

Compress-and-Forward Scheme for a Relay Network: Approximate Optimality and Connection to Algebraic Flows

Adnan Raja, and Pramod Viswanath

January 11, 2011

Abstract

We study a wireless relay network, with a single source and a single destination. Our main result is to show that an appropriate *compress-and-forward* scheme supports essentially the same reliable data rate as the quantize-map-and-forward and noisy network coding schemes [1, 2]; thus, it is approximately optimal – in the sense the data rate is a universal constant away from the cut-set upper bound. We characterize the compress-and-forward scheme through an abstract flow formulation, a generalization of flow on linking systems. This characterization allows for *efficient* computation of the *minimal* amount of information that has to flow through each node in the network. We also generalize these results to the network with multiple sources with independent messages intended for a single destination; and to the broadcast network with a single source intending to communicate independent messages to multiple destinations.

1 Introduction

The focus of this paper is a unicast wireless network: a single source node communicates reliably with a single destination node with the assistance of many relay nodes. The wireless network is modeled by a linear channel capturing *superposition* and *broadcast* and with additive Gaussian noise. This is known as the Gaussian relay network in the literature. Recently, the approximate capacity of the unicast Gaussian relay network was characterized in [1], where a quantize-map-and-forward scheme was proposed. It was shown that this scheme achieves within a constant gap of the cut-set bound, where the constant gap depends only on the size of the network and not on the channel parameters. In this scheme, each node quantizes the received signal, symbol by symbol, at the noise level and maps it to a random Gaussian codeword for forwarding. The scheme can also be modified to perform a vector quantization operation at each relay node [2, 3]. In particular, [2] gives an achievable scheme, called *noisy network coding*, for a general discrete memoryless network. In this scheme, the relay quantizes the received signal in blocks using vector-quantization; subsequently mapping each quantized codeword to a unique Gaussian codeword, which is re-transmitted by the relay.

A related scheme for the relay network is the compress-and-forward scheme which was first proposed for the simple one-relay channel in [4] and extended to relay networks in [5]. In compress-and-forward, the relay node *bins* the quantized received signal and subsequently maps the bin number to a unique Gaussian codeword, which is then retransmitted by the relay. In [4, 5], unlike the noisy network coding scheme, the decoding process required the destination to also decode the bin indices at the relays. Therefore, as shown in [2], this scheme is strictly sub-optimal as compared to noisy network coding.

In this paper, we propose a compress-and-forward scheme, where the destination node does not explicitly decode the bin indices as per [2]. We only need the destination to identify a unique source codeword that is consistent with *some* bin index at each relay. Our main result is to show that this scheme achieves the *same* data rate as the noisy network coding scheme of [2]. Further, we calculate the *minimal* compression rates at each relay node in a certain sense. This is done by introducing an abstract flow formulation over the communication network. This is a generalization of the concept of linking systems and flows introduced in [7, 8] in the context of the linear deterministic network. The flow value at each node can be thought of as the amount of information flowing through that node. We show that a compress-and-forward scheme can be constructed with compression rate at each node close to the flow value at that node. Further, the flow, and hence the approximately minimal compression rates, at each node can be computed *efficiently*.

The linear deterministic network was introduced in [1] as a model that captures many features of the wireless network. Random coding argument was used to show the existence of schemes that achieve capacity of the linear deterministic network [1, 2]. On the other hand, in [6, 7, 8], simpler and efficiently computable schemes have been proposed for this network. In particular, in [7, 8], the concept of flow was introduced for the linear deterministic network. The flow value at each node corresponded to the number of independent equations, that particular node needed to forward. This forwarding could be done with a simple permutation matrix at each node mapping the received vector to the transmit vector. Both the flow value at the node and the permutation mapping could be constructed in polynomial time. Our result can be viewed as the analog of the results of [6, 7, 8] in the context of the Gaussian network; see Figure 1.

The rest of the paper is organized as follows. In Section 2 we present our compress-and-forward scheme for the relay network and characterize the achievable rate as a function of the compression or binning rates at the relay nodes. In Section 3 we generalize the notion of flows and cuts with respect to arbitrary family of submodular functions defined over the network. A *max-flow min-cut* theorem is then proved. When specialized to matrices and the rank function, this theorem yields the main result of [7]. In Section 4, we use the notion of flows to characterize a particular compress-and-forward scheme for the relay network, which is *minimal* in terms of the compression rates at the relay nodes. In Section 5 we discuss the ramifications of our algebraic flow formulation to the important special cases of the Gaussian wireless relay network and the deterministic relay network. In Section 6, we generalize our results to the network with multiple sources with independent messages intended for a single destination; and to the broadcast network with a single source intending to communicate independent messages to multiple destinations.

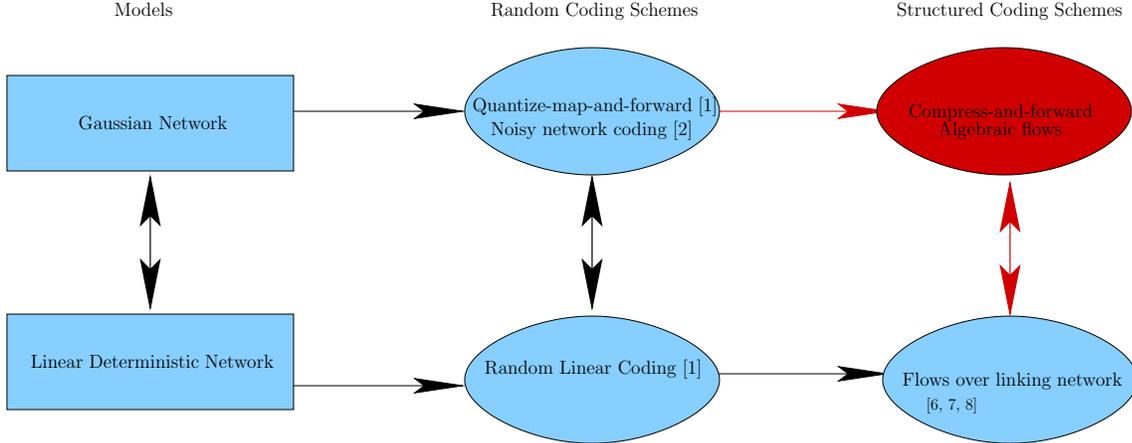


Figure 1: A depiction of the communication schemes on the Gaussian and linear deterministic networks. The main result of this paper is represented by the upper-right bubble in red.

2 Unicast Relay Network

Consider a communication network with a set of nodes \mathcal{V} . Each node in the network abstracts a *radio*, which can both transmit and receive (in full or half duplex modes). The traffic is *unicast*: a single source node is communicating reliably to a single destination node using the other nodes in the network as relays. We will be interested in a single-source single-destination relay network, which has a unique source node s and destination node d and the other nodes function as relay nodes. At any node v , the transmit alphabet is given by \mathcal{X}_v and the receive alphabet by \mathcal{Y}_v (supposed to be discrete sets, for the most part). Time is discrete and synchronized among all nodes. The transmit symbol at any time at a node v is given by $x_v \in \mathcal{X}_v$ and the receive symbol is given by $y_v \in \mathcal{Y}_v$. We will consider a *memoryless* network wherein the received symbol at any node at any given time depends (in a random fashion) only on the current transmitted symbols at other nodes.

A $(2^{TR}, T)$ coding scheme for the relay network, which communicates over T time instants, comprises of the following.

1. The *message* W , which is modeled as an independent random variable distributed uniformly on $[2^{TR}]$. W is known at the source node and is intended for the destination node.

2. The *source mapping* for each time $t \in [T]$,

$$f_{s,t} : (W \times \mathcal{Y}_s^{t-1}) \rightarrow \mathcal{X}_s. \quad (1)$$

3. The *relay mappings* for each $v \in \mathcal{V} \setminus \{s\}$ and $t \in [T]$,

$$f_{v,t} : \mathcal{Y}_v^{t-1} \rightarrow \mathcal{X}_v. \quad (2)$$

4. The *decoding map* at destination d ,

$$g_d : \mathcal{Y}_d^T \rightarrow \hat{W}. \quad (3)$$

The probability of error for destination d under this coding scheme is given by

$$P_e \stackrel{\text{def}}{=} \Pr\{\hat{W} \neq W\}. \quad (4)$$

A rate R (in bits per unit time) is said to be achievable if for any $\epsilon > 0$, there exists a $(2^{TR}, T)$ scheme that achieves a probability of error lesser than ϵ for all nodes, i.e., $P_e \leq \epsilon$. The capacity of the network is the supremum of all achievable rates.

It was shown in [1] that any arbitrary communication network can be converted into a layered network by coding over blocks of time. Each layer then captures the operations in the corresponding block of time. Further, if the nodes have half-duplex constraint, then this time-layering is done with a fixed transmit-receive schedule, which says which nodes are transmitting and which ones are listening in any block of time. It is then a secondary question to optimize over the schedule in order to get the maximum rate of transmission.

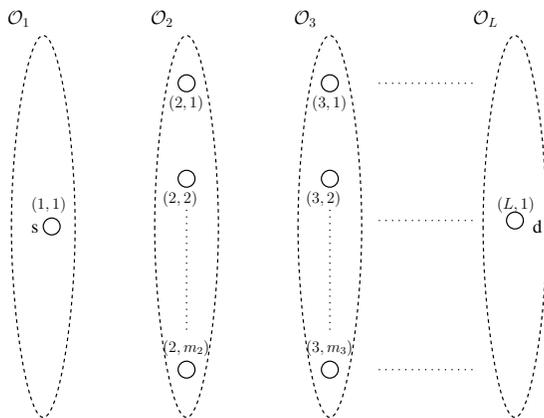


Figure 2: A layered network.

Henceforth, we will focus on an L -layered network as shown in Figure 2, so that

$$\mathcal{V} = \bigcup_{l=1}^L \mathcal{O}_l, \quad (5)$$

where \mathcal{O}_l denotes the m_l nodes in the l -th layer. The k -th node in the l -th layer will be denoted by the ordered pair (l, k) . The first layer has only one node which is the source node and is denoted by $(1, 1)$ or s . The last layer has only the destination node and is denoted by $(L, 1)$ or d . The nodes other than the source and the destination node will be referred to as the relay nodes and are denoted by \mathcal{V}_r .

In the layered network, the received symbol for a node in the $l + 1$ -th layer depends only on the transmit symbol from the nodes in the l -th layer. This dependency is given by the probability transition function

$$p(y_{\mathcal{O}_{l+1}} | x_{\mathcal{O}_l}) = \prod_{k=1}^{m_{l+1}} p(y_{(l+1,k)} | x_{\mathcal{O}_l}). \quad (6)$$

Here $x_{\mathcal{O}_l}$ is used to denote $\{x_v : v \in \mathcal{O}_l\}$. $y_{\mathcal{O}_l}$'s are similarly defined. This models the *communication channel* for the network.

In particular, if the probability transition function is a deterministic function, then the network is called a *deterministic network*. Further, if the transmit and received alphabets are modeled as vectors over finite fields and the deterministic function is modeled as a linear function, then the network is called a *linear deterministic network*. If the alphabet sets are complex and the probability transition function linear with an additive complex Gaussian noise, then the network is called a *Gaussian network*. The Gaussian network is of much practical interest and in [1] it was shown that the linear deterministic network captures many features of the Gaussian network.

2.1 Compress-and-Forward Scheme

In this section, we will describe in detail a compress-and-forward scheme, which involves quantization followed by a *binning* step before re-transmission at each relay node.

Consider the collection of independent random variables $\{X_{\mathcal{V}\setminus d}\} \sim \prod_{v \in \mathcal{V}\setminus d} p(x_v)$. The channel induces random variables $Y_{\mathcal{V}\setminus s}$. We also define random variables $\hat{Y}_{\mathcal{V}} \sim \prod_{v \in \mathcal{V}} p(\hat{y}_v|y_v)$. The scheme comprises of the following.

1. *Source codebook and encoding:* For each message $w \in [2^{TR}]$, the source generates a T -length sequence $x_s^T(w)$ using i.i.d. $p(x_s)$. Each message is assumed to be equally likely.
2. *Relay codebooks and mappings:* For every relay node $(l, k) \in \mathcal{V}_r$ a binned quantization codebook is generated with $2^{TR_{(l,k)}}$ bins. The binned quantization codebook is given by $\hat{y}_{(l,k)}^T(w_{(l,k)}, \bar{w}_{(l,k)})$, where $1 \leq w_{(l,k)} \leq 2^{TR_{(l,k)}}$ and $1 \leq \bar{w}_{(l,k)} \leq 2^{T\bar{R}_{(l,k)}}$. And it is generated using i.i.d. $p(\hat{y}_{(l,k)})$.

Every relay node also generates a transmission codebook of size $2^{TR_{(l,k)}}$, which consists of $x_{(l,k)}^T(w_{(l,k)})$ sequences generated using i.i.d. $p(x_{(l,k)})$.

On receiving $y_{(l,k)}^T$, the relay node finds a vector $\hat{y}_{(l,k)}^T(w_{(l,k)}, \bar{w}_{(l,k)})$ in the quantization codebook and transmits $x_{(l,k)}^T(w_{(l,k)})$ corresponding to the bin number of the quantization vector.

If the relay cannot find any quantization vector, it transmits a sequence corresponding to any bin uniformly at random. In order to ensure that the latter event has diminishing probability, we fix the total size of the quantization codebook to be of the order $2^{TI(Y_{(l,k)}, \hat{Y}_{(l,k)})}$. This fixes $\bar{R}_{(l,k)}$ such that,

$$\bar{R}_{(l,k)} = I(Y_{(l,k)}, \hat{Y}_{(l,k)}) - R_{(l,k)} + \epsilon_1, \quad (7)$$

for an arbitrary $\epsilon_1 > 0$.

3. *Decoding:*

On receiving y_d^T , the destination node finds a *unique* \hat{w} , and *any* $\hat{w}_{(l,k)}, \hat{\hat{w}}_{(l,k)}$, such that

$$\left(x_s^T(\hat{w}), \{ \hat{y}_{(l,k)}^T(\hat{w}_{(l,k)}, \hat{\hat{w}}_{(l,k)}), x_{(l,k)}^T(\hat{w}_{(l,k)}) \}_{(l,k) \in \mathcal{V}_r}, y_d^T \right) \in \mathcal{T}_\epsilon^T. \quad (8)$$

If it is successful, the destination declares \hat{w} as the decoded message; if not, the destination declares an error. The flexibility of allowing *any* $\hat{w}_{(l,k)}, \hat{\hat{w}}_{(l,k)}$, that is consistent with a unique \hat{w} such that the condition in Equation (8) is satisfied is the critical difference from the earlier works on compress-and-forward [4, 5].

2.1.1 Error Probability Analysis:

The compress-and-forward scheme is essentially parameterized by R , which is the rate of the source codebook and $R_{(l,k)}$'s, which is the binning rate at relay node (l, k) . The following theorem results from a careful analysis of the probability of error.

Theorem 1. *The compress-and-forward scheme is reliable with rate R and $R_{(l,k)}$'s if,*

$$R < R(\Omega^c \setminus \Phi) + I(\hat{Y}_\Phi; X_\Omega | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_\mathcal{V}), \quad (9)$$

$\forall \Omega \subseteq \mathcal{V}$ with $s \in \Omega$, and $\forall \Phi \subseteq \Omega^c$ with $d \in \Phi$, for some $\prod_{v \in \mathcal{V}} p(x_v)p(\hat{y}_v|y_v)$. Here $R(\mathcal{A}) \stackrel{\text{def}}{=} \sum_{v \in \mathcal{A}} R_v$. For notational convenience, we have assumed that $\hat{Y}_d = Y_d$, $X_d = \emptyset$, and $Y_s = \hat{Y}_s = \emptyset$.

Proof. See Appendix B. □

We first observe the following corollary of the above theorem which characterizes the overall rate that can be achieved by our compress-and-forward scheme

Corollary 1. *There exists a reliable compress-and-forward scheme achieving overall rate R if,*

$$R < \min_{\Omega \subseteq \mathcal{V}, s \in \Omega} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - I(\hat{Y}_\Omega; Y_\Omega | X_\mathcal{V}), \quad (10)$$

for some $\prod_{v \in \mathcal{V}} p(x_v)p(\hat{y}_v|y_v)$.

Proof. The compress-and-forward scheme with $R_{(l,k)} = I(Y_{(l,k)}, \hat{Y}_{(l,k)}) + \epsilon_1$ achieves this rate. □

Note that the above scheme is essentially the quantize-and-forward scheme, such that every quantized codeword is uniquely mapped to a re-transmission codeword at the relay node. In this paper, we aim to provide a compress-and-forward scheme with *minimal* binning rates at each node. The next corollary of Theorem 1 characterizes a lower bound on the binning rate across any layer.

Corollary 2. *Any compress-and-forward scheme achieving overall rate R must satisfy*

$$R(\mathcal{O}_l) > R. \quad (11)$$

Proof. To see this, note that letting $\Omega = \mathcal{O}_1 \cup \dots \cup \mathcal{O}_{l-1}$, and $\Phi = \mathcal{O}_{l+1} \cup \dots \cup \mathcal{O}_L$ in (9) gives

$$\begin{aligned} R(\mathcal{O}_l) &> R + I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}}) \\ &\geq R. \end{aligned}$$

□

If $R_{(l,k)}$ can be thought of as the rate of information flowing through the node (l, k) , then $R(\mathcal{O}_l)$ is the rate of total information flowing through the l -th layer and hence must be at least R .

We will see in the next two sections as to how the concept of a flow can be used to construct a compress-and-forward scheme with binning rates which closely meets the above lower bound.

3 Algebraic Flow

In this section, we generalize the concept of flow as described in [7] to an abstract network model, a special case of which is the discrete memoryless communication network. Our main result is a *max-flow min-cut theorem* for this abstract network model. This result is then used to shed insight into the compression rates at the relay nodes of the discrete memoryless communication network of interest.

Consider a network \mathcal{N} represented by the set of nodes \mathcal{V} . Further, we will consider an L -layered network, so that

$$\mathcal{V} = \bigcup_{l=1}^L \mathcal{O}_l, \quad (12)$$

where each layer has m_l nodes, i.e. $|\mathcal{O}_l| = m_l$. The k -th node in the l -th layer will be denoted by the ordered pair (l, k) .

Definition 1. A family of $L - 1$ functions $\{\rho_l : 1 \leq l \leq L - 1\}$, $\rho_l : 2^{\mathcal{O}_l} \times 2^{\mathcal{O}_{l+1}} \rightarrow \mathbb{R}^+$ are called channel functions for the layered network, if they satisfy the following properties:

1. ρ_l is bi-submodular, i.e., $\forall U_1, U_2 \subseteq \mathcal{O}_l, W_1, W_2 \subseteq \mathcal{O}_{l+1}$,

$$\rho_l(U_1 \cup U_2, W_1 \cap W_2) + \rho_l(U_1 \cap U_2, W_1 \cup W_2) \leq \rho_l(U_1, W_1) + \rho_l(U_2, W_2). \quad (13)$$

2. ρ_l is non-decreasing, i.e.

$$\rho_l(U, W) \leq \rho_l(U_1, W_1), \text{ for } U \cup W \subseteq U_1 \cup W_1. \quad (14)$$

3. If $U = \emptyset$ or $W = \emptyset$ then,

$$\rho_l(U, W) = 0. \quad (15)$$

These functions abstract a “channel” between two adjacent layers in the network. In the context of the discrete memoryless network, the natural candidate for the channel function is the *mutual information* $I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U})$. This is *indeed* a channel function, as shown below.

Proposition 1. Given a collection of independent random variables $\{X_{\mathcal{V} \setminus \mathcal{O}_L}\} \sim \prod_{v \in \mathcal{V} \setminus d} p(x_v)$ and the discrete memoryless network defined by (6), which induces random variables $Y_{\mathcal{V} \setminus \mathcal{O}_1}$,

$$\rho_l(U, W) = I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U}), \quad (16)$$

forms a family of channel functions.

Proof. See Appendix C. □

Next we define *flow* for the network with respect to the channel functions $\{\rho_l\}$.

Definition 2. A *flow* on a network is defined as a function $d : \mathcal{V} \rightarrow \mathbb{R}^+$ which satisfies the following two properties:

1. $\forall U \subseteq \mathcal{O}_l$, and $\forall W \subseteq \mathcal{O}_{l+1}$,

$$d(W) - d(\mathcal{O}_l \setminus U) \leq \rho_l(U, W). \quad (17)$$

2. Defining $d(\mathcal{A}) \stackrel{\text{def}}{=} \sum_{(l,k) \in \mathcal{A}} d(l, k)$,

$$d(\mathcal{O}_1) = d(\mathcal{O}_L). \quad (18)$$

The value of the flow is said to be $d(\mathcal{O}_1)$.

Traditionally, (i.e., in wireline networks) flow is defined over edges. Here, however, there are no natural equivalents of edges. As such, our flow is defined over nodes and restricted to satisfy certain conditions imposed by the channel functions. The quantity $d(l, k)$ abstracts the amount of “information” flowing through the node (l, k) . The first condition is essentially a capacity constraint resulting from the channel functions of the network. It states that the amount of information flowing from the nodes U to W is upper bounded by the channel function $\rho_l(U, W)$. The second condition states that the information in final layer must be the same as the final layer. This ensures that the first layer functions as the source and the last layer as the destination.

Finally, we define a cut for the network with the channel functions $\{\rho_l\}$.

Definition 3. For any $\Omega \subseteq \mathcal{V}$, the *cut* $C(\Omega) : 2^{\mathcal{V}} \rightarrow \mathbb{R}_+$ is defined as

$$C(\Omega) \stackrel{\text{def}}{=} \sum_{l=1}^{L-1} \rho_l(\Omega_l, \mathcal{O}_{l+1} \setminus \Omega_{l+1}), \quad (19)$$

where $\Omega_l \stackrel{\text{def}}{=} \Omega \cap \mathcal{O}_l$.

The main result of this section is to show that the maximum flow is equal to the minimum cut.

Theorem 2. *There exists a flow for the network with value f and with given flow values $\{d(1, k) : 1 \leq k \leq m_1\}$ for the first layer and $\{d(L, k) : 1 \leq k \leq m_L\}$ the last layer, if and only if*

$$f \leq \min_{\Omega \in \mathcal{V}} \{C(\Omega) + d(\mathcal{O}_1 \setminus \Omega_1) + d(\Omega_L)\}. \quad (20)$$

Further, this flow can be computed in polynomial time.

Proof. Our proof closely follows the proof of Theorem 14 in [7]. We prove by induction. For $L=2$, the theorem holds by definition. Consider $L > 2$. The induction hypothesis assumes that the theorem holds true for any network with less than L layers. Consider any $L_0 \in \{2, \dots, L-1\}$. Define networks \mathcal{N}_A and \mathcal{N}_B to be the sub-networks of \mathcal{N} with the set of vertices $\mathcal{V}_A = \cup_{l=1}^{L_0} \mathcal{O}_l$ and $\mathcal{V}_B = \cup_{l=L_0}^L \mathcal{O}_l$ respectively. Similarly, denote the cut for the two networks by C_A and C_B respectively.

By the induction hypothesis, there exists an f -flow for the network, if and only if there exists $\{d(L_0, k) : 1 \leq k \leq m_{L_0}\}$, such that

$$d(\mathcal{O}_{L_0}) = f, \quad (21)$$

$$d(\Omega_A \cap \mathcal{O}_1) - d(\Omega_A \cap \mathcal{O}_{L_0}) \leq C_A(\Omega_A), \quad \forall \Omega_A \subseteq \mathcal{V}_A, \text{ and} \quad (22)$$

$$d(\Omega_B \cap \mathcal{O}_{L_0}) - d(\Omega_B \cap \mathcal{O}_L) \leq C_B(\Omega_B), \quad \forall \Omega_B \subseteq \mathcal{V}_B. \quad (23)$$

The set of linear inequalities given by (22) and (23) can be rewritten as

$$d(T) \leq r_A(T) \stackrel{\text{def}}{=} \min \{C_A(\Omega_A) + d(\Omega_A^c \cap \mathcal{O}_1) : \Omega_A^c \cap \mathcal{O}_{L_0} = T\}, \quad (24)$$

$$d(T) \leq r_B(T) \stackrel{\text{def}}{=} \min \{C_B(\Omega_B) + d(\Omega_B \cap \mathcal{O}_L) : \Omega_B \cap \mathcal{O}_{L_0} = T\}, \quad (25)$$

$\forall T \subseteq \mathcal{O}_{L_0}$. We now have the following observation:

Lemma 1. *The functions $r_A(T)$ and $r_B(T)$ are*

- *submodular,*
- *non-decreasing, and*
- *satisfy $r_A(\emptyset) = 0$ and $r_B(\emptyset) = 0$.*

Therefore, we can associate the following polymatroids with the functions r_A and r_B .

$$P_A = \{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_A(U), \quad \forall U \in \mathcal{O}_{L_0}\} \quad (26)$$

$$P_B = \{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_B(U), \quad \forall U \in \mathcal{O}_{L_0}\}, \quad (27)$$

where $\mathbf{x} = [x(1) \dots x(m_{L_0})]$ and $x(U) \stackrel{\text{def}}{=} \sum_{u \in U} x(u)$. The conditions (21)-(23) are now equivalent to finding

$$[d(L_0, 1) \dots d(L_0, m_{L_0})] \in P_A \cap P_B, \quad (28)$$

such that $d(\mathcal{O}_{L_0}) = f$. It then follows from Edmond's polymatroid intersection ([9], Corollary 46.1b) that:

$$\max \{x(\mathcal{O}_{L_0}) : \mathbf{x} \in P_A \cap P_B\} = \min_{T \subseteq \mathcal{O}_{L_0}} \{r_A(\mathcal{O}_{L_0} \setminus T) + r_B(T)\}. \quad (29)$$

Therefore the required f -flow exists if and only if,

$$f \leq \min_{T \subseteq \mathcal{O}_{L_0}} \{r_A(\mathcal{O}_{L_0} \setminus T) + r_B(T)\} \quad (30)$$

$$= \min_{\Omega \in \mathcal{V}} \{C(\Omega) + d(\mathcal{O}_1 \setminus \Omega_1) + d(\Omega_L)\}. \quad (31)$$

Further, in Theorem 47.1 of [9] it is shown that the maximizing \mathbf{x} in (29) can be computed in *polynomial* time. Hence, the flow can also be computed in polynomial time. \square

Specializing Theorem 2 to the network with only one node each in the first and the last layer gives the following corollary.

Corollary 3. *If the network is such that both the first and the last layers have one node each, then there exists a flow for the network with value f if and only if*

$$f \leq \min_{\Omega: \mathcal{O}_1 \subseteq \Omega, \mathcal{O}_L \subseteq \Omega^c} C(\Omega). \quad (32)$$

We next use this corollary to construct a specific compress-and-forward scheme for the layered unicast Gaussian relay network.

4 A Compress-and-Forward Scheme From Flows

We now use our results on flows in Section 3 to find a feasible compress-and-forward scheme for the relay network. For the layered communication network and the choice of coding parameters (the random variables X 's and \hat{Y} 's) on it, as defined earlier in this section, the mutual information forms a natural family of channel functions given by

$$\rho_l(U, W) = I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U}), \quad (33)$$

for $U \subseteq \mathcal{O}_l$ and $W \subseteq \mathcal{O}_{l+1}$. From Proposition 1, we know that the functions ρ_l so defined, are channel functions (i.e., they satisfy the three properties of Definition 1). For any $\Omega \subseteq \mathcal{V}$, the corresponding cut value $C(\Omega)$ is now given by,

$$C(\Omega) = \sum_{l=1}^{L-1} I(X_{\Omega_l}; \hat{Y}_{\mathcal{O}_{l+1} \setminus \Omega_{l+1}} | X_{\mathcal{O}_l \setminus \Omega_l}) \quad (34)$$

$$= I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^e}). \quad (35)$$

The following corollary results from specializing Corollary 3 to the communication and channel functions defined above.

Corollary 4. *If*

$$f < \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}), \quad (36)$$

then there exists a flow $\{d_{(l,k)}\}$ in the network with value f .

We will use this flow to construct a compress-and-forward scheme for the relay network.

Theorem 3. *If*

$$R < \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} \left\{ I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Omega}; Y_{\Omega} | X_{\mathcal{V}}) \right\}, \quad (37)$$

for some $\prod_{v \in \mathcal{V}} p(x_v) p(\hat{y}_v | y_v)$, then there exists a compress-and-forward scheme for the relay network with rate R and binning rates given by,

$$R_{(l,k)} = d_{(l,k)} + I(\hat{Y}_{(l,k)}; Y_{(l,k)} | X_{\mathcal{V}_r}, X_s), \quad (38)$$

where $\{d_{(l,k)}\}$ is a flow on the network with value $f = R + \delta$, where

$$\delta = \max_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} I(\hat{Y}_{\Omega}; Y_{\Omega} | X_{\mathcal{V}}). \quad (39)$$

Proof. Note that

$$f = R + \delta \quad (40)$$

$$< \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}). \quad (41)$$

It follows from Corollary 4 that there exists a flow $\{d_{(l,k)}\}$ with value f . Now, $\forall \Omega \subset \mathcal{V}$ with $s \in \Omega$, and $\forall \Phi \subseteq \Omega^c$ with $d \in \Phi$, using the property of flows for the $L - 1$ layers.

$$d(\Phi_2) \leq I(\hat{Y}_{\Phi_2}; X_s), \quad (42)$$

$$d(\Phi_{l+1}) - d(\mathcal{O}_l \setminus \Omega_l) \leq I(\hat{Y}_{\Phi_{l+1}}; X_{\Omega_l} | X_{\mathcal{O}_l \setminus \Omega_l}), \quad \text{for } 1 < l < L - 1, \quad (43)$$

$$f - d(\mathcal{O}_{L-1} \setminus \Omega_{L-1}) \leq I(Y_d; X_{\Omega_{L-1}} | X_{\mathcal{O}_{L-1} \setminus \Omega_{L-1}}). \quad (44)$$

Adding together the $L - 1$ inequalities above gives

$$f - d(\Omega^c \setminus \Phi) \leq I(\hat{Y}_{\Phi}; X_{\Omega} | X_{\Omega^c}). \quad (45)$$

And therefore,

$$R \leq R(\Omega^c \setminus \Phi) + I(\hat{Y}_{\Phi}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}}). \quad (46)$$

□

The next corollary shows that the above scheme is an approximately minimal scheme in terms of the binning rates

Corollary 5. *For the compress-and-forward scheme of Theorem 3,*

$$R(\mathcal{O}_l) = R + I(\hat{Y}_{\mathcal{O}_l}; Y_{\mathcal{O}_l} | X_{\mathcal{V}}) + \delta. \quad (47)$$

Corollary 2 shows that $R(\mathcal{O}_l)$ is at least R for any compress-and-forward scheme. For the compress-and-forward constructed above from flows, the extra term is $I(\hat{Y}_{\mathcal{O}_l}; Y_{\mathcal{O}_l} | X_{\mathcal{V}}) + \delta$. We will show in the subsequent section that this gap is zero for deterministic network and can be bounded by a constant term which is independent of the channel parameters for the Gaussian network.

5 Important Special Cases

5.1 Deterministic Network

An important special case is the deterministic network. The received symbol at any node is a deterministic function of the transmitted symbols at the other nodes. Hence the relationship between the receive and transmit random variables is given by:

$$Y_{\mathcal{O}_{l+1}} = g_l(X_{\mathcal{O}_l}). \quad (48)$$

This network, in general, captures the effect of interference and broadcast, which are the key features in wireless networks.

For this special case, we can see that the optimal choice for the quantization random variables $\hat{Y}_{(l,k)} = Y_{(l,k)}$. The channel functions are now given by

$$\rho_l(U, W) = H(Y_W | X_{U^c}), \quad (49)$$

for $U \subseteq \mathcal{O}_l$, and $W \subseteq \mathcal{O}_{l+1}$. We now have the following corollary to Theorem 3:

Corollary 6. *If*

$$R < \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} \{H(Y_{\Omega^c} | X_{\Omega^c})\}, \quad (50)$$

for some $\prod_{v \in \mathcal{V}} p(x_v)$, then there exists a compress-and-forward scheme for the relay network with rate R and binning rates given by,

$$R_{(l,k)} = d_{(l,k)}, \quad (51)$$

where $\{d_{(l,k)}\}$ is a flow on the network with value R . Further for this scheme

$$R(\mathcal{O}_l) = R. \quad (52)$$

A further special case of the deterministic network is the linear deterministic function of [1] where the alphabet sets are vectors over finite field. The relationship between the transmit and receive symbols are given by:

$$Y_{\mathcal{O}_{l+1}} = G_l X_{\mathcal{O}_l}. \quad (53)$$

For this network, it can be shown that the optimal choice of X_v 's are i.i.d. uniform random variables. Specializing our results of Section 3 gives back the results of [7, 8], as mentioned earlier. Compared to our compress-and-forward scheme, a simple transmission scheme is given for this network in [7, 8]. This scheme is over a single time symbol. The operations at each node are linear operations captured by coding matrices at each node. Further, the coding matrices at the source and the relay nodes are simple permutation matrices and can be computed in polynomial time. However, their technique is limited to the linear deterministic network.

One of the main contributions of this paper has been to extend the idea of algebraic flows from [7, 8] to general communication networks and to interpret the flow values as the rate of the transmission codebook at each node. This interpretation is exact in case of deterministic network (as seen from Corollary 6), and is only approximate in general, with a certain gap (as seen from Theorem 3).

5.2 Gaussian Network

We will only consider the case with single antenna at each node, for simplicity. Each node has unit transmit power constraint. The relationship between the transmit and receive random variables are given by

$$Y_v = \sum_{u \in \mathcal{O}_l} h_{v,u} X_u + Z_v, \quad (54)$$

where $v \in \mathcal{O}_{l+1}$ and Z_v 's are i.i.d. Gaussian with unit variance.

Now, the following corollary specializes Theorem 3 to the Gaussian network.

Corollary 7. *If*

$$R < \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, d \in \Omega^c} \{\log |1 + H_{\Omega, \Omega^c}| - |\Omega|\}, \quad (55)$$

where H_{Ω, Ω^c} is the effective channel matrix between $\{s, \Omega\}$ and $\{d, \Omega^c\}$; then there exists a compress-and-forward scheme for the relay network with rate R and binning rates given by,

$$R_{(l,k)} = d_{(l,k)} + 1, \quad (56)$$

where $\{d_{(l,k)}\}$ is a flow on the network with value $f = R + |\mathcal{V}_r|$. Further for this scheme

$$R(\mathcal{O}_l) = R + |\mathcal{V}_r| + |\mathcal{O}_l|. \quad (57)$$

Proof. We let X_v 's to be i.i.d. Gaussian. Further, let $\hat{Y}_v = Y_v + \hat{Z}_v$, where \hat{Z}_v 's are also i.i.d. Gaussian with unit variance. The term $I(\hat{Y}_{(l,k)}; Y_{(l,k)} | X_{\mathcal{V}_r}, X_s)$ can now be evaluated as follows

$$I(\hat{Y}_{(l,k)}; Y_{(l,k)} | X_{\mathcal{V}_r}, X_s) = h(\hat{Y}_{(l,k)} | X_{\mathcal{V}_r}, X_s) - h(\hat{Y}_{(l,k)} | Y_{(l,k)}, X_{\mathcal{V}_r}, X_s) \quad (58)$$

$$= h(Z_{(l,k)} + \hat{Z}_{(l,k)}) - h(\hat{Z}_{(l,k)}) \quad (59)$$

$$= \log_2 2 \quad (60)$$

$$= 1 \text{ bit.} \quad (61)$$

□

Note that it has already been shown in [1, 2] that the above achievable rate is within a constant gap of the cut-set bound, where the gap again depends only on the number of nodes in the network.

6 Multi-source and Broadcast Relay Networks

6.1 Multi-source Relay Network

In this section we consider a communication network similar to Section 2, albeit with multiple source nodes each having independent messages for a common destination node d as shown in Figure 3. The noisy network coding scheme has been naturally extended to this case in [2]. For the linear deterministic model, the concept of flow in [7, 8] can also be naturally extended

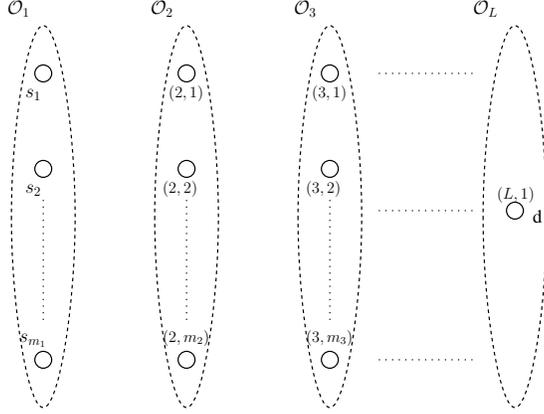


Figure 3: A layered multi-source network.

to the case of multi-source network. One way to do this is to convert the problem to the case of single-source single-destination by adding a super-node and connecting all the source nodes to the super-node by orthogonal links with capacities equal to the rate required for that source [12]. In this section, we extend the compress-and-forward scheme to the multi-source Gaussian network and as before connect it to algebraic flows.

As before, we will assume a block coding scheme with time-layering; therefore we will consider only layered networks. The source nodes together comprise the first layer and are denoted by $\{s_i\}_{i=1}^{m_1}$. We want to characterize the set of rate tuples $\{R_i\}_{i=1}^{m_1}$ achievable by the compress-and-forward scheme.

The compress-and-forward scheme for the multiple-source network is very similar to that described in Section 2.1. The difference being that now any source node s_i encodes its message $w_i \in [2^{TR_i}]$ into a T -length sequence generated using i.i.d. $p(x_{s_i})$. The relay mapping involves, as before for the unicast case, the relay nodes binning the quantized received vector before retransmitting the bin index. The binning rate at the relay node (l, k) is denoted by $R_{(l,k)}$. On receiving y_d^T , the destination node finds *unique* $\{\hat{w}_i\}_{i=1}^{m_1}$, and *any* $\hat{w}_{(l,k)}, \hat{\hat{w}}_{(l,k)}$, such that

$$\left(\left\{ x_{s_i}^T(\hat{w}_i) \right\}_{i=1}^{m_1}, \left\{ \hat{y}_{(l,k)}^T(\hat{w}_{(l,k)}, \hat{\hat{w}}_{(l,k)}) \right\}_{(l,k) \in \mathcal{V}_r}, y_d^T \right) \in \mathcal{T}_\epsilon^T. \quad (62)$$

If it is successful, the destination declares $\{\hat{w}_i\}_{i=1}^{m_1}$ as the decoded messages; if not, the destination declares an error.

The main result in this section is the following generalization of Theorem 3:

Theorem 4. *If (R_1, \dots, R_{m_1}) is a rate vector satisfying,*

$$R(\mathcal{J}) < \min_{\Omega \subseteq \mathcal{V}, \mathcal{J} \subseteq \Omega, d \in \Omega^c} \left\{ I(\hat{Y}_{\Omega^c}; X_{\Omega} | X_{\Omega^c}) - I(\hat{Y}_{\Omega}; Y_{\Omega} | X_{\mathcal{V}}) \right\}, \quad \forall \mathcal{J} \subseteq \mathcal{O}_1 \quad (63)$$

for some $\prod_{v \in \mathcal{V}} p(x_v) p(\hat{y}_v | y_v)$, then there exists a compress-and-forward scheme for the relay network with rates $\{R_i\}_{i=1}^{m_1}$ and binning rates given by,

$$R_{(l,k)} = d_{(l,k)} + I(\hat{Y}_{(l,k)}; Y_{(l,k)} | X_{\mathcal{V}}) + \delta, \quad (64)$$

where $\{d_{(l,k)}\}$ is a flow on the network with value $f = R$ and with flow values $d(1, i) = R_i$, where

$$\delta = \max_{\Omega \subseteq \mathcal{V}, d \in \Omega^c} I(\hat{Y}_\Omega; Y_\Omega | X_\mathcal{V}). \quad (65)$$

Proof. First of all we observe the following lemma:

Lemma 2. *The compress-and-forward scheme is reliable with rates $\{R_i\}_{i=1}^{m_1}$ and $R_{(l,k)}$'s if,*

$$R(\Omega_1) < R(\Omega^c \setminus \Phi) + I(\hat{Y}_\Phi; X_\Omega | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_\mathcal{V}), \quad (66)$$

$\forall \Omega \subseteq \mathcal{V}$ where $\Omega_1 \stackrel{\text{def}}{=} \Omega \cap \mathcal{O}_1$, and $\forall \Phi \subseteq \Omega^c$ with $d \in \Phi$, for some $\prod_{v \in \mathcal{V}} p(x_v)p(\hat{y}_v | y_v)$.

Proof. The proof of this lemma follows from the probability of error analysis, which is a generalization of the proof of Theorem 1 to the multi-source case. Without loss of generality we assume that the messages with index 1 is transmitted at the source and the index corresponding to the quantized vectors at each node is $(1, 1)$. We denote by $\mathcal{E}_{\{w_i\}_{i=1}^{m_1}, (w, \bar{w})_{\mathcal{V}_r}}$ the event that

$$\left(\{x_{s_i}^T(w_i)\}_{i=1}^{m_1}, \{\hat{y}_{(l,k)}^T(w_{(l,k)}, \bar{w}_{(l,k)})\}_{(l,k) \in \mathcal{V}_r}, y_d^T \right) \in \mathcal{T}_\epsilon^T. \quad (67)$$

The required probability of error is then given by

$$\mathbb{P}(\text{error}) = \mathbb{P}(\mathcal{E}_{\mathbf{1}, (1,1)_{\mathcal{V}_r}}^c) + \mathbb{P}\left(\bigcup_{\{w_i\}_{i=1}^{m_1} \neq \mathbf{1}} \mathcal{E}_{\{w_i\}_{i=1}^{m_1}, (w, \bar{w})_{\mathcal{V}_r}}\right), \quad (68)$$

where $\mathbf{1}$ denotes the vector of $(1, \dots, 1)$ of length m_1 .

From the properties of joint typicality, it can be shown that the former term goes to 0 and $T \rightarrow \infty$. The latter term can be upper bounded as following, by using cut-set partitions and union bound. Consider any $\Omega \subset \mathcal{V}$, and $\Phi \subseteq \Omega^c$; then

$$\mathbb{P}\left(\bigcup_{\{w_i\}_{i=1}^{m_1} \neq \mathbf{1}} \mathcal{E}_{\{w_i\}_{i=1}^{m_1}, (w, \bar{w})_{\mathcal{V}_r}}\right) \leq \sum_{\Omega, \Phi} P_{(\Omega, \Phi)}, \quad (69)$$

where $P_{(\Omega, \Phi)}$ is the probability corresponding to the typical event $\mathcal{E}_{\{w_i\}_{i=1}^{m_1}, (w, \bar{w})_{\mathcal{V}_r}}$ with $w_i \neq 1$ for $s_i \in \Omega$, $w_{(l,k)} \neq 1$ for only $(l, k) \in \Omega$ and $\bar{w}_{(l,k)} = 1$ for only $(l, k) \in \Phi$. $P_{(\Omega, \Phi)}$ can now be bounded as before to prove the lemma. \square

Next, we find a feasible compress-and-forward scheme using the flow formulation. The max-flow min-cut theorem specialized to the network with multiple-sources in the first layer and a single destination node in the last layer leads to the following lemma.

Lemma 3. *There exists a flow with value f for the network, with m_1 source nodes in the first layer with given flow values $\{d(1, k) : 1 \leq k \leq m_1\}$, and with 1 destination node in the last layer, if and only if*

$$d(\mathcal{J}) \leq \min_{\Omega \subseteq \mathcal{V}, d \in \Omega^c, \mathcal{J} \subseteq \Omega} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}), \quad \forall \mathcal{J} \subseteq \mathcal{O}_1. \quad (70)$$

The above lemma implies that there exists a flow $\{d_{(l,k)}\}$ on the network with value $f = R$ and with flow values $d(1,i) = R_i$. The properties of flow implies that the compress-and-forward scheme with $R_{(l,k)}$ given by (64) satisfies the conditions given by (66) and is therefore a reliable scheme. \square

6.2 Broadcasting in Relay Networks

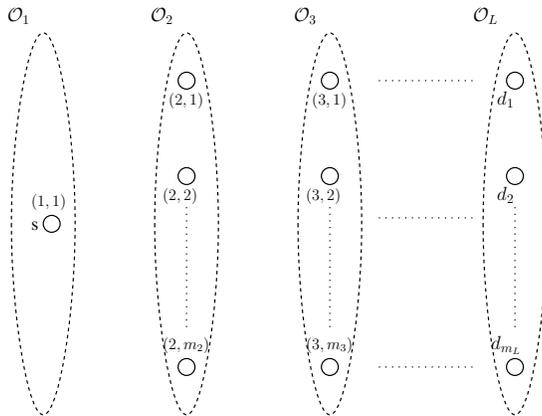


Figure 4: A layered broadcast network.

In this section, we consider the network scenario where one source node is communicating independent messages to multiple destinations as shown in Figure 4.

First of all, we observe that for this network with the layered linear deterministic model, the flow formulation of [7, 8] naturally generalizes. One way of seeing this is using the results of multi-source network and then the reciprocity in linear deterministic networks with linear coding [13]. On the other hand, there is no known scheme which achieves rates within a constant gap of the outer bound for discrete memoryless networks in general, unlike the noisy network coding scheme for the unicast and the multi-source network. However in [11], a scheme was proposed for a general deterministic network. The scheme proposed involved two steps. The *inner code* was a quantize-and-map operation at each node. The *outer code* was a Marton code over the end-to-end broadcast channel induced by the relaying scheme. In [11], the discrete superposition network, which is a deterministic network approximation network of the Gaussian relay network was then used to show that rates within a constant gap of the cut-set bound can be achieved, where the constant is independent of the channel coefficients and power constraints.

In this section, we follow the same approach as in [11]. We propose a compress-and-forward scheme for a general deterministic network, where the quantization is followed by a binning operation and only the bin index is transmitted. In effect, while the quantization earlier was at the noise level, the quantization in compress-and-forward scheme is to a much coarser level. The binning rate captures this level of quantization.

This compress-and-forward scheme holds in particular for the discrete superposition network which is obtained from the Gaussian relay network. As was done before in [11], we can use this network to perform the quantization operation at the relay for the Gaussian relay network followed then by the binning and retransmitting. In this manner, the compress-and-forward scheme is extended to the Gaussian relay network to achieve rates within a constant-gap of the cut-set bound.

For the rest of this section, we will consider only the deterministic network.

6.2.1 Compress-and-forward Scheme for the Deterministic Network

As before we will consider the layered setting alone, in which one source node s is in the first layer; and multiple destination nodes $\{d_i\}_{i=1}^{m_L}$ are in the last layer. Further we restrict ourselves to the deterministic network model of Section 5.1.

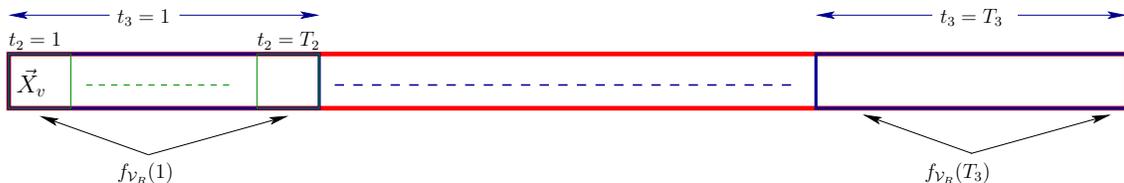


Figure 5: Coding scheme

Our random coding argument is almost similar to the one described in [11] and operates over three levels of nested blocks as shown in Figure 5. The only place it differs is at the relay mapping which now involves binning. We quickly outline the scheme from [11] elaborating only on the relay mapping. The random coding argument is described by a given product distribution $\prod_{v \in \mathcal{V}} p(X_v)$ (denoted by Q_p); the channel induces random variables $\{Y_v | v \in \mathcal{V}\}$. The coding scheme comprises of the following:

- The *message* W_i is broken into T_3 independent sub-messages $W_i(1), \dots, W_i(T_3)$, where each sub-message is encoded over a particular level-2 block.
- The *relay mappings* are done at the level-1 block. Every node blocks up T_1 received symbols and maps it to T_1 transmit symbols. This relay mapping is fixed for the duration of a level-2 block and it essentially creates an end-to-end deterministic broadcast channel at the level-2 block. The relay map $f_{(l,k)}(t_3)$ at the node (l, k) is done as follows: the set of all typical sequences $\mathcal{T}_\delta^{T_1}(Y_v)$ are binned into $2^{T_1 R_{(l,k)}}$ bins. $R_{(l,k)}$ is called the binning rate at the relay node. Each bin index is then mapped to a T_1 length sequence $y_{(l,k)}^{T_1}$ generated using i.i.d. $p(X_{(l,k)})$.
- The *encoding at the source* is done on the induced end-to-end deterministic broadcast channel for each level-2 block. Across different level-2 blocks, the mappings are generated in an i.i.d. manner and corresponding broadcast schemes are used for the different end-to-end broadcast channel induced by this operation. This is very much like a fading

broadcast channel. Here the random fading is introduced by the relay mappings. This is merely a proof technique which allows us to average the performance over random relay mappings.

- The *destinations* decode the sub-message corresponding to each level-2 block, independently and sequentially.

Our main result in this section is the following theorem:

Theorem 5. *For the deterministic broadcast network, a rate vector (R_1, \dots, R_{m_L}) satisfying*

$$R(\mathcal{J}) \leq \min_{\Omega \subseteq \mathcal{V}, s \in \Omega, \mathcal{J} \subseteq \Omega^c} H(Y_{\Omega^c} | X_{\Omega^c}), \quad \forall \mathcal{J} \subseteq \mathcal{O}_L, \quad (71)$$

for any product distribution $\prod_{v \in \mathcal{V}} p(X_v)$ is achievable by a compress-and-forward scheme with binning rates $R_{(l,k)} = d_{(l,k)}$, where $d_{(l,k)}$ is a flow on the network with value $R(\mathcal{O}_L)$ and with $d_{(L,i)} = R_i$, $\forall 1 \leq i \leq m_L$.

Proof. First of all, Theorem 2 guarantees the existence of such a flow $d_{(l,k)}$.

Next we need to show that the compress-and-forward scheme with the binning rates obtained from the flows achieves the required rate vector. Towards this, we first characterize the achievable rates under this compress-and-forward scheme with the chosen binning rates:

Proposition 2. *The achievable rate region under the above compress-and-forward scheme is given by:*

$$C(\mathcal{J}) < \frac{1}{T_1} H(Y_{\mathcal{J}}^{T_1} | \mathcal{F}_{\mathcal{V}_r}), \quad \forall \mathcal{J} \in \mathcal{O}_L. \quad (72)$$

Proof. We first consider the deterministic broadcast channel in the t_3 -th level-2 block, obtained by fixing the relay mapping to be $f_{\mathcal{V}_r}(t_3)$. Using the coding theorem for the deterministic broadcast channel, we know that the rate region for this is given by

$$C_{t_3}(\mathcal{J}) \leq \frac{1}{T_1} H(Y_{\mathcal{J}}^{T_1} | F_{\mathcal{V}_r} = f_{\mathcal{V}_r}(t_3)), \quad \forall \mathcal{J} \in \mathcal{O}_L. \quad (73)$$

Averaging across all the level-2 blocks proves the proposition. \square

Next, we prove the following lemma:

Lemma 4. *Given an arbitrary $\epsilon > 0$, $\exists T_1$ s.t. ,*

$$H(Y_{\mathcal{J}}^{T_1} | \mathcal{F}_{\mathcal{V}_r}) \geq T_1(R(\mathcal{J}) - \epsilon), \quad \forall \mathcal{J} \in \mathcal{O}_L. \quad (74)$$

Proof. The proof of this Lemma is along the lines of Lemma 1 in [11]. We consider the unicast network with the single source node and a single super destination node, which is obtained by combining the set \mathcal{J} of destination nodes. For this unicast network, we need to prove that a compress-and-forward scheme with binning rates given by $d_{(l,k)}$ can achieve at the least the rate $R(\mathcal{J})$. To prove this, we need to show that $d_{(l,k)}$ satisfies

$$R(\mathcal{J}) \leq d(\Omega^c \setminus \Phi) + H(Y_{\Phi} | X_{\Omega^c}), \quad (75)$$

$\forall \Omega \subseteq \mathcal{V}$, and $\forall \Phi \subseteq \Omega^c$, such that $s \in \Omega$, $\mathcal{J} \in \Phi$. To see this, note that the property of flow for each of the $L - 1$ layer implies:

$$d(\Phi_{l+1}) - d(\mathcal{O}_l \setminus \Omega_l) \leq H(Y_{\Phi_{l+1}} | X_{\mathcal{O}_l \setminus \Omega_l}), \quad \text{for } 1 \leq l \leq L - 1. \quad (76)$$

Adding together the $L - 1$ inequalities above gives

$$d(\Phi_L) - d(\Omega^c \setminus \Phi) \leq H(\hat{Y}_\Phi | X_{\Omega^c}). \quad (77)$$

And since $d(\Phi_L) \geq R(\mathcal{J})$,

$$R(\mathcal{J}) \leq d(\Omega^c \setminus \Phi) + H(Y_\Phi | X_{\Omega^c}). \quad (78)$$

Using Fano's inequality, as is done in [11], we can show that the mutual information per symbol between the source and the destination which is $H(Y_{\mathcal{J}}^{T_1} | \mathcal{F}_{\mathcal{V}_r})/T_1$ is close to $R(\mathcal{J})$. \square

It follows from the above lemma that the rate vector (R_1, \dots, R_{m_L}) lies within the region given by (72) and is therefore achievable. \square

7 Conclusion

In this paper, we showed that the compress-and-forward scheme with an appropriate condition on decodability at the destination node is approximately optimal for the Gaussian relay network. This scheme achieves identical reliable rate of communication as quantize-map-and-forward and noisy network coding schemes. An important aim in this work is to characterize the minimal compression rates at the relay nodes. This is done by an abstract algebraic flow formulation, that generalizes the one on the linking system in the context of the linear deterministic channel [8]. Computing the the flow, and hence the optimal compression rates at the relay nodes, can be done in polynomial time. We note, however, that this computation is a *centralized* procedure needing *global* channel state information. In contrast, the noisy network coding and quantize-map-and-forward schemes are distributed.

In [14], a practical coding and signaling framework is presented for the quantize-and-forward scheme for a small relay network. In their framework, LDPC codes are used at both the source and relay nodes. The decoding can be simplified to decoding over Tanner graphs, which incorporates both the tanner graphs from the LDPC encoders as well as the network structure and quantization operations. Our results suggest the minimal quantization and transmission rate to be used at relay nodes. Doing so, would reduce the complexity of the related Tanner graphs and hence lead to more efficient decoding algorithms.

A Proof of Lemma 1

We will prove the lemma for $r_B(T)$. The proof for $r_A(T)$ is similar.

1. Submodularity:

Let,

$$r_B(T^{(1)}) = C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \quad \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)} \quad (79)$$

$$r_B(T^{(2)}) = C_B(\Omega_B^{(2)}) + d(\Omega_B^{(2)} \cap \mathcal{O}_L), \quad \Omega_B^{(2)} \cap \mathcal{O}_{L_0} = T^{(1)}. \quad (80)$$

Since,

$$(\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cup T^{(2)}, \quad (81)$$

$$(\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cap T^{(2)}, \quad (82)$$

it follows that

$$r_B(T^{(1)} \cup T^{(2)}) \leq C_B(\Omega_B^{(1)} \cup \Omega_B^{(2)}) + d((\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_L), \quad (83)$$

$$r_B(T^{(1)} \cap T^{(2)}) \leq C_B(\Omega_B^{(1)} \cap \Omega_B^{(2)}) + d((\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_L). \quad (84)$$

By definition of cut and the bi-submodularity of ρ_l , it is easy to verify that $C_B(\Omega_B)$ is submodular. And since d is an additive function, it then follows that $r_B(T)$ is submodular.

2. Non-decreasing:

Consider $T^{(1)} \subseteq T^{(2)}$. Let

$$r_B(T^{(1)}) = C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \quad \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)}. \quad (85)$$

Let $\Omega_B = \Omega_B^{(1)} \cup T^{(2)} \setminus T^{(1)} \supseteq \Omega_B^{(1)}$, so that $\Omega_B \cap \mathcal{O}_{L_0} = T^{(2)}$. By the definition of cut and the non-decreasing property of ρ_l , it follows that $C_B(\Omega_B^{(1)}) \leq C_B(\Omega_B)$. Also $d(\Omega_B^{(1)} \cap \mathcal{O}_L) \leq d(\Omega_B \cap \mathcal{O}_L)$. Therefore

$$r_B(T^{(2)}) = C_B(\Omega_B) + d(\Omega_B \cap \mathcal{O}_L) \quad (86)$$

$$\geq C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L) \quad (87)$$

$$= r_B(T^{(1)}). \quad (88)$$

3. $r_B(\emptyset) = 0$:

When $T = \emptyset$, by letting $\Omega_B = \emptyset$, it follows that $r_B(\emptyset) = 0$.

B Proof of Theorem 1

Without loss of generality we assume that the message with index 1 is transmitted at the source and the index corresponding to the quantized vectors at each node is (1, 1). We will

find the probability of error that this message is wrongly decoded at the destination. We denote by $\mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}$ the event that

$$\left(x_s^T(w), \{\hat{y}_{(l,k)}^T(w_{(l,k)}, \bar{w}_{(l,k)}), x_{(l,k)}^T(w_{(l,k)})\}_{(l,k) \in \mathcal{V}_r}, y_d^T\right) \in \mathcal{T}_\epsilon^T. \quad (89)$$

The required probability of error is then given by

$$\mathbb{P}(\text{error}) = \mathbb{P}(\mathcal{E}_{1,(1,1)_{\mathcal{V}_r}}^c) + \mathbb{P}\left(\bigcup_{w \neq 1} \mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}\right). \quad (90)$$

From the properties of joint typicality, it can be shown that the former term goes to 0 and $T \rightarrow \infty$. The latter term can be upper bounded as following, by using cut-set partitions and union bound. Consider any $\Omega \subset \mathcal{V}_r$, and $\Phi \subseteq \mathcal{V}_r \setminus \Omega$; then

$$\mathbb{P}\left(\bigcup_{w \neq 1} \mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r} \mid \mathcal{E}_{1,(1,1)_{\mathcal{V}_r}}}\right) \leq \sum_{\Omega, \Phi} P_{(\Omega, \Phi)}, \quad (91)$$

where $P_{(\Omega, \Phi)}$ is the probability corresponding to the typical event $\mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}$ with $w \neq 1$, $w_{(l,k)} \neq 1$ for only $(l, k) \in \Omega$ and $\bar{w}_{(l,k)} = 1$ for only $(l, k) \in \Phi$. It can be shown that

$$\begin{aligned} P_{(\Omega, \Phi)} &= 2^{T(R+R(\Omega)+\bar{R}(\Phi^c))} 2^{T(H(Y_d, \hat{Y}_\Phi, \hat{Y}_{\Phi^c}, X_\Omega, X_{\Omega^c}, X_s) - H(X_\Omega, X_s) - H(Y_d, \hat{Y}_\Phi, X_{\Omega^c}) - \sum_{(l,k) \in \Phi^c} H(\hat{Y}_{(l,k)}))} \\ &= 2^{T(R+R(\Omega)+\bar{R}(\Phi^c))} 2^{T(H(Y_d, \hat{Y}_\Phi, \hat{Y}_{\Phi^c} | X_\Omega, X_{\Omega^c}, X_s) - H(Y_d, \hat{Y}_\Phi | X_{\Omega^c}) - \sum_{(l,k) \in \Phi^c} H(\hat{Y}_{(l,k)}))} \\ &= 2^{T(R+R(\Omega)+\bar{R}(\Phi^c))} 2^{-T(H(Y_d, \hat{Y}_\Phi | X_{\Omega^c}) - H(Y_d, \hat{Y}_\Phi | X_\Omega, X_{\Omega^c}, X_s) + \sum_{(l,k) \in \Phi^c} H(\hat{Y}_{(l,k)}) - H(\hat{Y}_{\Phi^c} | X_\Omega, X_{\Omega^c}, X_s))} \\ &= 2^{T(R+R(\Omega)+\bar{R}(\Phi^c))} 2^{-T(I(Y_d, \hat{Y}_\Phi; X_\Omega, X_s | X_{\Omega^c}) + \sum_{(l,k) \in \Phi^c} I(\hat{Y}_{(l,k)}; X_{\mathcal{V}_r}, X_s))} \end{aligned}$$

Using the Markovian property of the random variables, we have that

$$I(\hat{Y}_{(l,k)}; X_{\mathcal{V}_r}, X_s) = I(\hat{Y}_{(l,k)}; Y_{(l,k)}) - I(\hat{Y}_{(l,k)}; Y_{(l,k)} | X_{\mathcal{V}_r}, X_s), \quad (92)$$

and using (7) we have

$$P_{(\Omega, \Phi)} = 2^{T(R - R(\Omega^c \setminus \Phi) - I(Y_d, \hat{Y}_\Phi; X_\Omega, X_s | X_{\Omega^c}) + I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}_r}, X_s))}. \quad (93)$$

Therefore $P_{(\Omega, \Phi)} \rightarrow 0$, if

$$R < R(\Omega^c \setminus \Phi) + I(Y_d, \hat{Y}_\Phi; X_\Omega, X_s | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}_r}, X_s). \quad (94)$$

C Proof of Proposition 1

We need to show that $I(X_U; \hat{Y}_W | X_{\mathcal{O}_i \setminus U})$ satisfies the three properties of channel functions. Firstly we show that it is bi-submodular.

$$I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U}) = H(\hat{Y}_W | X_{\mathcal{O}_l \setminus U}) - H(\hat{Y}_W | X_{\mathcal{O}_l}) \quad (95)$$

$$= H(\hat{Y}_W, X_{\mathcal{O}_l \setminus U}) - H(X_{\mathcal{O}_l \setminus U}) - H(\hat{Y}_W | X_{\mathcal{O}_l}). \quad (96)$$

The submodularity of entropy [10] implies that $H(\hat{Y}_W, X_{\mathcal{O}_l \setminus U})$ is bi-submodular. The submodularity of entropy follows from the fact that given collection of random variables Υ_1 and Υ_2 , we have

$$H(\Upsilon_1) + H(\Upsilon_2) - H(\Upsilon_1 \cup \Upsilon_2) - H(\Upsilon_1 \cap \Upsilon_2) = I(\Upsilon_1 \setminus \Upsilon_2; \Upsilon_2 \setminus \Upsilon_1 | \Upsilon_1 \cap \Upsilon_2) \quad (97)$$

$$\geq 0. \quad (98)$$

The product form of the random variables implies that $H(X_{\mathcal{O}_l \setminus U})$ and $H(\hat{Y}_W | X_{\mathcal{O}_l})$ are modular or additive. Therefore, $I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U})$ is bi-submodular.

Next, we show the non-decreasing property. Given $U_1 \subseteq U \subseteq \mathcal{O}_l$ and $W_1 \subseteq W \subseteq \mathcal{O}_{l+1}$, we have

$$I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \setminus U}) = H(X_U | X_{\mathcal{O}_l \setminus U}) - H(X_U | X_{\mathcal{O}_l \setminus U} \hat{Y}_W) \quad (99)$$

$$\geq H(X_U | X_{\mathcal{O}_l \setminus U}) - H(X_U | X_{\mathcal{O}_l \setminus U} \hat{Y}_{W_1}) \quad (100)$$

$$= I(X_U; \hat{Y}_{W_1} | X_{\mathcal{O}_l \setminus U}) \quad (101)$$

$$= H(\hat{Y}_{W_1} | X_{\mathcal{O}_l \setminus U}) - H(\hat{Y}_{W_1} | X_{\mathcal{O}_l}) \quad (102)$$

$$\geq H(\hat{Y}_{W_1} | X_{\mathcal{O}_l \setminus U_1}) - H(\hat{Y}_{W_1} | X_{\mathcal{O}_l}) \quad (103)$$

$$= I(X_{U_1}; \hat{Y}_{W_1} | X_{\mathcal{O}_l \setminus U_1}), \quad (104)$$

where both the inequalities follow from the fact that conditioning reduces entropy.

The third property is readily seen.

Acknowledgments

We would like to thank Chandra Chekuri for the many useful discussions.

References

- [1] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse “Wireless network information flow: A deterministic Approach,” submitted to the *IEEE Trans. Info. Theory*.
- [2] S. H. Lim, Y-H. Kim, A. El Gamal, and S-Y. Chung, “Noisy Network Coding,” <http://arxiv.org/abs/1002.3188>.
- [3] A. Ozgur, and S. N. Diggavi, “Approximately achieving Gaussian relay network capacity with lattice codes,” in *Proceedings of IEEE Intl. Symp. on Info. theory 2010, Austin, Texas* June 2010. See also <http://arxiv.org/abs/1005.1284>.

- [4] T. M. Cover, and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Info. Theory*, vol. 25, no. 5, pp. 572584, Sep. 1979.
- [5] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Trans. Info. Theory*, vol. 51, no. 9, pp. 30373063, Sep. 2005.
- [6] A. Amadruz, and C. Fragouli, “Combinatorial algorithms for wireless information flow,” in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Jan. 2009.
- [7] S. M. S. Yazdi and S. A. Savari, “A combinatorial study of linear deterministic relay networks,” in *Proceedings of Allerton Conference on Communications, Control, and Computing*, Sep. 2009.
- [8] M. X. Goemans, S. Iwata, and R. Zenklusen, “An algebraic framework for wireless information flow,” in *Proceedings of Allerton Conference on Communications, Control, and Computing*, Sep. 2009.
- [9] A. Schrijver, “Combinatorial Optimization,” Springer, Berlin, 2003.
- [10] A. K. Kelmans and B. N. Kimelfeld, “Multiplicative submodularity of a matrixs principal minor as a function of the set of its rows and some combinatorial applications,” *Discrete Mathematics* Vol. 44(1), pp. 113-116, 1980.
- [11] S. Kannan, A. Raja, and P. Viswanath, “Approximately optimal wireless broadcasting,” submitted to *IEEE Trans. Info. Theory*, <http://arxiv.org/abs/>?
- [12] M. M. Butt, G. Caire, and R. R. Müller, “Linear Finite-Field Deterministic Networks with Many Sources and One Destination,” <http://arxiv.org/abs/1001.2464>.
- [13] A. Raja, V. Prabhakaran, P. Viswanath, “Reciprocity in Linear Deterministic Networks under Linear Coding,” *Proc. IEEE Info. Theory Workshop*, Volos, Greece, Jun. 2009
- [14] V. Nagpal, I. Wang, M. Joragovanovic, D. Tse, and B. Nikolić “Quantize-Map-and-Forward Relaying: Coding and System Design,” in *Proceedings of Allerton Conference on Communications, Control, and Computing*, Sep. 2010.