

Compressive Sensing Over Networks

Soheil Feizi
MIT

Email: sfeizi@mit.edu

Muriel Médard
MIT

Email: medard@mit.edu

Michelle Effros
Caltech

Email: effros@caltech.edu

Abstract—In this paper, we demonstrate some applications of compressive sensing over networks. We make a connection between compressive sensing and traditional information theoretic techniques in source coding and channel coding. Our results provide an explicit trade-off between the rate and the decoding complexity. The key difference of compressive sensing and traditional information theoretic approaches is at their decoding side. Although optimal decoders to recover the original signal, compressed by source coding have high complexity, the compressive sensing decoder is a linear or convex optimization. First, we investigate applications of compressive sensing on distributed compression of correlated sources. Here, by using compressive sensing, we propose a compression scheme for a family of correlated sources with a modularized decoder, providing a trade-off between the compression rate and the decoding complexity. We call this scheme *Sparse Distributed Compression*. We use this compression scheme for a general multicast network with correlated sources. Here, we first decode some of the sources by a network decoding technique and then, we use a compressive sensing decoder to obtain the whole sources. Then, we investigate applications of compressive sensing on channel coding. We propose a coding scheme that combines compressive sensing and random channel coding for a high-SNR point-to-point Gaussian channel. We call this scheme *Sparse Channel Coding*. We propose a modularized decoder providing a trade-off between the capacity loss and the decoding complexity. At the receiver side, first, we use a compressive sensing decoder on a noisy signal to obtain a noisy estimate of the original signal and then, we apply a traditional channel coding decoder to find the original signal.

I. INTRODUCTION

Data compression has been a research topic of many scholars in past years. However, recently, the field of compressive sensing, originated in [1], [2] and [3], looked at the compression problem from another point of view. In this paper, we try to make a bridge between compressive sensing which may be viewed as a signal processing technique and both source coding and channel coding. By using this connection, we propose some non-trivial practical coding schemes that provide a trade-off between the rate and the decoding complexity.

Compressive sensing has provided a low complexity approximation to the signal reconstruction. Information theoretic has been mostly concerned with accuracy of the signal reconstruction under rate constraints. In this paper, we seek to provide new connections which use compressive sensing for traditional information theory problems such as Slepian-Wolf compression and channel coding in order to provide

mechanisms to explicitly trade-off between the decoding complexity and the rate.

Some previous work investigated the connection between compressive sensing and information theory. For example, reference [4] studied the minimum number of noisy measurements required to recover a sparse signal by using Shannon information theory bounds. Reference [5] investigated the contained information in noisy measurements by viewing the measurement system as an information theoretic channel and using the rate distortion function. Also, reference [6] studied the trade-offs between the number of measurements, the signal sparsity level, and the measurement noise level for exact support recovery of sparse signals by using an analogy between support recovery and communication over the Gaussian multiple access channel.

In this paper, we want to investigate applications of compressive sensing on source coding and channel coding. First, in Section III, we consider distributed compression of correlated sources. In this section, by using compressive sensing, we propose a compression scheme for a Slepian-Wolf problem with a family of correlated sources. The proposed decoder is a modularized decoder, providing a trade-off between the compression rate and the decoding complexity. We call this scheme *Sparse Distributed Compression*. Then, We use this compression scheme for a general multicast network with correlated sources. Here, we first decode some of the sources by a network decoding technique and then, we use the compressive sensing decoder to obtain the whole sources.

Then, in Section IV, we investigate applications of compressive sensing on channel coding. We propose a coding scheme that combines compressive sensing and random channel coding for a high-SNR point-to-point Gaussian channel. We call this scheme *Sparse Channel Coding*. We propose a modularized decoder, providing a trade-off between the capacity loss and the decoding complexity (i.e., the higher the capacity loss, the lower the decoding complexity.). The idea is to add intentionally some correlation to transmitted signals in order to decrease the decoding complexity. At the receiver side, first, we use a compressive sensing decoder on a noisy signal to obtain a noisy estimate of the original signal and then, we apply a traditional channel coding decoder to find the original signal.

II. TECHNICAL BACKGROUND AND PRIOR WORK

In this section, we review some prior results, in both compressive sensing and distributed compression, which will

be used in the rest of the paper.

A. Compressive Sensing Background

Let $\mathbf{X} \in \mathbb{R}^n$ be a signal vector. We say this signal is k -sparse if k of its coordinates are non-zero and the rest are zero. Define $\alpha = \frac{k}{n}$ as the sparsity ratio of the signal. Suppose Φ is a $n \times n$ measurement matrix (hence, $\mathbf{Y} = \Phi \mathbf{X}$ is the measurement vector.). If Φ is non-singular, by having \mathbf{Y} and Φ , one can recover \mathbf{X} . However, [2] and [3] showed that, one can recover the original sparse signal by having far fewer measurements than n . Suppose for a given n and k , m satisfies the following:

$$m \geq \rho k \log(n/k) \quad (1)$$

where ρ is a constant. Let \mathcal{S} be the power set of $\{1, 2, \dots, m\}$, and \mathcal{S}_m represent a member of \mathcal{S} with cardinality m (e.g., $\mathcal{S}_m = \{1, 2, \dots, m\}$). A $m \times n$ matrix $\Phi_{\mathcal{S}_m}$ is formed of rows of Φ whose indices belong to \mathcal{S}_m . $\Phi_{\mathcal{S}_m}$ represents an incomplete measurement matrix.

Reference [7] showed that, if m satisfies (1) and $\Phi_{\mathcal{S}_m}$ satisfies the *Restricted Isometry Property* (which will be explained later), having these m measurements is sufficient to recover the original sparse vector. In fact, it has been shown in [7] that, \mathbf{X} is the solution of the following linear programming:

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_{\mathcal{L}_1} \\ \text{subject to} \quad & \mathbf{Y} = \Phi_{\mathcal{S}_m} \mathbf{X}, \end{aligned} \quad (2)$$

where $\|\cdot\|_{\mathcal{L}_1}$ represents the \mathcal{L}_1 norm of a vector. Let us refer to the complexity of this optimization by $CX_{cs}^{noiseless}(m, n)$. We say $\Phi_{\mathcal{S}_m}$ satisfies the restricted isometry property (RIP) iff there exists $0 < \delta_k < 1$ such that, for any vector $\mathbf{V} \in \mathbb{R}^n$, we have:

$$(1 - \delta_k) \|\mathbf{V}\|_{\mathcal{L}_2}^2 \leq \|\Phi_{\mathcal{S}_m} \mathbf{V}\|_{\mathcal{L}_2}^2 \leq (1 + \delta_k) \|\mathbf{V}\|_{\mathcal{L}_2}^2. \quad (3)$$

Reference [7] showed if,

$$\delta_k + \delta_{2k} + \delta_{3k} < 1, \quad (4)$$

then, the optimization (2) can recover the original k -sparse signal. This condition leads to (1). Intuitively, if $\Phi_{\mathcal{S}_m}$ satisfies the RIP condition, it preserves the Euclidean length of k sparse signals. Hence, these sparse signals cannot be in the null-space of $\Phi_{\mathcal{S}_m}$. Moreover, any k subset of columns of $\Phi_{\mathcal{S}_m}$ are nearly orthogonal if $\Phi_{\mathcal{S}_m}$ satisfies RIP. Most of matrices which satisfy RIP are random matrices. Reference [8] showed a connection between compressive sensing, n -widths, and the Johnson-Lindenstrauss Lemma. Through this connection, [9] provided an elegant way to find δ_k by using the concentration theorem for random matrices and proposed some data-base friendly matrices (binary matrices) satisfying RIP.

Now, suppose our measurements are noisy (i.e., $\mathbf{Y} = \Phi_{\mathcal{S}_m} \mathbf{X} + \mathbf{Z}$, where $\|\mathbf{Z}\|_{\mathcal{L}_2} = \epsilon$). Then, reference [1] showed that if $\Phi_{\mathcal{S}_m}$ satisfies RIP (3) and \mathbf{X} is sufficiently sparse, the

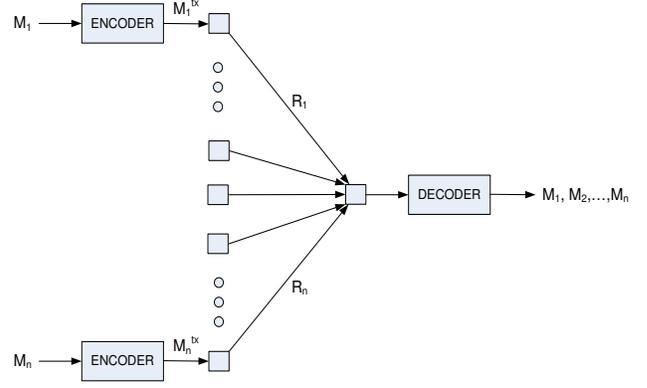


Fig. 1. A depth one multicast tree with n sources.

following optimization recovers the original signal up to the noise level:

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_{\mathcal{L}_1} \\ \text{subject to} \quad & \|\mathbf{Y} - \Phi_{\mathcal{S}_m} \mathbf{X}\|_{\mathcal{L}_2} \leq \epsilon. \end{aligned} \quad (5)$$

Hence, if \mathbf{X}^* is a solution of (5), then $\|\mathbf{X} - \mathbf{X}^*\|_{\mathcal{L}_2} \leq \beta\epsilon$, where β is a constant. To have this, the original vector \mathbf{X} should be sufficiently sparse. Specifically, [1] showed that, if

$$\delta_{3k} + 3\delta_{4k} < 2, \quad (6)$$

then, the results hold. We refer to the complexity of this optimization by $CX_{cs}^{noisy}(m, n)$.

B. Distributed Compression Background

In this section, we briefly review some prior results in distributed compression. In Section III, we will use these results along with compressive sensing results, making a bridge these two subjects.

Consider the network depicted in Figure 1. This is a depth one tree network with n sources. Suppose each source i , has messages referred by M_i , wishes to send to the receiver. To do this, each source i sends its message with a rate R_i (i.e., say \mathbf{m}_i has m messages of source i . This source maps \mathbf{m}_i to $\{1, 2, \dots, 2^{mR_i}\}$). Let $R = \sum_{i=1}^n R_i$. According to Slepian-Wolf compression ([10]), one needs to have,

$$R \geq H(M_1, M_2, \dots, M_n). \quad (7)$$

By using linear network coding, reference [11] extended Slepian-Wolf compression for a general multicast problem. Note that, the proposed decoders in [10] and [11] (a minimum entropy or a maximum probability decoder) have high complexity ([12]). For a depth one tree network, references [13], [14], [15], [16], [17] and [18] provide some low complexity decoding techniques for Slepian-Wolf compression. However, none of these methods can explicitly provide a trade-off between the rate and the decoding complexity. For a general multicast network, it has been shown in [19] that, there is no separation between Slepian-Wolf compression and network

coding. For this network, a low complexity decoder for an optimal compression scheme has not been proposed yet.

III. COMPRESSIVE SENSING AND DISTRIBUTED SOURCE CODING

In this part, we propose a distributed compression scheme for correlated sources providing a trade-off between the decoding complexity and the compression rate. References [10] and [11] propose optimal distributed compression schemes for correlated sources, for depth-one trees and general networks, respectively. However, their proposed compression schemes need to use a minimum entropy or a maximum probability decoder, which has high complexity ([12]). On the other hand, compressive sensing provides a low complexity decoder, by using a convex optimization, to recover a sparse vector from incomplete and contaminated observations (e.g., [1], [2], [3] and [7]). Our aim is to use compressive sensing over networks to design compression schemes for correlated sources.

Note that, our proposed techniques can be applied on different network topologies, with different assumptions and constraints. Here, to illustrate the main ideas, first we consider a noiseless tree network with depth one and with correlated sources. Then, in Section III-B, we extend our techniques to a general multicast network.

A. A one-stage tree with n -correlated sources

Consider the network shown in Figure 1, which is a one-stage noiseless multicast tree network with n -sources. First, consider a case where sources are independent. Hence, the problem reduces to a classical source coding problem. Say each source is transmitting with a rate R_i . Denote $R = \sum_{i=1}^n R_i$. It is well-known that the following sum-rate for this network is the minimum required sum-rate:

$$R \geq \sum_{i=1}^n H(M_i), \quad (8)$$

where M_i is the message random variable of source i . Let us refer to the complexity of its decoder at the receiver as $CX_{indep}(n)$, since sources are independent.

Now, we consider a case where sources are correlated. We formulate the correlation of sources as follows:

Definition 1. Suppose $\mu_{i,t}$ represents a realization of the i^{th} source message random variable at time t (i.e., a realization of M_i at time t). Hence, the vector $\bar{\mu}_t = \{\mu_{1,t}, \dots, \mu_{n,t}\}$ is the sources' message vector at time t . We drop the subscript t when no confusion arises. Suppose sources are correlated so that there exists a transform matrix under which the sources' message vector is k -sparse (i.e., there exists a $n \times n$ transform matrix Φ and a k -sparse vector $\bar{\mu}'_t$ such that $\bar{\mu}_t = \Phi \bar{\mu}'_t$). Let \mathcal{S} be the power set of $\{1, 2, \dots, n\}$, and \mathcal{S}_m denote members of \mathcal{S} whose cardinality are m . Suppose $\Phi_{\mathcal{S}_m}$, defined as in Section II-A, satisfies RIP (3), where m satisfies (1) for a given n and k . We have, $\bar{\mu}_{t,\mathcal{S}_m} = \Phi_{\mathcal{S}_m} \bar{\mu}'_t$, where $\bar{\mu}_{t,\mathcal{S}_m}$ is components of $\bar{\mu}_t$ whose indices belong to \mathcal{S}_m . We refer to these sources as k -sparsely correlated sources.

If we know zero locations of $\bar{\mu}'_t$ and if the transform matrix is non-singular, by having each subset containing k sources, the whole vector $\bar{\mu}_t$ and $\bar{\mu}'_t$ can be computed. Hence, by the data processing inequality, we have,

$$H(M_{i_1}, \dots, M_{i_k}) + nH_b(\alpha) = H(M_1, \dots, M_n), \quad (9)$$

for and $1 \leq i_1 < i_2 < \dots < i_k \leq n$ where M_{i_j} is the message of source i_j . Also, by [2], having m sources, where m satisfies (1) for a given k and n , allows us to recover the whole vector $\bar{\mu}'_t$ and therefore, $\bar{\mu}_t$. Hence,

$$H(M_{i_1}, \dots, M_{i_m}) = H(M_1, \dots, M_n). \quad (10)$$

Example 2. For an example of k -sparsely correlated sources, suppose each coordinates of $\bar{\mu}'_t$ is zero with probability $1-\alpha$, otherwise it is uniformly distributed over $\{1, 2, \dots, 2^R\}$. Suppose the entries of Φ are independent realizations of Bernoulli random variables:

$$\Phi_{i,j} = \begin{cases} \frac{1}{\sqrt{m}} & \text{with prob. } \frac{1}{2} \\ -\frac{1}{\sqrt{m}} & \text{with prob. } \frac{1}{2} \end{cases}. \quad (11)$$

Thus, as shown in [9], $\Phi_{\mathcal{S}_m}$ satisfies RIP (3). Also, we have,

$$\begin{aligned} H(M_1, \dots, M_n) &= nH_b(\alpha) + k(R+1) \\ &\approx nH_b(\alpha) + kR. \end{aligned} \quad (12)$$

The entropy of each source random variable for large n can be computed approximately as follows:

$$H(M_i) \approx R + \frac{1}{2} \log(k). \quad (13)$$

Note that, the individual entropies of k -sparsely correlated sources are roughly the same. Now, we want to use compressive sensing to have a distributed source coding providing a trade-off between the compression rate and the decoding complexity.

We shall show that, if sources are k -sparsely correlated, if we have m of them at the receiver at each time t , by using a linear programming, all sources can be recovered at that time. Hence, instead of sending n correlated sources, one needs to transmit m of them where $m \ll n$. We assume that the entropies of sources are the same. Hence, each source transmits with probability $\gamma = \frac{m}{n}$. Therefore, by the law of large numbers, for large enough n , we have m transmitting sources. For the case of non-equal source entropies, one can a priori choose m sources with the lowest entropy as the transmitting sources.

In Table III-A, we compare four different schemes:

- The first scheme which Theorem 3 is about, is called *Sparse Distributed Compression* with independent coding (SDCIC). The idea is to send m sources from n of them, assuming they are independent (i.e., their correlation is not used in coding). In this case, at the receiver, first we decode these m transmitted messages

TABLE I
COMPARISON OF DISTRIBUTED COMPRESSION METHODS OF
CORRELATED SOURCES

Compression Methods	Minimum Min-Cut Rate	Decoding Complexity
SDCIC	$\sum_{j=1}^m H(M_{i_j})$	$CX_{indep}(m) + CX_{cs}^{noiseless}(m, n)$
Slepian-Wolf (SW)	$H(M_1, \dots, M_n)$	$CX_{sw}(n)$
Combination of SDC and SW	$H(M_1, \dots, M_n)$	$CX_{sw}(m) + CX_{cs}^{noiseless}(m, n)$
Naive Correlation Ignorance	$\sum_{i=1}^n H(M_i)$	$CX_{indep}(n)$

and then, we use a compressive sensing decoder to recover the whole sources. The decoding complexity is $CX_{indep}(m) + CX_{cs}^{noiseless}(m, n)$. The required min-cut rate between each receiver and sources is $\sum_{j=1}^m H(M_{i_j})$.

- In the second scheme referred in Table III-A as the *Slepian-Wolf* coding, one performs an optimal source coding on correlated sources. The required min-cut rate between sources and the receiver is $H(M_1, \dots, M_n)$ which is less than $\sum_{j=1}^m H(M_{i_j})$. However, at the receiver, we need to have a Slepian-Wolf decoder with complexity $CX_{sw}(n)$.
- One can have a combination of sparse distributed compression and Slepian-Wolf compression. To do this, instead of n sources, we transmit m of them by using a Slepian-Wolf coding. The required min-cut rate would be $H(M_{i_1}, \dots, M_{i_m})$ which by (10), $H(M_1, \dots, M_n) = H(M_{i_1}, \dots, M_{i_m}) \leq \sum_{j=1}^m H(M_{i_j})$. At the receiver, first we use a Slepian-Wolf decoder with complexity $CX_{sw}(m)$ to decode the m transmitted sources' information. Then, we use a compressive sensing decoder to recover the whole sources.
- The fourth method is a naive way of transmitting n correlated sources so that we have an easy decoder at the receiver. We call this scheme a *Naive Correlation Ignorance* method. In this method, we simply ignore the correlation in the coding scheme. In this naive scheme, the required min-cut rate is $\sum_{i=1}^n H(M_i)$ with a decoding complexity equal to $CX_{indep}(n)$. Note that, in sparse distributed compression with independent coding, the required min-cut rate is much less than this scheme ($\sum_{j=1}^m H(M_{i_j}) \ll \sum_{i=1}^n H(M_i)$).

The following theorem is about sparse distributed compression with independent coding:

Theorem 3. *For a depth one tree network with n -sources, k -sparsely correlated, the required min-cut rate between the receiver and sources in sparse distributed compression with independent coding (SDCIC) is,*

$$R \geq \sum_{j=1}^m H(M_{i_j}), \quad (14)$$

where m is the smallest number satisfies (1) for a given n and k and i_j is the index of j^{th} active source. Also, the decoding complexity of SDCIC method is $CX_{indep}(m) + CX_{cs}^{noiseless}(m, n)$.

Proof: For a given n (number of sources) and k (the

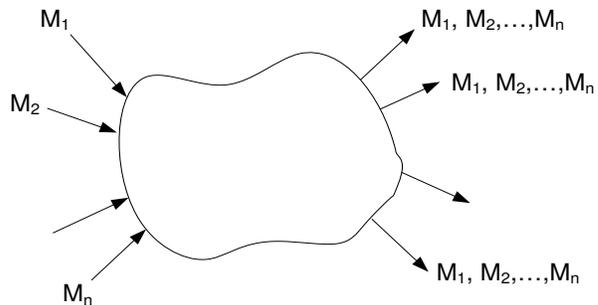


Fig. 2. A general multicast network with n sources.

sparsity factor), we choose m to satisfy (1). Now, suppose each source is transmitting its block with probability γ where $\gamma = \frac{m}{n}$. Hence, by the law of large numbers, the probability that we have m transmitting sources' blocks goes to one for sufficiently large n . Without loss of generality, say sources 1, ..., m are transmitting (i.e., active sources) and other sources are idle. If $R_i \geq H(M_i)$ for active sources and zero for idle sources, active sources can transmit their messages to the receiver and the required min-cut rate between the receiver and sources would be $R \geq \sum_{j=1}^m H(M_{i_j})$. Let \mathcal{S}_m indicate active sources' indices. First, at the receiver, we decode these m active sources. The complexity of this decoding part is $CX_{indep}(m)$ because we did not use their correlation in the coding scheme. Hence, we have $\overline{\mu_{t, \mathcal{S}_m}}$, active sources' messages at time t at the receiver (in this case, $\overline{\mu_{t, \mathcal{S}_m}} = (\mu_{1,t}, \dots, \mu_{m,t})$). Then, by having $\overline{\mu_{t, \mathcal{S}_m}}$ and using the sparsity of $\overline{\mu_t'}$, the following optimization can recover the whole sources:

$$\begin{aligned} \min \quad & \|\overline{\mu_t'}\|_{\mathcal{L}_1} \\ \text{subject to} \quad & \overline{\mu_{t, \mathcal{S}_m}} = \Phi_{\mathcal{S}_m} \overline{\mu_t'}. \end{aligned} \quad (15)$$

The overall decoding complexity is $CX_{indep}(m) + CX_{cs}^{noiseless}(m, n)$. ■

B. A General Multicast Network with Correlated Sources

In this section, we extend results of Section III-A to a general noiseless multicast network. Note that, for a depth one tree network, references [13], [14], [15], [16], [17] and [18] provide some low complexity decoding techniques for Slepian-Wolf compression. However, none of these methods can explicitly provide a trade-off between the rate and the decoding complexity. For a general multicast network, it has been shown in [19] that, there is no separation between Slepian-Wolf compression and network coding. For this network, a low complexity decoder for an optimal compression scheme has not been proposed yet. There are few practical approaches and they rely on small topologies [20]. In this section, by using compressive sensing, we propose a coding scheme which provides a trade-off between the compression rate and the decoding complexity.

Consider a multicast network shown in Figure 2 which has n sources and T receivers. If sources are independent, reference [21] showed that the minimum required min-cut rate between each receiver and sources is as (8). This rate can be achieved by using linear network coding over the network. Reference [11] showed that random linear network coding can perform arbitrarily closely to this rate bound. Say $CX_{indep}^{nc}(n)$ is the complexity of its decoder.

Now, consider a case when sources are correlated. Reference [11] extended Slepian-Wolf compression result ([10]) to a multicast network by using network coding. The proposed decoder is a minimum entropy or a maximum probability decoder whose complexity is high ([12]). We refer to its complexity by $CX_{sw}^{nc}(n)$.

In this section, we consider sources to be k -sparsely correlated. We want to use compressive sensing along with network coding to propose a distributed compression scheme providing a trade-off between the compression rate and the decoding complexity.

In the following, we compare four distributed compression methods for a general multicast network:

- In the first scheme, we use sparse distributed compression with independent coding. With high probability, we have m active sources. Then, we perform network coding on these active sources over the network, assuming they are independent. At the receiver, first we decode these m active sources and then, by a compressive sensing decoder, we recover the whole sources. The decoding complexity is $CX_{indep}^{nc}(m) + CX_{cs}^{noiseless}(m, n)$. The required min-cut rate between each receiver and sources is $\sum_{j=1}^m H(M_{i_j})$.
- In the second scheme, we use an extended version of Slepian-Wolf compression for a multicast network ([22]). Reference [22] considers a vector linear network code that operates on blocks of bits. The required min-cut rate between sources and the receiver is $H(M_1, \dots, M_n)$. At the receiver, one needs to have a minimum entropy or a maximum probability decoder with complexity $CX_{sw}^{nc}(n)$.
- A combination of sparse distributed compression and Slepian-Wolf compression can provide a trade-off between the compression rate and the decoding complexity. For example, instead of n sources, one can transmit m of them by using a Slepian-Wolf coding for multicast networks. The required min-cut rate would be $H(M_{i_1}, \dots, M_{i_m})$. At the receiver, first we use a Slepian-Wolf decoder with complexity $CX_{sw}^{nc}(m)$ to decode m active sources. Then, we use a compressive sensing decoder to recover the whole sources.
- In a naive correlation ignorance method, we simply ignore the correlation in the coding scheme. In this naive scheme, the required min-cut rate is $\sum_{i=1}^n H(M_i)$ with a decoding complexity equal to $CX_{indep}^{nc}(n)$.

The following theorem is about sparse distributed compression with independent coding for a multicast network:

Theorem 4. *For a general multicast network with n -sources,*

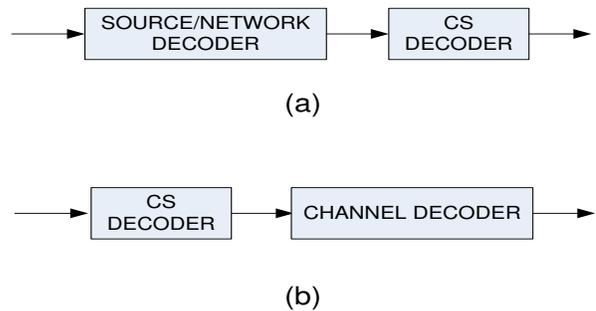


Fig. 3. A modularized decoder for a) compressive sensing with source coding, (b) compressive sensing with channel coding.

k -sparsely correlated, the required min-cut rate between each receiver and sources in sparse distributed compression with independent coding is,

$$R \geq \sum_{j=1}^m H(M_{i_j}), \quad (16)$$

where m is the smallest number satisfies (1) for a given n and k . Also, the decoding complexity is $CX_{indep}^{nc}(m) + CX_{cs}^{noiseless}(m, n)$.

Proof: The proof is similar to the one of Theorem 3. The only difference is that, here, one needs to perform linear network coding on active sources over the network without using the correlation among them. At the receiver, first, these m active sources' blocks are decoded (the decoding complexity of this part is $CX_{indep}^{nc}(m)$). Then, a compressive sensing decoder (15) is used to recover the whole sources. ■

Remark: In this section, we explained how compressive sensing can be used with distributed source coding. At the receiver side, first, we decode the active sources by a network decoding technique and then, we use a compressive sensing decoder to recover the whole sources. This high level modularized decoding scheme of sparse distributed compression is depicted in Figure 3-a. On the other hand, when we use compressive sensing for channel coding (which will be explained in detail in Section IV), we switch these two decoding modules (as shown in Figure 3-b). In other words, first, we use a compressive sensing decoder on a noisy signal to obtain a noisy estimate of the original signal and then, we use a channel decoder to find the original message from this noisy estimate.

IV. COMPRESSIVE SENSING AND CHANNEL CODING

In this section, we make a bridge between compressive sensing, which is more a signal processing technique, and channel coding. We consider a high-SNR Gaussian point-to-point channel depicted in Figure 4. We propose a coding scheme that combines compressive sensing and random channel coding with a modularized decoder to provide a trade-off between the capacity loss and the decoding complexity (i.e., the higher the capacity loss, the lower the decoding

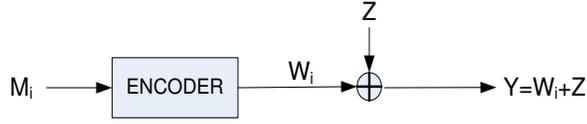


Fig. 4. A Point-to-Point Gaussian Channel

complexity.). We call this scheme *Sparse Channel Coding*. We add intentionally some correlation to transmitted signals to decrease the decoding complexity. This is an example of how compressive sensing can be used with channel coding and could be extended to other types of channels.

Consider a point to point channel with additive Gaussian noise with noise power N and transmission power constraint P , shown in Figure 4. Suppose we are in a high SNR regime (i.e., $\frac{P}{N} \gg 1$). The capacity of this channel is,

$$C \approx \frac{1}{2} \log\left(\frac{P}{N}\right). \quad (17)$$

An encoding and a decoding scheme for this channel can be found in [12]. As explained in [12], at the receiver, one needs to use a m -dimensional maximum likelihood decoder where m , the code length, is arbitrarily large. We refer to the complexity of such a decoder as $CX_{ml}(m)$. Our aim is to design a coding scheme to have a trade-off between the rate and the decoding complexity.

Suppose m is arbitrarily large. Choose n and k to satisfy (6).

Definition 5. A $(2^{mR}, m)$ sparse channel code for a point-to-point Gaussian channel with power constraint P satisfies the following:

- *Encoding process:* message i from the set $\{1, 2, \dots, 2^{mR}\}$ is assigned to a m -length vector \mathbf{W}_i , satisfying the power constraint; that is, for each codeword, we have,

$$\frac{1}{m} \sum_{j=1}^m W_{ij}^2 \leq P, \quad (18)$$

where W_{ij} is the j^{th} coordinate of codeword \mathbf{W}_i .

- *Decoding process:* the receiver receives \mathbf{Y} , a noisy version of the transmitted codeword \mathbf{W}_i (i.e., $\mathbf{Y} = \mathbf{W}_i + \mathbf{Z}$, where \mathbf{Z} is a Gaussian noise vector with i.i.d. coordinates with power N). A decoding function g maps \mathbf{Y} to the set $\{1, 2, \dots, 2^{mR}\}$. The decoding complexity is $CX_{ml}(k) + CX_{cs}^{noisy}(m, n)$.

A rate R is achievable if $P_e^m \rightarrow 0$ as $m \rightarrow \infty$, where,

$$P_e^m = \frac{1}{2^{mR}} \sum_{i=1}^{2^{mR}} \Pr(g(\mathbf{Y}) \neq i | M = i), \quad (19)$$

and M represents the transmitted message.

Theorem 6. For a high-SNR Gaussian point-to-point channel described in Definition 5, the following rate is achievable, while the decoding complexity is $CX_{ml}(k) + CX_{cs}^{noisy}(m, n)$:

$$R = \frac{k}{m}C + \frac{n}{m}H_b(\alpha) + \frac{k}{2m} \log\left(\frac{1}{\beta(1+\delta_k)}\right), \quad (20)$$

where C is the channel capacity (17), α , β and δ_k are parameters defined in Section II-A, and $H_b(\cdot)$ is the binary entropy function.

Before expressing the proof, let us make some remarks:

- if we have (1) (i.e., $m/n = \rho\alpha \log(1/\alpha)$), the achievable rate can be approximated as follows:

$$R \approx \frac{1}{\log(1/\alpha)}C + \frac{1}{\alpha \log(1/\alpha)}H_b(\alpha) \quad (21)$$

where $\alpha = k/n$ is the sparsity ratio. The first term of (21) shows that the capacity loss factor is a log function of the sparsity ratio. The term $\frac{1}{\alpha \log(1/\alpha)}H_b(\alpha)$ is the rate gain that we obtain by using compressive sensing in our scheme. Note that, since the SNR is sufficiently high, the overall rate would be less than the capacity.

- The complexity term $CX_{ml}(k)$ is an exponential function, while $CX_{cs}^{noisy}(m, n)$ is a polynomial.

Proof: We use random coding and compressive sensing arguments to propose a concatenated channel coding scheme satisfying (20). First, we explain the encoding part:

- Each message i is mapped randomly to a k -sparse vector \mathbf{X}_i with length n so that its non-zero coordinates are i.i.d. drawn from $\mathcal{N}(0, \frac{m}{k(1+\delta_k)}P)$. n is chosen to satisfy (6), for a given k and m .
- To obtain \mathbf{W}_i , we multiply \mathbf{X}_i by a $m \times n$ matrix (Φ_{S_m}) satisfying the RIP condition (3) (i.e., $\mathbf{W}_i = \Phi_{S_m} \mathbf{X}_i$). We send \mathbf{W}_i through the channel.

In fact, it is a concatenated channel coding scheme [23]. The outer layer of this coding is the sparsity pattern of \mathbf{X}_i . The inner layer of this coding is based on random coding around each sparsity pattern. To satisfy transmission power constraint P , we generate each non-zero coordinate of \mathbf{X}_i by $\mathcal{N}(0, \frac{m}{k(1+\delta_k)}P)$. Note that, the encoding matrix Φ_{S_m} is known to both encoding and decoding sides.

At the receiver, we have a noisy version of the transmitted signal. Suppose message i has been transmitted. Hence, $\mathbf{Y} = \mathbf{W}_i + \mathbf{Z} = \Phi_{S_m} \mathbf{X}_i + \mathbf{Z}$. The decoding part is as follows:

- Since \mathbf{X}_i is k -sparse, first, we use a compressive sensing decoder to find $\tilde{\mathbf{X}}_i$ such that $\|\mathbf{X}_i - \tilde{\mathbf{X}}_i\|_{\mathcal{L}_2}^2 \leq \beta mN$, as follows, where β is the parameter of (5):

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_{\mathcal{L}_1} \\ \text{subject to} \quad & \frac{1}{m} \|\mathbf{Y} - \Phi_{S_m} \mathbf{X}_i\|_{\mathcal{L}_2}^2 \leq N. \end{aligned} \quad (22)$$

We assume that the complexity of this convex optimization is smaller than the one of a maximum likelihood decoder.

- Since we are in a high-SNR regime, by having $\tilde{\mathbf{X}}_i$, we can find the sparsity pattern of \mathbf{X}_i . It gives us the outer layer code of message i . The higher the outer layer code rate, the higher the required SNR. We shall develop this

argument with more detail later (26).

- Having the sparsity pattern of \mathbf{X}_i , we use a maximum likelihood decoder in a k -dimensional space (i.e., non-zero coordinates of \mathbf{X}_i) to find non-zero coordinates of \mathbf{X}_i (the inner layer code). The complexity of this decoder is denoted by $CX_{ml}(k)$.

Before presenting the error probability analysis, let us discuss different terms of (20). Since in our coding scheme, for each message i , we send a k -sparse signal by sending a m -length vector, we have the fraction $\frac{k}{m}$ before the capacity term C . The capacity term C in (20) comes from the inner layer coding scheme. The term $\frac{n}{m}H_b(\alpha)$ comes from the outer layer coding scheme. Note that, since α is a small number, $H_b(\alpha)$ is close to one. Also, the ratio $\frac{n}{m}$ depends on the outer layer code rate. Here, we assumed the SNR is sufficiently high so that we can use all possible outer layer codes. The term $\frac{k}{2m} \log(\frac{1}{\beta(1+\delta_k)})$ is because of the power constraint and the RIP condition. Note that, if one performs a time-sharing based channel coding, a rate $\frac{k}{m}C$ is achievable with a decoding complexity $CX_{ml}(k)$. By using the compressive sensing, we obtain additional rate terms because of the outer layer coding with approximately the same complexity.

We proceed by the error probability analysis. Define Ξ_i as the sparsity pattern of \mathbf{X}_i (i.e., $\Xi_{ij} = 0$ if $X_{ij} = 0$, otherwise, $\Xi_{ij} = 1$). Also, define $\tilde{\Xi}_i$ as the decoded sparsity pattern of message i . Without loss of generality, assume that message 1 was transmitted. Thus, $\mathbf{Y} = \Phi_{S_m} \mathbf{X}_1 + \mathbf{Z}$. Define the following events:

$$\begin{aligned} E_0 &= \left\{ \frac{1}{m} \sum_{j=1}^m W_{1j}^2 > P \right\} \\ E_i &= \left\{ \frac{1}{n} \|\mathbf{X}_i - \tilde{\mathbf{X}}_i\|_{\mathcal{L}_2}^2 \leq \frac{\beta m}{n} N \right\} \\ E_{p_1} &= \{\tilde{\Xi}_1 \neq \Xi_1\}. \end{aligned} \quad (23)$$

Hence, an error occurs when E_0 occurs (i.e., the power constraint is violated), or E_{p_1} occurs (i.e., the outer layer code is wrongly decoded), or E_1^c occurs (i.e., the underlying sparse signal of the transmitted message \mathbf{X}_1 and its decoded noisy version $\tilde{\mathbf{X}}_1$ are in a distance greater than the noise level), or one of E_i occurs while $i \neq 1$. Say M is the transmitted message (in this case $M = 1$) and \hat{M} is the decoded message. Let \mathcal{E} denote the event $M \neq \hat{M}$. Hence,

$$\begin{aligned} Pr(\mathcal{E}|M=1) &= Pr(\mathcal{E}) \\ &= Pr(E_0 \cup E_{p_1} \cup E_1^c \cup \bigcup_{j=2}^{2^{mR}} E_j) \quad (24) \\ &\leq Pr(E_0) + Pr(E_{p_1}) + Pr(E_1^c) \\ &\quad + \sum_{j=2}^{2^{mR}} Pr(E_j), \end{aligned}$$

by union bounds of probabilities. We bound these probabilities term by term as follows:

- By the law of large numbers, we have $Pr(\|\mathbf{X}_1\|_{\mathcal{L}_2}^2 > \frac{m}{1+\delta_k}P + \epsilon) \rightarrow 0$ as $n \rightarrow \infty$. By using the RIP condition (3) and with probability one, we have,

$$\|\Phi_{S_m} \mathbf{X}_1\|_{\mathcal{L}_2}^2 \leq (1 + \delta_k) \|\mathbf{X}_1\|_{\mathcal{L}_2}^2 \leq mP. \quad (25)$$

Hence, $Pr(E_0) \rightarrow 0$ as $n \rightarrow \infty$.

- After using the compressive sensing decoder mentioned in (22), we have $\tilde{\mathbf{X}}_1$ such that $\frac{1}{n} \|\mathbf{X}_1 - \tilde{\mathbf{X}}_1\|_{\mathcal{L}_2}^2 \leq \frac{\beta m}{n} N$. Suppose this error is uniformly distributed over different coordinates. We use a threshold comparison to determine the sparsity pattern $\tilde{\Xi}_1$. The probability of error in determining $\tilde{\Xi}_1$ determines how high the SNR should be. Intuitively, if we use all possible outer layer codes (all possible sparsity patterns), we should not make any error in determining the sparsity pattern of each coordinate j (i.e., $\tilde{\Xi}_{1j}$). Hence, we need sufficiently high SNR for a given n . On the other hand, if we decrease the outer layer code rate, the required SNR would be lower than the case before. Here, to illustrate how the required SNR can be computed, we assume that we use all possible sparsity patterns. Say $E_{p_{j1}}$ is the event of making an error in determining whether the j^{th} coordinate is zero or not. We say $\tilde{\Xi}_{1j} = 0$ if $X_{ij} < \tau$. Otherwise, $\tilde{\Xi}_{1j} = 1$. Hence, for a given threshold τ , we have,

$$\begin{aligned} Pr(E_{p_{j1}}|\Xi_{1j}=0) &= \phi\left(\frac{\tau}{\sqrt{\frac{\beta m}{n}N}}\right) \\ Pr(E_{p_{j1}}|\Xi_{1j}=1) &= 1 - 2\phi\left(\frac{\tau}{\sqrt{P + \frac{\beta m}{n}N}}\right) \end{aligned}$$

where $\phi(\cdot)$ is the cumulative distribution function of a normal random variable. Hence,

$$\begin{aligned} Pr(E_{p_1}) &= 1 - Pr(E_{p_1}^c) \quad (26) \\ &= 1 - \left(1 - \phi\left(\frac{\tau}{\sqrt{\frac{\beta m}{n}N}}\right)\right)^{n(1-\alpha)} \\ &\quad \times \left(2\phi\left(\frac{\tau}{\sqrt{P + \frac{\beta m}{n}N}}\right)\right)^{n\alpha} \end{aligned}$$

Note that, one can choose τ and $\frac{P}{N}$ large enough to have $Pr(E_{p_1})$ arbitrarily small.

- By [2], $Pr(E_1^c)$ is zero.
- At the last step, we need to bound $Pr(E_j)$ for $j \neq 1$. We have,

$$\begin{aligned} Pr(E_j) &= Pr(\Xi_j = \Xi_1)Pr(E_j|\Xi_j = \Xi_1) \\ &\quad + Pr(\Xi_j \neq \Xi_1)Pr(E_j|\Xi_j \neq \Xi_1) \\ &= \alpha^k(1 - \alpha)^{n-k}Pr(E_j|\Xi_j = \Xi_1). \end{aligned} \quad (27)$$

Note that, $Pr(E_j|\Xi_j \neq \Xi_1)$ goes to zero in a high SNR regime. By bounding the noise power of non-zero coordinates by $\beta m N$ (the whole noise power) and using random coding argument in a k dimensional space, we

have,

$$Pr(E_j | \Xi_j = \Xi_1) \stackrel{\leq}{\leq} \frac{1}{(1 + \frac{1}{\beta(1+\delta_k)} \frac{P}{N})}, \quad (28)$$

where $\stackrel{\leq}{\leq}$ indicates the right hand side is less or equal than the left one in an exponential rate. Therefore, by (27) and (28), and doing some manipulations, we have,

$$P(E_j) \stackrel{\leq}{\leq} 2^{-n(H_b(\alpha) + \frac{\alpha}{2} \log(\frac{1}{\beta(1+\delta_k)} \frac{P}{N}))}, \quad (29)$$

for $j = 2, \dots, 2^m R$.

Thus, in a sufficiently high SNR regime, by (24) and (29), we have,

$$\begin{aligned} P_e^m = Pr(\mathcal{E}) &= Pr(\mathcal{E} | M = 1) \\ &\stackrel{\leq}{\leq} 2^m R 2^{-n(H_b(\alpha) + \frac{\alpha}{2} \log(\frac{1}{\beta(1+\delta_k)} \frac{P}{N}))} \\ &\stackrel{\leq}{\leq} 2^m (R - \frac{n}{m} H_b(\alpha) - \frac{k}{m} C - \frac{k}{2m} \log(\frac{1}{\beta(1+\delta_k)} \frac{P}{N})). \end{aligned} \quad (30)$$

For any given $\epsilon > 0$, we can choose n large enough to have $R \leq \frac{k}{m} C + \frac{n}{m} H_b(\alpha) + \frac{k}{2m} \log(\frac{1}{\beta(1+\delta_k)} \frac{P}{N}) - \epsilon$ achievable. ■

V. CONCLUSIONS

In this paper, we demonstrated some applications of compressive sensing over networks. We made a connection between compressive sensing and traditional information theoretic techniques in source coding and channel coding in order to provide mechanisms to explicitly trade-off between the decoding complexity and the rate. Although optimal decoders to recover the original signal compressed by source coding have high complexity, the compressive sensing decoder is a linear or convex optimization. First, we investigated applications of compressive sensing on distributed compression of correlated sources. For a depth one tree network, references ([13]-[18]) provide some low complexity decoding techniques for Slepian-Wolf compression. However, none of these methods can explicitly provide a trade-off between the rate and the decoding complexity. For a general multicast network, reference [11] extended Slepian-Wolf compression for a multicast problem. It is showed in [19] that, there is no separation between Slepian-Wolf compression and network coding. For this network, a low complexity decoder for an optimal compression scheme has not been proposed yet. Here, by using compressive sensing, we proposed a compression scheme for a family of correlated sources with a modularized decoder providing a trade-off between the compression rate and the decoding complexity. We called this scheme *Sparse Distributed Compression*. We used this compression scheme for a general multicast network with correlated sources. Here, we first decoded some of the sources by a network decoding technique and then, we used a compressive sensing decoder to obtain the whole sources.

Next, we investigated applications of compressive sensing on channel coding. We proposed a coding scheme that combines compressive sensing and random channel coding for a high-SNR point-to-point Gaussian channel. We called

this scheme *Sparse Channel Coding*. Our coding scheme provides a modularized decoder to have a trade-off between the capacity loss and the decoding complexity. The idea is to add intentionally some correlation to transmitted signals in order to decrease the decoding complexity. At the decoder side, first we used a compressive sensing decoder to get an estimate of the original sparse signal, and then we used a channel coding decoder in the subspace of non-zero coordinates.

REFERENCES

- [1] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, p. 969, 2007.
- [3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [4] M. Akçakaya and V. Tarokh, "Shannon theoretic limits on noisy compressive sampling," *arXiv*, vol. 711.
- [5] S. Sarvotham, D. Baron, and R. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, 2006.
- [6] Y. Jin, Y.-H. Kim, and B. D. Rao, "Performance tradeoffs for exact support recovery of sparse signals," in *2010 IEEE International Symposium on Information Theory, ISIT 2010*, TX, US, 2010, pp. 1558–1562.
- [7] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [8] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [9] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.
- [10] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [11] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [12] I. Csiszar and J. Körner, in *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.
- [13] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," in *dcc*. Published by the IEEE Computer Society, 1999, p. 158.
- [14] T. Coleman, A. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Transactions on Information Theory*, vol. 52, no. 8, 2006.
- [15] R. Gallager, "Low-density parity-check codes," *Information Theory, IRE Transactions on*, vol. 8, no. 1, pp. 21–28, 1962.
- [16] T. Tian, J. Garcia-Frias, and W. Zhong, "Compression of correlated sources using LDPC codes," in *Data Compression Conference, 2003. Proceedings. DCC 2003*, 2003.
- [17] C. Berrou, A. Glavieux, P. Thitimajshima *et al.*, "Near Shannon limit error-correcting coding and decoding," in *Proceedings of ICC*, vol. 93, 1993, pp. 1064–1070.
- [18] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Communications Letters*, vol. 5, no. 10, pp. 417–419, 2001.
- [19] A. Ramamoorthy, K. Jain, P. Chou, and M. Effros, "Separating distributed source coding from network coding," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2785–2795, 2006.
- [20] G. Maierbacher, J. Barros, and M. Médard, "Practical source-network decoding," in *Wireless Communication Systems, 2009. ISWCS 2009. 6th International Symposium on*. IEEE, 2009, pp. 283–287.
- [21] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1204–1216, 2000.
- [22] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. on networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [23] G. Forney, *Concatenated codes*. Citeseer, 1966.