# Compress-and-Forward Scheme for Relay Networks: Backword Decoding and Connection to Bisubmodular Flows

Adnan Raja, *Member, IEEE*, and Pramod Viswanath, *Fellow, IEEE*

***Abstract*— In this paper, a compress-and-forward scheme with backward decoding is presented for the unicast wireless relay network. The encoding at the source and relay is a generalization of the noisy network coding (NNC) scheme. While it achieves the same reliable data rate as NNC scheme, the backward decoding allows for a better decoding complexity as compared with the joint decoding of the NNC scheme. Characterizing the layered decoding scheme is shown to be equivalent to characterizing an information flow for the wireless network. A node-flow for a graph with bisubmodular capacity constraints is presented and a max-flow min-cut theorem is presented. This generalizes many well-known results of flows over capacity constrained graphs studied in computer science literature. The results for the unicast relay network are generalized to the network with multiple sources with independent messages intended for a single destination.**

***Index Terms*— Wireless communication, relay networks, capacity, bisubmodular flows, max flow min-cut.**

## I. Introduction

THE primary focus of this paper is a unicast wireless relay network: a single source node intends to communicate reliably with a single destination node with the assistance of many relay nodes. The communication channels are wireless; transmitted signals from a node are *broadcasted* to all other nodes; received signals at a node is a linear *superposition* of the transmit signals with a random additive noise, which has the familiar Gaussian distribution.

In [2] a quantize-map-forward scheme was presented for the wireless relay network. It was shown that this scheme is approximately optimal, i.e. it gives a reliability criterion for rates within a constant gap of the cutset 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. The quantized symbols accumulated together in a block are then mapped to a transmit codeword at that node. These transmission codebooks at every node are generated independently of each other.

In [3], a related scheme was presented for the wireless relay network. Here, the coding and quantization is done in a structured manner using lattices. The scheme was shown to achieve performance similar to the quantize-map-forward scheme of [2] in terms of the reliable rates.

In [1], a *noisy network coding* scheme in the more general setting of the discrete memoryless network was presented for the unicast relay network and also generalized to the case of multicast and multiple sources with single destination. In this scheme, the relay quantizes the received signal in blocks using vector-quantization, subsequently mapping each quantized codeword to a unique codeword, which is re-transmitted by the relay. Specialized to the wireless network, the noisy network coding can be thought of as a vector version of the quantize-map-forward scheme, where each relay does a vector quantization rather than the scalar quantization proposed in [2].

In [4], an alternate approach was provided, wherein the discrete superposition network was used as a digital interface for the wireless network and the scheme was constructed by lifting the scheme for the discrete superposition network. The discrete superposition network provided the quantization interface for this scheme.

In this paper, a compress-and-forward scheme is presented for a relay network in the general setting of the discrete memoryless network. This encoding is similar to the noisy network scheme, but the relay mapping is generalized, so that the relay node compresses the received signal in blocks, on top of the vector quantization in NNC. The additional compression does not increase the achievable rate beyond the rate achievable by NNC; however, the first main result of this paper is that, if the compression rates are chosen appropriately then a lower complexity backward decoding achieves approximately the same rate. The above result was also proved independently in [5]. The second important result in this paper is to show that this appropriate choice of compression rates can be computed efficiently by computing a node-flow on a bisubmodular capacitated graph. The flow formulation captures the rate of *actual* information that should

Manuscript received January 10, 2011; revised October 20, 2013; accepted April 28, 2014. Date of publication July 17, 2014; date of current version August 14, 2014. This work was supported in part by the National Science Foundation under Grant CCF-1017430 and Grant ECCS-1232257, in part by the Army Research Office under Grant W911NF-14-1-0220, and in part by Intel Corporation, Santa Clara, CA, USA. This paper was presented at the 2011 IEEE International Symposium on Information Theory.

A. Raja was with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61801 USA. He is now with Fastback Networks, San Jose, CA 95131 USA (e-mail: araja2@illinois.edu).

P. Viswanath is with the Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Champaign, IL 61801 USA (e-mail: pramodv@illinois.edu).

Communicated by C. Fragouli, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2334328

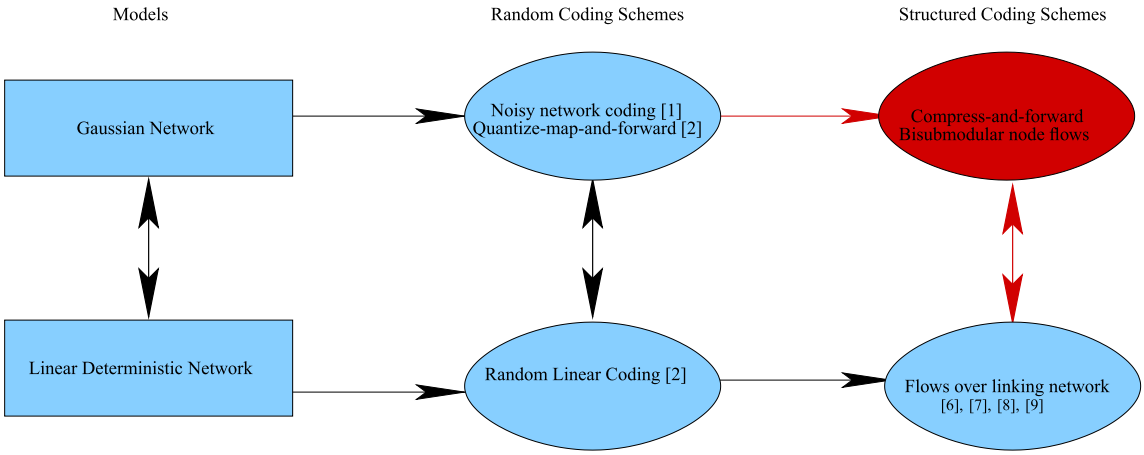


Fig. 1. A depiction of the communication schemes on the Gaussian and linear deterministic networks. The main result of this paper is represented by the upper-right bubble in red.

flow through each node to support a given rate of flow of information from the source to the destination. In other words, this paper shows that backward decoding does almost as good as joint decoding, if the relay nodes compress their signals to capture the right amount of information that should flow through that given the network topology.

The paper presents a max-flow min-cut result for a node-flow on a bisubmodular capacitated graph. This is related to many well-known results of flows over capacity constrained graphs studied in computer science literature, albeit with two differences; the first one being that the flow is defined over nodes rather than the conventional approach of defining over edges; and the second is that the graphs are restricted to layered graphs alone. The first difference is a fundamental difference. Flows over graphs are conventionally defined as numbers over edges of the graph, such that for every node the incoming-flow is equal to the outgoing-flow. Since the motivation here is to model the wireless network where there are no physical edges, it is more appropriate to define node-flow rather than edge-flow; the relation being that the node-flow represents the incoming-flow or outgoing-flow at the node. The second is less fundamental and the restriction to layered graphs is done only because the block-coding scheme for the relay network can be studied by considering a virtual layered network. the layering offers a convenient way of defining the bisubmodular capacity functions on the layered graph.

The bisubmodular capacitated graph presented here is motivated by the ideas of linking systems and flows introduced in [6]–[9] in the context of the linear deterministic network. The linear deterministic network was introduced in [2] 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 [6], [7] developed a polynomial time algorithm that discovers the relay encoding strategy using a notion of linear independence between channels. Taking this concept forward, in [8] and [9], the concept of flow was introduced for the linear deterministic network. The flow value at each node in this network corresponds to the number of independent equations, that particular node needs to forward. The result in this paper can be viewed as a loose analog of these results in the context of the Gaussian network; see Figure 1. The additional structure of the linear deterministic channel, is used in [6]–[9] to show that a single-block coding scheme where a simple permutation matrix at each node mapping the received vector to the transmit vector is optimal. Both the flow values at the node and the permutation mapping were constructed in polynomial time.

The rest of the paper is organized as follows. In Section II the compress-and-forward scheme for the relay network is described and characterized. A lower-complexity layered decoding is presented and the achievable rates are characterized. It is shown that this decoding scheme does as well as the joint decoding scheme. To prove this result, the notion of node-flows for a bisubmodular capacitated graph is developed in Section III. In Section IV, the results are generalized to the network with multiple sources with independent messages intended for a single destination. In Section V 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.

## II. Unicast Relay Network

A communication network is represented by 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 $y_v \in \mathcal{Y}_v$. *Memoryless* network will be considered here wherein the received symbol at any node at any given time depends

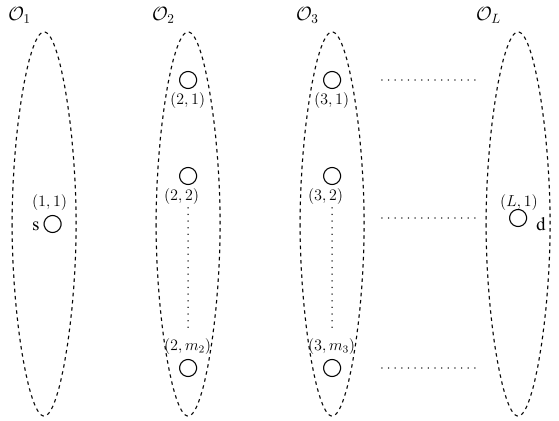


Fig. 2. A layered network.

(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 $t \in [T]$,
$$f_{s,t} : (W \times \mathcal{Y}_s^{t-1}) \to \mathcal{X}_s. \tag{1}$$
3) The *relay mappings* for each $v \in \mathcal{V} \backslash \{s\}$ and $t \in [T]$,
$$f_{v,t} : \mathcal{Y}_v^{t-1} \to \mathcal{X}_v. \tag{2}$$
4) The *decoding map* at destination $d$,
$$g_d : \mathcal{Y}_d^T \to \hat{W}. \tag{3}$$

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

$$P_e \stackrel{\text{def}}{=} \Pr\{\hat{W} \neq W\}. \tag{4}$$

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 $\epsilon$ 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 [2] 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.

Henceforth, the focus will be only on an $L$-layered network as shown in Figure 2, so that

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

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 $\mathcal{V}_r$, i.e.,

$$\mathcal{V}_r = \bigcup_{l=2}^{L-1} \mathcal{O}_l. \tag{6}$$

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. Therefore, for the layered network the channel which is denoted by a transition probability function can be simplified into a product across layers as follows:

$$p\left(y_{\mathcal{V}} | x_{\mathcal{V}}\right) = \prod_{l=1}^{L-1} p\left(y_{\mathcal{O}_{l+1}} | x_{\mathcal{O}_l}\right). \tag{7}$$

The noise across each relay node is assumed to be independent, which implies that the channel function for each layer is further given by,

$$p(y_{\mathcal{O}_{l+1}} | x_{\mathcal{O}_l}) = \prod_{k=1}^{m_{l+1}} p(y_{(l+1,k)} | x_{\mathcal{O}_l}). \tag{8}$$

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 layered network.

In particular, if the received symbol is a deterministic function of the transmitted symbols, i.e.,

$$y_{\mathcal{O}_{l+1}} = g_l\left(x_{\mathcal{O}_l}\right), \tag{9}$$

then the network is called a *deterministic network*. Further, if the transmit and received symbols are restricted to vectors over finite fields and the deterministic function is modeled as a linear function, such that

$$y_{\mathcal{O}_{l+1}} = G_l x_{\mathcal{O}_l}, \tag{10}$$

then the network is called a *linear deterministic network*. If the network is a wireless network, then the alphabet sets are complex and the probability transition function linear with an additive complex Gaussian noise $z_v$, such that,

$$y_v = \sum_{u \in \mathcal{O}_l} h_{v,u} x_u + z_v, \tag{11}$$

where $v \in \mathcal{O}_{l+1}$. The wireless network is the one with the most practical interest and in [2] it was shown that the linear deterministic network captures many features of the wireless network.

### A. Compress-and-Forward Scheme

In this section, the compress-and-forward scheme is described and it's performance is characterized. It is a block-encoded scheme where each node performs its operation over blocks of time symbols. The relay node quantizes (or compresses) the symbols it receives over a block of time

to finite bits. These bits are then transmitted in the next block. The compression rate at a relay node is defined to be the rate of transmission of the compressed bits.

Assuming that uniformly sized blocks of $T$ symbols are used by each node for this operation, a compress-and-forward scheme is parametrized by $\left(T, R, \{r_v\}_{v\in\mathcal{V}_r}\right)$, where $R$ is the overall rate of communication and $r_v$'s are the compression rates at the relay nodes. A rate vector $\left(R, \{r_v\}_{v\in\mathcal{V}_r}\right)$ is said to be feasible w.r.t. the compress-and-forward scheme, if for any arbitrary $\epsilon > 0$, there exists a compress-and-forward scheme $\left(T, R, \{r_v\}_{v\in\mathcal{V}_r}\right)$ which achieves a probability of error less than $\epsilon$.

The following theorem characterizes the feasible region of $\left(R, \{r_v\}_{v\in\mathcal{V}_r}\right)$ for the compress-and-forward scheme.

*Theorem 1: A rate vector $\left(R, \{r_v\}_{v\in\mathcal{V}_r}\right)$ is feasible if for some collection of random variables $\left\{X_\mathcal{V}, \hat{Y}_\mathcal{V}\right\}$, henceforth denoted by $Q_p$, which is distributed as*

$$p(X_\mathcal{V}, \hat{Y}_\mathcal{V}, Y_\mathcal{V}) = \left(\prod_{v\in\mathcal{V}} p(X_v)\right) p(Y_\mathcal{V}|X_\mathcal{V}) \left(\prod_{v\in\mathcal{V}} p(\hat{Y}_v|Y_v)\right), \tag{12}$$

*the vector $\left(R, \{r_v\}_{v\in\mathcal{V}_r}\right)$ satisfies*

$$R < r(\Omega^c\backslash\Phi) + I(\hat{Y}_\Phi; X_\Omega|X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c}|X_\mathcal{V}), \tag{13}$$

$\forall\ \Omega, \Phi,$ *s.t.,* $S \in \Omega \subseteq \mathcal{V}, D \in \Phi \subseteq \Omega^c$, *where* $r(\mathcal{A}) \stackrel{\text{def}}{=} \sum_{v\in\mathcal{A}} r_v$.

Note 1: The choice $\hat{Y}_D = Y_D$ is always optimal for (13).

Note 2: In the usual cut-set definition, the node-set is partitioned into two sets; a set containing the source $\Omega$ and the complementary set $\Omega^c$, containing the destination. However, here the node set is partition into a set containing the source - $\Omega$, a set containing the destination - $\Omega^c$, and the rest.

*Proof:* The proof is by random coding technique. A random ensemble of coding scheme is defined using the collection of random variables $Q_p$ distributed as given by (12). A scheme in the ensemble is generated as follows.

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. $p(X_S)$.
2) *Relay Codebooks and Mappings:* For every relay node $v \in \mathcal{V}_r$ a binned quantization codebook is generated with $2^{Tr_v}$ bins. The binned quantization codebook is given by $\hat{y}_v^T(w_v, \bar{w}_v)$, where $w_v \in [2^{Tr_v}]$ and $\bar{w}_v \in [2^{T\bar{r}_v}]$. And it is generated using i.i.d. $p(\hat{Y}_v)$.
   Every relay node also generates a transmission codebook of size $2^{Tr_v}$, which consists of $x_v^T(w_v)$ sequences generated using i.i.d. $p(X_v)$.
   On receiving $y_v^T$, the relay node finds a vector $\hat{y}_v^T(w_v, \bar{w}_v)$ in the quantization codebook that is jointly typical with $y_v^T$, and transmits $x_v^T(w_v)$ 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. The probability that this latter event is arbitrarily is small is ensured by letting
   $$\bar{r}_v = I(Y_v; \hat{Y}_v) - r_v + \epsilon_1, \tag{14}$$
   for an arbitrarily small $\epsilon_1 > 0$. This ensures that the total size of the quantization codebook is of the order $2^{TI(Y_v,\hat{Y}_v)}$.
3) *Decoding:* On receiving $y_D^T$, the destination node finds a *unique* $\hat{w}$, and *any* $\{(\hat{w}_v, \hat{\bar{w}}_v)\}_{v\in\mathcal{V}_r}$, such that
   $$\left(x_S^T(\hat{w}), \left\{\hat{Y}_v^T(\hat{w}_v, \hat{\bar{w}}_v), x_v^T(\hat{w}_v)\right\}_{v\in\mathcal{V}_r}, y_D^T\right) \in \mathcal{T}_\epsilon^T. \tag{15}$$
   If it is successful, the destination declares $\hat{w}$ as the decoded message; if not, the destination declares an error.

The theorem follows by the standard argument of showing that the average probability of error, averaged over the ensemble of codes and over all messages, goes to 0 as $T$ tends to infinity. The details of the error probability analysis are in Appendix A.

□

In the usual communication problem setup, one is interested in only maximizing the overall communication rate $R$. The following corollary of the above theorem establishes the achievable rate by the compress-and-forward scheme.

*Corollary 1: The communication rate $R$ is achievable by the compress-and-forward scheme if*

$$R < \min_{\Omega\subseteq\mathcal{V}, S\in\Omega} I(\hat{Y}_{\Omega^c}; X_\Omega|X_{\Omega^c}) - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V},), \tag{16}$$

*for some collection of random variables $Q_p$.*

*Proof:* The corollary can be proved by showing that the RHS of (16) is always greater than the RHS of (13). To shows this, we will choose the compression rates of the compress-and-forward scheme to be $r_v = I(Y_v; \hat{Y}_v) + \epsilon_1$. (Note that this is the maximum choice for the compression rate as this makes $\bar{r}_v = 0$ in (14)). With this choice of compression rates,

$$\begin{aligned}
\text{RHS of (13)} &> I(Y_{\Omega^c\backslash\Phi}; \hat{Y}_{\Omega^c\backslash\Phi}) + I(\hat{Y}_\Phi; X_\Omega|X_{\Omega^c}) \\
&\quad - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c}|X_\mathcal{V}) && (17)\\
&= I(\hat{Y}_\Phi; X_\Omega|X_{\Omega^c}) + I(Y_{\Omega^c\backslash\Phi}; \hat{Y}_{\Omega^c\backslash\Phi}) \\
&\quad - I(Y_{\Omega^c\backslash\Phi}; \hat{Y}_{\Omega^c\backslash\Phi}|X_\mathcal{V}) - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (18)\\
&= H(\hat{Y}_\Phi|X_{\Omega^c}) - H(\hat{Y}_\Phi|X_\mathcal{V}) + H(Y_{\Omega^c\backslash\Phi}) \\
&\quad - H(\hat{Y}_{\Omega^c\backslash\Phi}|X_\mathcal{V}) - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (19)\\
&> H(\hat{Y}_\Phi|X_{\Omega^c}) - H(\hat{Y}_\Phi|X_\mathcal{V}) + H(Y_{\Omega^c\backslash\Phi}|X_{\Omega^c}) \\
&\quad - H(\hat{Y}_{\Omega^c\backslash\Phi}|X_\mathcal{V}) - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (20)\\
&= H(\hat{Y}_\Phi|X_{\Omega^c}) + H(Y_{\Omega^c\backslash\Phi}|X_{\Omega^c}) - H(\hat{Y}_{\Omega^c}|X_\mathcal{V}) \\
&\quad - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (21)\\
&> H(Y_{\Omega^c}|X_{\Omega^c}) - H(\hat{Y}_{\Omega^c}|X_\mathcal{V}) \\
&\quad - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (22)\\
&= I(Y_{\Omega^c}; X_\Omega|X_{\Omega^c}) - I(\hat{Y}_\Omega; Y_\Omega|X_\mathcal{V}) && (23)\\
&= \text{RHS of (16)}. && (24)
\end{aligned}$$

□

It should be noted that the achievable rate in (16) is the same as the one obtained in noisy network coding scheme in [1]. This is not surprising as by allowing the compression rates to be large enough, the scheme essentially reduces to the noisy

network coding scheme, where every quantized codeword is uniquely mapped to a re-transmission codeword at the relay node.

### B. A Layered Decoding Scheme

A maximum likelihood decoder maximizes the probability of the received vector conditioned on the transmitted codeword at the source. Note that the jointly-typical-set decoding is a proof technique for the random coding argument and it upper-bounds the error probability that can be achieved by the maximum likelihood (ML) decoder.

$$\text{ML decoder:} \quad \hat{w} = \text{argmax}_w \, p\big(y_D^T | x_S^T(w)\big). \tag{25}$$

The conditional probability depends on the channel model and the operations (quantization, compression and mapping) at each node. Therefore implementing a ML decoder has very high complexity. In this section, a layered decoding architecture is presented for the compress-and-forward scheme which operates layer-by-layer and decodes the compressed bits transmitted by each relay node. As will be discussed in Section VI, the layered decoding can help reduce the decoding complexity.

*Layered Decoding Scheme:* The decoder at the destination node operates backwards layer-by-layer. First, it decodes the messages (or compressed bits) transmitted by the nodes in the layer $\mathcal{O}_{L-1}$. Then using these decoded messages, it decodes the messages in the layer $\mathcal{O}_{L-2}$. This process continues till the destination node eventually decodes the source message. Note that the layered decoding scheme is the same as the backward decoding for the block-encoding schemes in relay networks.

The following theorem characterizes the feasible region of $\big(R, \{r_v\}_{v\in\mathcal{V}_r}\big)$.

*Theorem 2: A rate vector $\big(R, \{r_v\}_{v\in\mathcal{V}_r}\big)$ is feasible for the compress-and-forward scheme, under the layered decoding scheme, if for some $Q_p$ the vector $\big(R, \{r_v\}_{v\in\mathcal{V}_r}\big)$ satisfies*

$$r(U) \le I(X_U; Y_D | X_{\mathcal{O}_{L-1}\setminus U}), \ \forall\, U \subseteq \mathcal{O}_{L-1}, \tag{26}$$

$$\begin{aligned} r(U) - r(\mathcal{O}_{l+1}\setminus V) \le\ & I(X_U; \hat{Y}_V | X_{\mathcal{O}_l\setminus U}) \\ & - I(\hat{Y}_{\mathcal{O}_{l+1}\setminus V}; Y_{\mathcal{O}_{l+1}\setminus V} | X_{\mathcal{O}_l}), \\ & \forall\, U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}, 2 \le l \le L-2, \end{aligned} \tag{27}$$

$$\begin{aligned} R - r(\mathcal{O}_2\setminus V) \le\ & I(X_S; \hat{Y}_V) - I(\hat{Y}_{\mathcal{O}_{l+1}\setminus V}; Y_{\mathcal{O}_{l+1}\setminus V} | X_S), \\ & \forall V \subseteq \mathcal{O}_2. \end{aligned} \tag{28}$$

*Proof:* The proof is by backward induction. Assuming that the destination has decoded the messages transmitted by the relay nodes in layer $\mathcal{O}_{l+1}$, the probability of error for decoding the messages from the layer $\mathcal{O}_l$ is considered. To do so, a hypothetical layered network as shown in Figure 3 is considered. This network consists of the layers $\mathcal{O}_l$ and $\mathcal{O}_{l+1}$ and in addition a layer with an aggregator node $A$. A node $v_{(l+1,j)}$ in layer $\mathcal{O}_{l+1}$ is connected to the aggregator node with wired link of capacity $r_{v_{(l+1,j)}}$ bits per symbol. This layer represents the forward part of the network beyond layer $\mathcal{O}_{l+1}$.

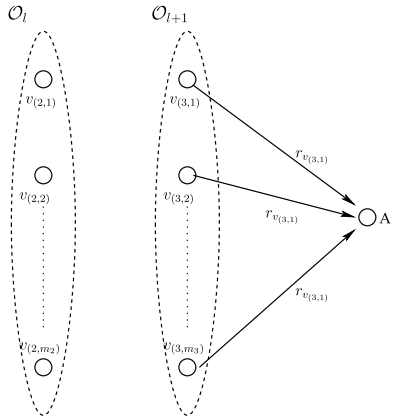


Fig. 3. A hypothetical network.

This network is now a multiple-source single-destination relay network, with all the nodes in layer $O_l$ being source nodes and the aggregator node as the destination node. The node $v_{(l,j)}$ has a message for the aggregator node with rate $r_{v_{(l,j)}}$. The noisy network coding scheme [1] assures that the messages can be decoded with arbitrarily small probability of error, if

$$r(U) - r(\mathcal{O}_{l+1}\setminus V) \le I(X_U; \hat{Y}_V | X_{\mathcal{O}_l\setminus U}) - I(\hat{Y}_{V^c}; Y_{V^c} | X_{\mathcal{O}_l}), \tag{29}$$

$\forall\, U \subseteq \mathcal{O}_l, V \subseteq \mathcal{O}_{l+1}$, where the above inequality corresponds to the cut $\Omega = U \bigcup V^c$.

□

Note that the layered decoding scheme is weaker than the ML decoding scheme. Therefore the feasible region under the layered decoding scheme should be a strict subset of the feasible region under the ML decoding scheme.

However, the following theorem shows that the compress-and-forward scheme with layered decoding achieves similar communication rate as the noisy network coding scheme.

*Theorem 3: The communication rate $R$ is achievable by the compress-and-forward scheme with layered decoding if for some collection of random variables $Q_p$,*

$$R < \min_{\Omega\subseteq\mathcal{V}, S\in\Omega} I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - \kappa_1, \tag{30}$$

*where the constant $\kappa_1$ is given by the recursive relation,*

$$\kappa_l = I(\hat{Y}_{\mathcal{O}_{l+1}}; Y_{\mathcal{O}_{l+1}} | X_{\mathcal{O}_l}) + \kappa_{l+1} |\mathcal{O}_{l+1}|, \tag{31}$$

*and $\kappa_{L-1} = 0$.*

*Proof:* The above theorem will be proved by characterizing an *information flow* for the network in the Section III-B. □

Note that the conditions of Theorem 2 can be interpreted as a flow decomposition for the layered network. If $R$ is the information that flows from the source to the destination, then the flow decomposition gives the effective amount of information that flows through each node. If the compression rate at each relay node is made approximately equal to the information flowing through that node, then the layered decoding where the destination ends up decoding the effective information at each node has a chance to work. Thus, in order to choose the right compression rates at each node, a flow decomposition for

the network must be obtained. These notions are made more precise in the next section.

## III. Flows With Bisubmodular Capacity Constraints

Maximum flow problems are extensively studied in graph theory and combinatorial optimization [10]. The problems are most often motivated from the study of transportation and communication networks. A directed graph $(\mathcal{V}, \mathcal{E})$ consists of the set of vertices or nodes $\mathcal{V}$ and the set of edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. Traditionally, flow is defined to be a non-negative function over the set of all edges which satisfy the *flow-conservation law* at each vertex other than the source and the destination node. Further, the flow over any edge is less than the capacity of that the edge. The classic max-flow min-cut result of [11] characterizes the maximum flow from the source to destination node and shows it to be equal to the min-cut of the graph. In order to distinguish from the concept of the node-flow that will be introduced here, such a flow is called an edge flow over an edge-capacitated graph. Beginning from the single commodity result of [11], various extensions of these problems have been considered. In particular, the edge-capacitated graph was extended to a polymatroidal network [12], where the flow is constrained not only by the edge-capacities but by joint capacities on sets of incoming and outgoing edges at every vertex. A special case is the node-capacitated graph [13], where the constraints on the flow are on the sum-total of the incoming and outgoing flow at each node.

In this section, the concept of a node-flow in the context of a layered graph with bisubmodular constraints on the flows is introduced. The node-flows can be related to the edge-flows with flow-conservation at the node. Note that the conservation law for edge-flow enforces that the net incoming flow at any node is equal to the net outgoing flow at the node and this quantity can be viewed as the node-flow for a node. The bisubmodular constraints can be viewed as generalizations of the polymatroidal constraints of [12]. The definitions here are motivated by the layered coding scheme for the wireless network, which was presented in the previous chapter. The main result is a max-flow min-cut theorem for the single-commodity node-flow for a graph with bisubmodular capacity constraints. The result is closely related to, and can be viewed as a generalization of, the flow introduced in the context of the linear deterministic networks and polylinking systems in [8] and [9].

### A. A Max-Flow Min-Cut Theorem

In this section, the max-flow min-cut theorem is proved for *single-commodity node-flow* on a *layered graph* with *bisubmodular capacity* constraints.

*Layered graph:* A layered graph is considered, which is represented by a set of nodes $\mathcal{V}$, which can be decomposed into subsets $\mathcal{O}_l, 1 \leq l \leq L$ as shown in Figure 2. The layering is ensured by the edges of the graph, which connect nodes in any layer $l$ to nodes in the subsequent layer $l+1$. Since the edges do not play any role in the problem here, beyond ensuring the layering, they will henceforth be neglected. The first layer $\mathcal{O}_1$ has a single node, which is the source node and the last layer $\mathcal{O}_L$ has a single node, which is the destination node.

*Bisubmodular Capacity Functions:* The bisubmodular capacity functions are defined for the layered graph using a family of $L-1$ functions $\{\rho_l : 1 \leq l \leq L-1\}$, $\rho_l : 2^{\mathcal{O}_l} \times 2^{\mathcal{O}_{l+1}} \to \mathbb{R}^+$, which satisfy the following properties:

1) $\rho_l$ is bisubmodular, i.e., $\forall U_1, U_2 \subseteq \mathcal{O}_l, V_1, V_2 \subseteq \mathcal{O}_{l+1}$,

$$\begin{aligned}&\rho_l(U_1 \cup U_2, V_1 \cap V_2) + \rho_l(U_1 \cap U_2, V_1 \cup V_2)\\ &\quad \leq \rho_l(U_1, V_1) + \rho_l(U_2, V_2).\end{aligned} \tag{32}$$

2) $\rho_l$ is non-decreasing, i.e.

$$\rho_l(U, V) \leq \rho_l(U_1, V_1), \text{ for } U \cup V \subseteq U_1 \cup V_1. \tag{33}$$

3) If $U = \emptyset$ or $V = \emptyset$, then

$$\rho_l(U, V) = 0. \tag{34}$$

*Node-flow:* The node-flow for the layered graph is defined as a function $f : \mathcal{V} \to \mathbb{R}^+$ which satisfies the capacity constraints, i.e.,

$$\begin{aligned}&f(V) - f(\mathcal{O}_l \backslash U) \leq \rho_l(U, V),\\ &\quad \forall\ U \subseteq \mathcal{O}_l,\ V \subseteq \mathcal{O}_{l+1}, \forall l \in [L-1],\end{aligned} \tag{35}$$

where $f(A)$ is an over-loaded notation, such that when $A \subseteq \mathcal{V}$ then $f(A) \stackrel{\text{def}}{=} \sum_{v \in A} f(v)$. Further, the destination node must sink the flow from the source. Therefore $f(D) = f(S)$.

The max-flow problem is to find the maximum $f(S)$ that can be supported given the capacity constraints on the graph. An efficient algorithm to compute the flow at each node given any $f(S)$ that can be supported is also sought.

An upper bound on the max-flow is given by the cut function.

*Cut function*: The cut function $C : 2^{\mathcal{V}} \to \mathbb{R}_+$ is defined as

$$C(\Omega) \stackrel{\text{def}}{=} \sum_{l=1}^{L-1} \rho_l(\Omega_l, \mathcal{O}_{l+1} \backslash \Omega_{l+1}), \tag{36}$$

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

Clearly,

$$\max f(S) \leq \min_{\Omega \subseteq \mathcal{V}} C(\Omega). \tag{37}$$

The next theorem shows that the min-cut is achievable. The proof is constructive and gives and efficient method of computing the flow.

*Theorem 4:*

$$\max f(S) = \min_{\Omega \subseteq \mathcal{V}} C(\Omega). \tag{38}$$

*Proof:* The proof is based on the polymatroid intersection theorem. The details are in Appendix B. □

The max-flow min-cut theorem for node-flows with bisubmodular constraints presented here is closely related to the max-flow min-cut results of [8] and [9]. [8] considered linear deterministic networks, which led to bisubmodular capacity functions arising from the rank of a matrix. [9] considered polylinking systems, where the bisubmodular capacity functions are given by the polylinking function. The results of

[9] generalized the results of [8] by showing that a linear deterministic network is a special case of polylinking system.

The max-flow min-cut theorem can be easily generalized to the following two cases:

- *Multi-source:* Consider a layered graph with $J$ source nodes in $\mathcal{O}_1$ and a single destination node in $\mathcal{O}_L$, such that $f(\mathcal{O}_1) = f(D)$. For this case, the following corollary generalizes Theorem 4.
  *Corollary 2:* $\{f(v)|v \in \mathcal{O}_1\}$ *is a feasible flow iff,*
  $$f(\mathbf{\Omega}_1) \leq C(\mathbf{\Omega}), \qquad \forall\ \mathbf{\Omega} \subseteq \mathcal{V}, \tag{39}$$
  *where* $\mathbf{\Omega}_1 \stackrel{\text{def}}{=} \mathbf{\Omega} \cap \mathcal{O}_1$.
- *Multi-destination:* Consider a layered graph with a single source node in $\mathcal{O}_1$ and J destination nodes in $\mathcal{O}_L$, such that $f(S) = f(\mathcal{O}_L)$. For this case, the following corollary generalizes Theorem 4.
  *Corollary 3:* $\{f(v)|v \in \mathcal{O}_L\}$ *is a feasible flow iff,*
  $$f(\mathbf{\Omega}_L) \leq C(\mathbf{\Omega}), \qquad \forall\ \mathbf{\Omega} \subseteq \mathcal{V}, \tag{40}$$
  *where* $\mathbf{\Omega}_L \stackrel{\text{def}}{=} \mathbf{\Omega} \cap \mathcal{O}_L$.

Note that the proof for the multiple sources (or destinations) case follows by adding a hypothetical supernode $A$ in layer 0 (or $L+1$) with capacity functions $\rho_0$ (or $\rho_L$) given by $\rho_0(A, V) = \sum f(v),\ \forall\ V \subseteq \mathcal{O}_1$ (or $\rho_L(V, A) = \sum f(v),\ \forall\ V \subseteq \mathcal{O}_L$).

### *B. Proof of Theorem 3: A Compress-and-Forward Scheme From Flows*

In this section, Theorem 3 is proved by establishing a connection between the compression rates of the compress-and-forward scheme with the layered decoding and the node-flows with bisubmodularity constraints. Recall that the achievable rates for the compress-and-forward with the layered decoding scheme are given by (26)–(28), which appear very much like the bisubmodular capacity constraints.

To make this connection more precise, first observe the following proposition.

*Proposition 1: Given the collection of random variables* $Q_p$ *distributed as given by* (12)*, the family of* $L-1$ *functions* $\rho_l : \mathcal{O}_l \times \mathcal{O}_{l+1} \to \mathbb{R}^+,\ \forall l \in [L-1]$ *defined by*

$$\rho_l(U, V) \stackrel{\text{def}}{=} I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \backslash U}) \tag{41}$$

*forms a family of bisubmodular capacity functions.*

*Proof:* Appendix D. □

For any $\mathbf{\Omega} \subseteq \mathcal{V}$, the corresponding cut value $C(\mathbf{\Omega})$ is now given by

$$C(\mathbf{\Omega}) = \sum_{l=1}^{L-1} I(X_{\mathbf{\Omega}_l}; \hat{Y}_{\mathcal{O}_{l+1} \backslash \mathbf{\Omega}_{l+1}} | X_{\mathcal{O}_l \backslash \mathbf{\Omega}_l}) \tag{42}$$

$$= I(\hat{Y}_{\mathbf{\Omega}_c}; X_{\mathbf{\Omega}} | X_{\mathbf{\Omega}^c}). \tag{43}$$

Theorem 4 is then used construct a flow $f(v)$ for this network, such that

$$f(S) \leq \min_{\mathbf{\Omega}} I(\hat{Y}_{\mathbf{\Omega}_c}; X_{\mathbf{\Omega}} | X_{\mathbf{\Omega}^c}), \qquad S \in \mathbf{\Omega}, D \in \mathbf{\Omega}^c, \tag{44}$$

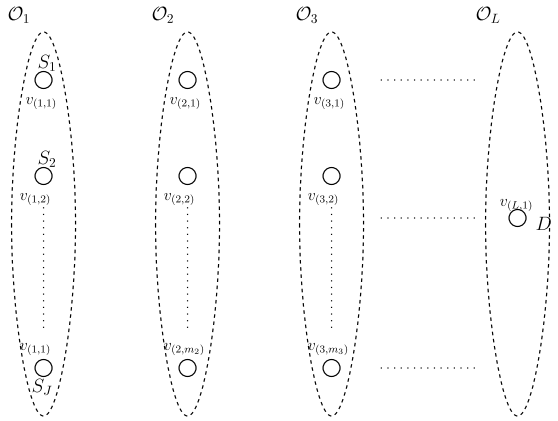


Fig. 4. A layered multi-source network.

and

$$f(V) - f(\mathcal{O}_l \backslash U) \leq \rho_l(U, V), \quad \forall\ U \subseteq \mathcal{O}_l,\ V \subseteq \mathcal{O}_{l+1}, \forall l \in [L-1]. \tag{45}$$

For any $v \in \mathcal{O}_l, l \in [L-1]$, let

$$r_v = f(v) - \kappa_l, \tag{46}$$

and $R = f(S) - \kappa_1$, where $\kappa_l$ is given by (31).

Then $\forall U \neq \emptyset \subseteq \mathcal{O}_l,\ V \subseteq \mathcal{O}_{l+1}$,

$$r(U) - r(\mathcal{O}_{l+1} \backslash V) = f(U) - f(\mathcal{O}_{l+1} \backslash V) - |U|\kappa_l + |\mathcal{O}_{l+1} \backslash V|\kappa_{l+1} \tag{47}$$

$$\leq \rho_l(U, V) - \kappa_l + |\mathcal{O}_{l+1}|\kappa_{l+1} \tag{48}$$

$$= \rho_l(U, V) - I(\hat{Y}_{\mathcal{O}_{l+1}}; Y_{\mathcal{O}_{l+1}} | X_{\mathcal{O}_l}) \tag{49}$$

$$\leq I(X_U; \hat{Y}_V | X_{\mathcal{O}_l \backslash U}) - I(\hat{Y}_{\mathcal{O}_{l+1} \backslash V}; Y_{\mathcal{O}_{l+1} \backslash V} | X_{\mathcal{O}_l}). \tag{50}$$

Therefore $\left(R, \{r_v\}_{v \in \mathcal{V}_r}\right)$ satisfies (26)–(28). And therefore Theorem 2 implies that the rate $R = f(S) - \kappa_1$ is achievable. This proves Theorem 3.

## IV. Generalizations to Multi-Source Networks

The communication network with multiple source nodes $\{S_i | i \in [J]\}$ is illustrated in Figure 4. The source node $S_i$ has independent message $W_i$ at rate $R_i$. There is a common destination node $D$. The multi-source relay network was perhaps first studied in [14], [15], where the rate region for the deterministic case and an approximate rate region for the Gaussian case were established. The noisy network coding scheme of [1] extends to this case as well. In fact this result was used for each layer to analyze the layered decoding scheme in the proof of Theorem 2.

The results of the compress-and-forward scheme and the layered decoding scheme can be generalized to the communication network with multiple source nodes and a common destination node.

The following corollary extends the results of the compress-and-forward scheme for the unicast network to the multi-source relay network.

*Theorem 5: The communication rates* $\vec{R} = (R_1, \ldots, R_J)$ *are achievable by the compress-and-forward scheme (with joint decoding) for the multi-source single destination network if, for some collection of random variables* $Q_p$ *which is distributed as* (12)*, the rates satisfy*

$$R(\Omega_1) < I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - I(\hat{Y}_\Omega; Y_\Omega | X_\mathcal{V}), \quad \forall\ \Omega,\ s.t.,\ \Omega \subseteq \mathcal{V}, D \in \Omega^c, \tag{51}$$

*where* $\Omega_1 \stackrel{\text{def}}{=} \Omega \cap \mathcal{O}_1$.

*Further, with the layered decoding scheme, the rates* $\vec{R} = (R_1, \ldots, R_J)$ *are achievable if*

$$R(\Omega_1) < I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - |\Omega_1|\kappa_1, \tag{52}$$

*where* $\kappa_1$ *is given by* (31)*.*

The results can be proved by adding a hypothetical supernode in layer 0, which is connected to the source nodes with orthogonal wired links such that the wired link to node $S_i$ is of rate $R_i$.

## V. Special Cases

### A. Wireless Network

For the special case of the Wireless network described by (11), the achievable rates can be compared to the cutset bound [16].

*Corollary 4: If* $\vec{R} = (R_1, \ldots, R_J)$ *is in the cutset bound, then rates* $\vec{R} - 3|\mathcal{V}|\vec{1}$ *are achievable by the compress-and-forward scheme (with joint decoding) for the multi-source single destination Gaussian network. Further, with the layered decoding scheme, the rates* $\vec{R} - (2|\mathcal{V}| + \kappa_1^g)\vec{1}$ *are achievable, where*

$$\kappa_l^g = |\mathcal{O}_{l+1}|(\kappa_{l+1}^g + 1), \tag{53}$$

*and* $\kappa_{L-1}^g = 0$*.*

*Proof:* To characterize the achievable rate we will use a joint Gaussian $Q_p$. Let $X_\mathcal{V} \sim \mathcal{CN}(0, I)$. As also noted in [1], a good choice for $\hat{Y}_v$ for the Gaussian network is given by

$$\hat{Y}_v = Y_v + \hat{Z}_v, \tag{54}$$

where $\hat{Z}_v \sim \mathcal{CN}(0, 1)$ is independent across nodes.

The particular choice of $\hat{Y}_v$ implies that the quantization is done at the noise level. This also agrees with the philosophy in [2] and [4], where the quantization was done at the noise level to show approximate optimality; in [2], scalar quantization was done at the noise level, and in [4], quantization was done using the discrete superposition network, which was a model obtained from the wireless network by clipping the signal at the noise level.

As shown in [1], with this choice of $\hat{Y}_v$ and with $X_\mathcal{V} \sim \mathcal{CN}(0, I)$,

$$I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) = \log \left| I + \frac{H_{\Omega\Omega^c} H_{\Omega\Omega^c}^*}{2} \right| \tag{55}$$

$$\geq \log \left| I + H_{\Omega\Omega^c} H_{\Omega\Omega^c}^* \right| - \frac{|\Omega^c|}{2}. \tag{56}$$

And further,

$$I(\hat{Y}_v; Y_v | X_\mathcal{V}) \leq 1, \tag{57}$$

which implies

$$I(\hat{Y}_\Omega; Y_\Omega | X_\mathcal{V}) \leq |\Omega|. \tag{58}$$

This allows us to lower bound the RHS of (51) as follows,

$$I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) - I(\hat{Y}_\Omega; Y_\Omega | X_\mathcal{V}) > \log \left| I + H_{\Omega\Omega^c} H_{\Omega\Omega^c}^* \right| - |\mathcal{V}|. \tag{59}$$

Lemma 6.6 in [2] lower bounds the quantity $\log \left| I + H_{\Omega\Omega^c} H_{\Omega\Omega^c}^* \right|$ to within $2|\mathcal{V}|$ of the cut-set bound. The corollary then follows from Theorem 5. □

*Remark 1: We showed that the layered decoding scheme is approximately optimal like the joint-decoding scheme, i. e. the gap from capacity can be bounded by a constant, which only depends on the size of the network and not the channel characteristics or power. However, in general, our bounds on the gap from capacity with the layered-decoding scheme can be much larger than the bounds on the gap from capacity for the joint-decoding scheme. For example, consider a layered network with L-layer of relay nodes each with n relay nodes in each layer. With joint decoding scheme, the gap is* $\mathrm{O}(nL)$*. With layered decoding, the gap is dominated by* $\kappa_1^g = \mathrm{O}\left(n^L\right)$*.*

### B. Deterministic Network

For the special case of the deterministic network described by (9), the optimal choice of $\hat{Y}_v$ is $Y_v$ and with this choice

$$I(\hat{Y}_{\Omega^c}; X_\Omega | X_{\Omega^c}) = H(\hat{Y}_{\Omega^c} | X_{\Omega^c}). \tag{60}$$

And further,

$$I(\hat{Y}_v; Y_v | X_\mathcal{V}) = 0. \tag{61}$$

Therefore, specializing the results of Theorem 5 leads to the following corollary.

*Corollary 5: For the multi-source single-destination deterministic network,* $\vec{R} = (R_1, \ldots, R_J)$ *is achievable by the compress-and-forward scheme with the layered decoding scheme if for some collection of random variables* $Q_p$ *which is distributed as* (12)*,*

$$\vec{R} \in \bar{\mathcal{C}}(Q_p), \tag{62}$$

*where* $\bar{\mathcal{C}}(Q_p)$ *is the cutset bound evaluated under the product distribution for the network [2].*

Specializing further to the linear deterministic region, it can be shown that the product distribution (with uniformly distributed $X_v$ over all input alphabets) maximizes the cutset bound, thereby showing that all rates in the cutset bound are achievable.

## VI. Discussion: Layered Decoding vs. Joint Decoding

As mentioned previously, layered decoding could reduce the decoding complexity as compared to joint decoding. This advantage comes in at a potential decrease in the rate that can be achieved. The rate achieved by layered decoding is, in general, always lesser than the rate achieved by the joint-decoding.

The comparison of the decoding complexity between the joint decoding and the layered decoding will be done with

respect to an exhaustive search ML decoder. In practical implementation a more structured codebook is sought, which simplifies the ML decoding complexity. For example, in [17], the ML decoder is implemented for a simple one-relay network with binary LDPC codes and a reduced quantizer operation. The ML decoding can be reduced to belief-propagation over a large Tanner graph, which comprises the Tanner graphs of the LDPC codes for each node, the quantization and mapping operation, and the network itself. If this simplified encoding scheme is extended to a network with multiple layers of relay nodes, the resulting graph would be humongous, and the decoding complexity would be large. It is very much possible that the layered decoding will help to reduce this complexity by breaking the joint decoding over a large graph, into decoding over smaller sub-graphs corresponding to each layer, thereby reducing the complexity.

Assuming the maximum likelihood (ML) decoding is done by an exhaustive search as given by (15), the decoding complexity of the joint decoding is the product of the codebooks of all the nodes. Therefore the complexity of the joint-decoding is given by

$$\mathfrak{C}_{\text{joint}} = 2^{RT} \prod_{v \in \mathcal{V}_r} n_{Q,v}, \tag{63}$$

where $n_{Q,v}$ is the number of quantization points in the relay quantization codebook. With the compress-and-forward scheme with the layered decoding, the complexity is reduced to

$$\mathfrak{C}_{\text{layered}} = \sum_{l=1}^{L-1} 2^{r(\mathcal{O}_l)T} \prod_{v \in \mathcal{O}_{l+1}} n_{Q,v}. \tag{64}$$

## VII. Conclusion

In this paper, the compress-and-forward scheme is analyzed for the unicast relay network. It is shown that while it achieves the same overall rate as NNC, it allows for a lower complexity layered/backward decoding algorithm. However, this requires each relay node to compress their information to the right amount. This paper also presents a computationally efficient way of finding the optimal compression rates at each relay node using a node-flow formulation over a bisubmodular constrained graph.

## Appendix A
## Probability of Error Analysis for CF Scheme

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

$$\left(x_s^T(w), \left\{\hat{y}_{(l,k)}^T(w_{(l,k)}, \bar{w}_{(l,k)}), x_{(l,k)}^T(w_{(l,k)})\right\}_{(l,k)\in\mathcal{V}_r}, y_d^T\right) \in \mathcal{T}_\epsilon^T. \tag{65}$$

Here $(w, \bar{w})_{\mathcal{V}_r}$ is shorthand for $\{(w_v, \bar{w}_v)\,|v \in \mathcal{V}_r\}$. The error event is the union of two terms and is given by

$$\left(\bigcup_{w_{\mathcal{V}_r}, \bar{w}_{\mathcal{V}_r}} \mathcal{E}_{1,(w,\bar{w})_{\mathcal{V}_r}}\right)^c \bigcup \left(\bigcup_{w\neq 1, w_{\mathcal{V}_r}, \bar{w}_{\mathcal{V}_r}} \mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}\right). \tag{66}$$

The first term corresponds to the event that the transmitted message is not jointly typical and the second term corresponds to some other message other than the transmitted being jointly typical. The first event can be upper bounded by $\mathcal{E}^c_{1,(1,1)_{\mathcal{V}_r}}$. For any $\Omega \subseteq \mathcal{V}_r$, and $\Phi \subseteq \mathcal{V}_r \backslash \Omega$, let

$$\begin{aligned}\mathfrak{S}_{\Omega,\Phi} \stackrel{\text{def}}{=} \big\{&(w, (w, \bar{w})_{\mathcal{V}_r})| \\ &w \neq 1, \\ &w_{(l,k)} \neq 1 \forall (l,k) \in \Omega, \\ &w_{(l,k)} = 1, \bar{w}_{(l,k)} \neq 1 \forall (l,k) \in \Omega^c \backslash \Phi, \\ &w_{(l,k)} = 1, \bar{w}_{(l,k)} = 1 \forall (l,k) \in \Phi\big\},\end{aligned} \tag{67}$$

and

$$\mathcal{E}_{\Omega,\Phi} \stackrel{\text{def}}{=} \bigcup_{\mathfrak{S}_{\Omega,\Phi}} \mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}. \tag{68}$$

The second event can be equivalently written as,

$$\left(\bigcup_{w\neq 1, w_{\mathcal{V}_r}, \bar{w}_{\mathcal{V}_r}} \mathcal{E}_{w,(w,\bar{w})_{\mathcal{V}_r}}\right) = \bigcup_{\Omega,\Phi} \mathcal{E}_{\Omega,\Phi}, \tag{69}$$

The probability or error by union bound can be upper bounded by,

$$\mathbb{P}(error) \leq \mathbb{P}(\mathcal{E}^c_{1,(1,1)_{\mathcal{V}_r}}) + \sum_{\Omega,\Phi} \mathbb{P}\left(\mathcal{E}_{\Omega,\Phi}\right). \tag{70}$$

From the properties of joint typicality, it can be shown that the first term goes to 0 and $T \to \infty$. It can be shown that

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_{\Omega,\Phi}\right) &\doteq 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} 2^{T(H(Y_d,\hat{Y}_\Phi,\hat{Y}_{\Phi^c},X_\Omega,X_{\Omega^c},X_s))} \\ &\quad 2^{-T(H(X_\Omega,X_s)+H(Y_d,\hat{Y}_\Phi,X_{\Omega^c})+\sum_{(l,k)\in\Phi^c} H(\hat{Y}_{(l,k)}))} \qquad (71) \\ &= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} 2^{T(H(Y_d,\hat{Y}_\Phi,\hat{Y}_{\Phi^c},X_\Omega,X_{\Omega^c},X_s))} \\ &\quad 2^{-T(H(Y_d,\hat{Y}_\Phi|X_{\Omega^c})+\sum_{(l,k)\in\Phi^c} H(\hat{Y}_{(l,k)}))} \qquad (72) \\ &= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} \\ &\quad 2^{-T(H(Y_d,\hat{Y}_\Phi|X_{\Omega^c})-H(Y_d,\hat{Y}_\Phi|X_\Omega,X_{\Omega^c},X_s))} \\ &\quad 2^{-T(\sum_{(l,k)\in\Phi^c} H(\hat{Y}_{(l,k)})-H(\hat{Y}_{\Phi^c}|X_\Omega,X_{\Omega^c},X_s))} \qquad (73) \\ &= 2^{T(R+r(\Omega)+\bar{r}(\Phi^c))} 2^{-T(I(Y_d,\hat{Y}_\Phi;X_\Omega,X_s|X_{\Omega^c}))} \\ &\quad 2^{-T(\sum_{(l,k)\in\Phi^c} I(\hat{Y}_{(l,k)};X_{\mathcal{V}_r},X_s))}. \qquad (74)\end{aligned}$$

Here $r(A) \stackrel{\text{def}}{=} \sum_{v\in A} r_v$. Using the Markovian property of the random variables, we have that

$$I(\hat{Y}_{(l,k)}; X_{\mathcal{V}_r}, X_s) = I(\hat{Y}_{(l,k)}; Y_{(l,k)}) - I(\hat{Y}_{(l,k)}; Y_{(l,k)}|X_{\mathcal{V}_r}, X_s), \tag{75}$$

and using (14) we have

$$\begin{aligned}&\mathbb{P}\left(\mathcal{E}_{\Omega,\Phi}\right) \\ &= 2^{T(R-r(\Omega^c\backslash\Phi)-I(Y_d,\hat{Y}_\Phi;X_\Omega,X_s|X_{\Omega^c})+I(\hat{Y}_{\Phi^c};Y_{\Phi^c}|X_{\mathcal{V}_r},X_s))}. \qquad (76)\end{aligned}$$

Therefore $\mathbb{P}\left(\mathcal{E}_{\Omega,\Phi}\right) \to 0$, if

$$R < r(\Omega^c \backslash \Phi) + I(Y_d, \hat{Y}_\Phi; X_\Omega, X_s | X_{\Omega^c}) - I(\hat{Y}_{\Phi^c}; Y_{\Phi^c} | X_{\mathcal{V}_r}, X_s). \quad (77)$$

## Appendix B
## Proof of Theorem 4

The theorem will be proved in a slightly general setting, allowing multiple nodes in layer $\mathcal{O}_1$ and layer $\mathcal{O}_L$. Assuming that the flow values for these layers $\mathcal{O}_1$ and $\mathcal{O}_L$ are given and satisfy

$$f(\mathcal{O}_1) = f(\mathcal{O}_L), \quad (78)$$

$$f(\Omega_1) - f(\Omega_L) \leq C(\Omega), \quad \forall\ \Omega \subseteq \mathcal{V}, \quad (79)$$

the flow for all intermediate layers will be constructed.

The proof is by inductive construction.

For $L = 2$, there are no intermediate layers and the theorem holds by definition. Consider $L > 2$. The induction hypothesis assumes that the flow can be constructed with fewer than $L$ layers and the flow for the boundary layers are specified with the constraints given by (79).

Consider any $L_0 \in \{2, \ldots, L-1\}$. 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.

Next, a flow for the layer $\mathcal{O}_{L_0}$ will be constructed which satisfies the following conditions.

$$f(\mathcal{O}_{L_0}) = f(\mathcal{O}_1), \quad (80)$$

$$f(\Omega_A \cap \mathcal{O}_1) - f(\Omega_A \cap \mathcal{O}_{L_0}) \leq C_A(\Omega_A), \quad \forall\ \Omega_A \subseteq \mathcal{V}_A, \text{ and} \quad (81)$$

$$f(\Omega_B \cap \mathcal{O}_{L_0}) - f(\Omega_B \cap \mathcal{O}_L) \leq C_B(\Omega_B), \quad \forall\ \Omega_B \subseteq \mathcal{V}_B. \quad (82)$$

The induction hypothesis would then guarantee that the flows for the intermediate layers in the sub-networks $\mathcal{N}_A$ and $\mathcal{N}_B$ can be constructed.

Using (80), the set of linear inequalities given by (81) can be written as,

$$f(\Omega_A^c \cap \mathcal{O}_{L_0}) - f(\Omega_A^c \cap \mathcal{O}_1) \leq C_A(\Omega_A), \quad \forall\ \Omega_A \subseteq \mathcal{V}_A, \quad (83)$$

where $\Omega_A^c = \mathcal{V}_A \backslash \Omega_A$. For any fixed $T \subseteq \mathcal{O}_{L_0}$, the collection of inequalities where $\Omega_A^c \cap \mathcal{O}_{L_0} = T$, can be concisely represented as,

$$f(T) \leq \min\left\{C_A(\Omega_A) + f(\Omega_A^c \cap \mathcal{O}_1) : \Omega_A^c \cap \mathcal{O}_{L_0} = T\right\}. \quad (84)$$

Defining

$$r_A(T) \stackrel{\text{def}}{=} \min\left\{C_A(\Omega_A) + f(\Omega_A^c \cap \mathcal{O}_1) : \Omega_A^c \cap \mathcal{O}_{L_0} = T\right\}, \quad (85)$$

the set of linear inequalities given by (81) can be concisely written as,

$$f(T) \leq r_A(T), \quad \forall\ T \subseteq \mathcal{O}_{L_0}. \quad (86)$$

Similary, defining

$$r_B(T) \stackrel{\text{def}}{=} \min\left\{C_B(\Omega_B) + f(\Omega_B \cap \mathcal{O}_L) : \Omega_B \cap \mathcal{O}_{L_0} = T\right\}, \quad (87)$$

the set of linear inequalities given by (82) can be concisely written as,

$$f(T) \leq r_B(T), \quad \forall\ T \subseteq \mathcal{O}_{L_0}. \quad (88)$$

The following properties for the functions $r_A(T)$ and $r_B(T)$ can be established.

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

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

*Proof:* Appendix C. □

Define the following polymatroids with the functions $r_A$ and $r_B$.

$$P_A = \left\{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_A(U),\ \forall\ U \in \mathcal{O}_{L_0}\right\} \quad (89)$$

$$P_B = \left\{\mathbf{x} \in \mathbb{R}_+^{m_{L_0}} : x(U) \leq r_B(U),\ \forall\ U \in \mathcal{O}_{L_0}\right\}, \quad (90)$$

where $\mathbf{x} = [x(1) \ldots x(m_{L_0})]$ and $x(U) \stackrel{\text{def}}{=} \sum_{u \in U} x(u)$. The conditions (80)–(82) are now equivalent to finding

$$[f(L_0, 1) \ldots f(L_0, m_{L_0})] \in P_A \cap P_B, \quad (91)$$

such that $f(\mathcal{O}_{L_0}) = f(\mathcal{O}_1)$. It then follows from Edmond's polymatroid intersection ( [10], Corollary 46.1c) that:

$$\max\left\{x(\mathcal{O}_{L_0}) : \mathbf{x} \in P_A \cap P_B\right\} = \min_{T \subseteq \mathcal{O}_{L_0}} \left\{r_A(\mathcal{O}_{L_0} \backslash T) + r_B(T)\right\}. \quad (92)$$

Therefore the required flow exists since

$$f(\mathcal{O}_1) \leq \min_{T \subseteq \mathcal{O}_{L_0}} \left\{r_A(\mathcal{O}_{L_0} \backslash T) + r_B(T)\right\} \quad (93)$$

$$= \min_{\Omega \in \mathcal{V}} \left\{C(\Omega) + f(\mathcal{O}_1 \backslash \Omega_1) + f(\Omega_L)\right\}. \quad (94)$$

Further, in [10, Th. 47.1] it is shown that the maximizing $\mathbf{x}$ in (92) can be computed in *polynomial* time in the dimension of $\mathbf{x}$. Hence, the flow can also be computed in polynomial time in the number of nodes.

## Appendix C
## Proof of Lemma 1

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

1) Submodularity:
Let,

$$r_B(T^{(1)}) = C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \quad \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)} \quad (95)$$

$$r_B(T^{(2)}) = C_B(\Omega_B^{(2)}) + d(\Omega_B^{(2)} \cap \mathcal{O}_L), \quad \Omega_B^{(2)} \cap \mathcal{O}_{L_0} = T^{(1)}. \quad (96)$$

Since,

$$(\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cup T^{(2)}, \quad (97)$$

$$(\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_{L_0} = T^{(1)} \cap T^{(2)}, \quad (98)$$

it follows that

$$r_B(T^{(1)} \cup T^{(2)}) \leq C_B(\Omega_B^{(1)} \cup \Omega_B^{(2)}) + d((\Omega_B^{(1)} \cup \Omega_B^{(2)}) \cap \mathcal{O}_L), \quad (99)$$

$$r_B(T^{(1)} \cap T^{(2)}) \leq C_B(\Omega_B^{(1)} \cap \Omega_B^{(2)}) + d((\Omega_B^{(1)} \cap \Omega_B^{(2)}) \cap \mathcal{O}_L). \quad (100)$$

By definition of cut and the bi-submodularity of $\rho_l$, 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 sub modular.

2) Non-decreasing:
Consider $T^{(1)} \subseteq T^{(2)}$. Let

$$r_B(T^{(1)}) = C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L), \quad \Omega_B^{(1)} \cap \mathcal{O}_{L_0} = T^{(1)}. \quad (101)$$

Let $\Omega_B = \Omega_B^{(1)} \cup T^{(2)} \backslash T^{(1)} \supseteq \Omega_B^{(1)}$, so that $\Omega_B \cap \mathcal{O}_{L_0} = T^{(2)}$. By the definition of cut and the non-decreasing property of $\rho_l$, 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

$$r_B(T^{(2)}) = C_B(\Omega_B) + d(\Omega_B \cap \mathcal{O}_L) \quad (102)$$

$$\geq C_B(\Omega_B^{(1)}) + d(\Omega_B^{(1)} \cap \mathcal{O}_L) \quad (103)$$

$$= r_B(T^{(1)}). \quad (104)$$

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

## Appendix D
## Proof of Proposition 1

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

$$I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \backslash U}) = H(\hat{Y}_W | X_{\mathcal{O}_l \backslash U}) - H(\hat{Y}_W | X_{\mathcal{O}_l}) \quad (105)$$

$$= H(\hat{Y}_W, X_{\mathcal{O}_l \backslash U}) - H(X_{\mathcal{O}_l \backslash U}) - H(\hat{Y}_W | X_{\mathcal{O}_l}). \quad (106)$$

The submodularity of entropy [18] implies that $H(\hat{Y}_W, X_{\mathcal{O}_l \backslash U})$ is bi-submodular.

The submodularity of entropy follows from the fact that given collection of random variables $\Upsilon_1$ and $\Upsilon_2$, we have

$$H(\Upsilon_1) + H(\Upsilon_2) - H(\Upsilon_1 \cup \Upsilon_2) - H(\Upsilon_1 \cap \Upsilon_2) = I(\Upsilon_1 \backslash \Upsilon_2; \Upsilon_2 \backslash \Upsilon_1 | \Upsilon_1 \cap \Upsilon_2) \quad (107)$$

$$\geq 0. \quad (108)$$

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

$$I(X_U; \hat{Y}_W | X_{\mathcal{O}_l \backslash U}) = H(X_U | X_{\mathcal{O}_l \backslash U}) - H(X_U | X_{\mathcal{O}_l \backslash U} \hat{Y}_W) \quad (109)$$

$$\geq H(X_U | X_{\mathcal{O}_l \backslash U}) - H(X_U | X_{\mathcal{O}_l \backslash U} \hat{Y}_{W_1}) \quad (110)$$

$$= I(X_U; \hat{Y}_{W_1} | X_{\mathcal{O}_l \backslash U}) \quad (111)$$

$$= H(\hat{Y}_{W_1} | X_{\mathcal{O}_l \backslash U}) - H(\hat{Y}_{W_1} | X_{\mathcal{O}_l}) \quad (112)$$

$$\geq H(\hat{Y}_{W_1} | X_{\mathcal{O}_l \backslash U_1}) - H(\hat{Y}_{W_1} | X_{\mathcal{O}_l}) \quad (113)$$

$$= I(X_{U_1}; \hat{Y}_{W_1} | X_{\mathcal{O}_l \backslash U_1}), \quad (114)$$

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

The third property is readily seen.

## Acknowledgments

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

## References

[1] S. Lim, Y.-H. Kim, A. E. Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[2] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.

[3] A. Ozgur and S. N. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 669–673.

[4] M. Anand and P. R. Kumar, "A digital interface for Gaussian relay and interference networks: Lifting codes from the discrete superposition model," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2548–2564, May 2011.

[5] X. Wu and L.-L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, May 2013.

[6] A. Amadruz and C. Fragouli, "Combinatorial algorithms for wireless information flow," in *Proc. 20th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*, Philadelphia, PA, USA, Jan. 2009, pp. 555–564.

[7] J. B. Ebrahimi and C. Fragouli, "Combinatiorial algorithms for wireless information flow," *ACM Trans. Algorithms (TALG)*, vol. 9, no. 1, p. 8, 2012.

[8] S. M. S. Yazdi and S. A. Savari, "A max-flow/min-cut algorithm for linear deterministic relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3005–3015, May 2011.

[9] M. X. Goemans, S. Iwata, and R. Zenklusen, "A flow model based on polylinking system," *Math. Program.*, vol. 135, pp. 1–2, Oct. 2011.

[10] A. Schrivjer, *Combinatorial Optimization*. New York, NY, USA: Springer-Verlag, 2003.

[11] L. R. Ford, Jr. and D. R. Fulkerson, "Maximal flow through a network," *Can. J. Math.*, vol. 8, pp. 399–404, Jan. 1956.

[12] E. L. Lawler and C. U. Martel, "Computing maximal 'Polymatroidal' network flows," *Math. Oper. Res.*, vol. 7, no. 3, pp. 334–347, Aug. 1982.

[13] U. Feige, M. T. Hajiaghayi, and J. R. Lee, "Improved approximation algorithms for minimum weight vertex separators," *SIAM J. Comput.*, vol. 38, no. 2, pp. 629–657, May 2008.

[14] E. Perron, "Information-theoretic secrecy for wireless networks," Ph.D. dissertation, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, Aug. 2009.

[15] E. Perron, S. N. Diggavi, and I. E. Telatar, "On noise insertion strategies for wireless network secrecy," in *Proc. Inform. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2009, pp. 77–84.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Telecommunications and Signal Processing). New York, NY, USA: Wiley, 2006.

[17] V. Nagpal, I.-H. Wang, M. Joragovanovic, D. N. C. Tse, and B. Nikolić, "Quantize-map-and-forward relaying: Coding and system design," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2010, pp. 443–450.

[18] A. K. Kelmans and B. N. Kimelfeld, "Multiplicative submodularity of a matrix's principal minor as a function of the set of its rows and some combinatorial applications," *Discrete Math.*, vol. 44, no. 1, pp. 113–116, 1980.

**Adnan Raja**, biography not available at the time of publication.

**Pramod Viswanath**, biography not available at the time of publication.