(F_{d_m} \left(\mathbf{X}_1^{(\ell)}, ..., \mathbf{X}_k^{(\ell)}, \tilde{\mathbf{Z}}_1^{(\ell)}, ..., \tilde{\mathbf{Z}}_N^{(\ell)} \right) \right) \right\|^2,$$

for some functions F_{d_m}, $d_m \in \mathcal{D}$. We then define ℓ_b as the value of $\ell \in \{0, ..., b-1\}$ that maximizes

$$\mathsf{E} \left\| \mathbf{X}_m^{(\ell)} - g_{d_m} \left(F_{d_m} \left(\mathbf{X}_1^{(\ell)}, ..., \mathbf{X}_k^{(\ell)}, \tilde{\mathbf{Z}}_1^{(\ell)}, ..., \tilde{\mathbf{Z}}_N^{(\ell)} \right) \right) \right\|,$$

for each $b = 1, 2, ...$ This time, we use Lemma 1 to establish that $\left(\tilde{Z}_1^{(\ell_b)}[t], \tilde{Z}_2^{(\ell_b)}[t], ..., \tilde{Z}_N^{(\ell_b)}[t] \right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{K})$. Then, since the mapping

$$\left\{ \mathbf{X}_1^{(\ell)}, ..., \mathbf{X}_k^{(\ell)}, \tilde{\mathbf{Z}}_1^{(\ell)}, ..., \tilde{\mathbf{Z}}_N^{(\ell)} \right\} \mapsto \left\| \mathbf{X}_m^{(\ell)} - g_{d_m} \left(F_m \left(\mathbf{X}_1^{(\ell)}, ..., \mathbf{X}_k^{(\ell)}, \tilde{\mathbf{Z}}_1^{(\ell)}, ..., \tilde{\mathbf{Z}}_N^{(\ell)} \right) \right) \right\|^2,$$

for $m = 1, ..., k$, is continuous almost everywhere, we have

$$\begin{aligned} &\left\| \mathbf{X}_m^{(\ell_b)} - g_{d_m} \left(F_m \left(\mathbf{X}_1^{(\ell_b)}, ..., \mathbf{X}_k^{(\ell_b)}, \tilde{\mathbf{Z}}_1^{(\ell_b)}, ..., \tilde{\mathbf{Z}}_N^{(\ell_b)} \right) \right) \right\|^2 \\ &= \left\| \mathbf{X}_m^{(0)} - g_{d_m} \left(F_m \left(\mathbf{X}_1^{(0)}, ..., \mathbf{X}_k^{(0)}, \tilde{\mathbf{Z}}_1^{(\ell_b)}, ..., \tilde{\mathbf{Z}}_N^{(\ell_b)} \right) \right) \right\|^2 \\ &\xrightarrow{d} \left\| \mathbf{X}_m^{(0)} - g_{d_m} \left(F_m \left(\mathbf{X}_1^{(0)}, ..., \mathbf{X}_k^{(0)}, \tilde{\mathbf{Z}}_1, ..., \tilde{\mathbf{Z}}_N \right) \right) \right\|^2, \end{aligned}$$

as $b \to \infty$, where $\tilde{\mathbf{Z}}_i = \tilde{Z}_i[0 : n-1]$ for $i = 1, ..., N$, and $\left\{ \left(\tilde{Z}_1[k], ..., \tilde{Z}_N[k] \right) \right\}_{k=0}^{n-1}$ is an i.i.d. sequence of $\mathcal{N}(\mathbf{0}, \mathbf{K})$ vectors. As in Section III-A, the Dominated Convergence Theorem is then used to conclude that, for large enough b,

$$\begin{aligned} \frac{1}{nb} \sum_{\ell=0}^{b-1} \mathsf{E} \left\| \mathbf{X}_m^{(\ell)} - g_{d_m} \left(F_m \left(\mathbf{X}_1^{(\ell)}, ..., \mathbf{X}_k^{(\ell)}, \tilde{\mathbf{Z}}_1^{(\ell)}, ..., \tilde{\mathbf{Z}}_N^{(\ell)} \right) \right) \right\|^2 \\ \leq D_m + \epsilon. \end{aligned}$$

IV. Conclusions

We showed that, in the problem of distributed compression of correlated sources over a network, the worst-case source under a covariance constraint is Gaussian, and, if we have an additive noise network, then the worst-case noise under a covariance constraint is also Gaussian. Moreover, we describe a systematic way of converting coding schemes designed under Gaussian assumptions into coding schemes that can handle non-Gaussian assumptions. The idea behind the construction of such schemes is simple conceptually, using DFT-based linear transformations, which renders it algorithmically tractable.

A research direction stemming from this work concerns finding outer bounds to the distortion region of Gaussian problems. This could be done by choosing special source or noise distributions (e.g., discrete), for which it is easier to obtain non-trivial outer bounds. Similar ideas can be found in [9], where the worst-case noise result in [5] was used to show that the capacity of a class of networks under a deterministic channel model is an outer bound to the capacity of the same network under an AWGN channel model.

V. Acknowledgements

The research of H. Asnani and T. Weissman is supported by Scott A. and Geraldine D. Macomber Stanford Graduate Fellowship and NSF Center for Science of Information (CSOI). The research of A. S. Avestimehr and I. Shomorony is supported by NSF Grants CAREER-0953117, CCF-1161720, and AFOSR Young Investigator Program Award.

References

[1] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1520–1529, September 1996.

[2] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, November 2001.

[3] S. Shamai and S. Verdu, "Worst-case power-constrained noise for binary-input channels," *IEEE Transactions on Information Theory*, vol. 38, no. 5, pp. 1494–1511, September 1992.

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

[5] I. Shomorony and A. S. Avestimehr, "Worst-case additive noise in wireless networks," *accepted for publication in IEEE Transactions on Information Theory*, 2012.

[6] I. Shomorony, A. S. Avestimehr, H. Asnani, and T. Weissman, "Worst-case source for distributed compression with quadratic distortion," *In Proc. of Information Theory Workshop (ITW)*, 2012.

[7] H. Asnani, I. Shomorony, A. S. Avestimehr, and T. Weissman, "Network compression: Worst-case analysis," *Preprint, available online at http://people.ece.cornell.edu/avestimehr/papers/wcNetComp.pdf.*

[8] P. Billingsley, *Probability and Measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1995.

[9] I. Shomorony and A. S. Avestimehr, "On the role of deterministic models in $K \times K \times K$ wireless networks," *Information Theory Workshop*, 2012.