Wireless Network Information Flow: A Deterministic Approach

A. Salman Avestimehr, *Member, IEEE*, Suhas N. Diggavi, *Member, IEEE*, and David N. C. Tse, *Fellow, IEEE*

***Abstract*—In a wireless network with a single source and a single destination and an arbitrary number of relay nodes, what is the maximum rate of information flow achievable? We make progress on this long standing problem through a two-step approach. First, we propose a deterministic channel model which captures the key wireless properties of signal strength, broadcast and superposition. We obtain an exact characterization of the capacity of a network with nodes connected by such deterministic channels. This result is a natural generalization of the celebrated max-flow min-cut theorem for wired networks. Second, we use the insights obtained from the deterministic analysis to design a new *quantize-map-and-forward* scheme for Gaussian networks. In this scheme, each relay quantizes the received signal at the noise level and maps it to a random Gaussian codeword for forwarding, and the final destination decodes the source's message based on the received signal. We show that, in contrast to existing schemes, this scheme can achieve the cut-set upper bound to within a gap which is independent of the channel parameters. In the case of the relay channel with a single relay as well as the two-relay Gaussian diamond network, the gap is 1 bit/s/Hz. Moreover, the scheme is universal in the sense that the relays need no knowledge of the values of the channel parameters to (approximately) achieve the rate supportable by the network. We also present extensions of the results to multicast networks, half-duplex networks, and ergodic networks.**

***Index Terms*—Information flow, network capacity, network information theory, relay networks, wireless networks.**

I. Introduction

TWO main distinguishing features of wireless communication are:

- *broadcast*: wireless nodes communicate over the air and signals from any one transmitter are heard by multiple nodes with possibly different signal strengths.
- *superposition*: a wireless node receives signals from multiple simultaneously transmitting nodes, with the received signals all superimposed on top of each other.

Because of these effects, links in a wireless network are never isolated but instead interact in seemingly complex ways. On the one hand, this facilitates the spread of information among users in a network; on the other hand it can be harmful by potentially creating signal interference among users. This is in direct contrast to wired networks, where transmitter-receiver pairs can be thought of as isolated point-to-point links. Starting from the max-flow-min-cut theorem of Ford–Fulkerson [1], there has been significant progress in understanding information flow over wired networks. Much less, however, is known for wireless networks.

The linear additive Gaussian channel model is a commonly used model to capture signal interactions in wireless channels. Over the past couple of decades, capacity study of Gaussian networks has been an active area of research. However, due to the complexity of the Gaussian model, except for the simplest networks such as the one-to-many Gaussian broadcast channel and the many-to-one Gaussian multiple access channel, the capacity of most Gaussian networks is still unknown. For example, even the capacity of a Gaussian single-relay network, in which a point to point communication is assisted by one relay, has been open for more than 30 years. In order to make progress on this problem, we take a two-step approach. We first focus on the signal interaction in wireless networks rather than on the noise. We present a new deterministic channel model which is analytically simpler than the Gaussian model but yet still captures three key features of wireless communication: channel strength, broadcast. and superposition. A motivation to study such a model is that in contrast to point-to-point channels where noise is the only source of uncertainty, networks often operate in the *interference-limited* regime where the noise power is small compared to signal powers. Therefore, for a first level of understanding, our focus is on such signal interactions rather than the background noise. Like the Gaussian model, our deterministic model is linear, but unlike the Gaussian model, operations are on a finite-field (the simplicity of *scalar* finite-field channel models has also been noted in [2]). We provide a complete characterization of the capacity of a network of nodes connected by such deterministic channels. This result is a natural generalization of the max-flow min-cut theorem for wired networks.

The second step is to utilize the insights from the deterministic analysis to find "approximately optimal" communication schemes for Gaussian relay networks. The analysis for deterministic networks not only gives us insights for potentially successful coding schemes for the Gaussian case, but also gives tools for the proof techniques used. We show that in

Manuscript received July 27, 2009; revised June 02, 2010; accepted August 31, 2010. Date of current version March 16, 2011. D. Tse and A. Avestimehr were supported in part by the National Science Foundation under Grants 0326503, 0722032, and 0830796, and in part by a gift from Qualcomm, Inc. S. Diggavi was supported in part by the Swiss National Science Foundation NCCR-MICS Center.

A. S. Avestimehr is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: avestimehr@ece.cornell.edu).

S. N. Diggavi is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: suhas@ee.ucla.edu).

D. N. C. Tse is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

Communicated by M. Franceschetti, 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.2011.2110110

Gaussian networks, an *approximate* max-flow min-cut result can be shown, where the capacity approximation is within an additive constant which is universal over the values of the channel parameters (but could depend on the number of nodes in the network). For example, the additive gap for both the single-relay network and for the two-relay diamond network is 1 bit/s/Hz. This is the first result we are aware of that provides such performance guarantees on relaying schemes. To highlight the strength of this result, we demonstrate that none of the existing strategies in the literature, like amplify-and-forward, decode-and-forward and Gaussian compress-and-forward, yield such a universal approximation for arbitrary networks. Instead, a scheme, which we term *quantize-map-and-forward*, provides such a universal approximation.

In this paper we focus on unicast and multicast communication scenarios. In the unicast scenario, one source wants to communicate to a single destination. In the multicast scenario, one source wants to transmit the *same* message to multiple destinations. Since in these scenarios, all destination nodes are interested in the same message, there is effectively only one information stream in the network. Due to the broadcast nature of the wireless medium, multiple copies of a transmitted signal *are* received at different relays and superimposed with other received signals. However, since they are all a function of the same message, they are not considered as interference. In fact, the quantize-map-and-forward strategy exploits this broadcast nature by forwarding all the available information received at the various relays to the final destination. This is in contrast to more classical approaches of dealing with simultaneous transmissions by either avoiding them through transmit scheduling or treating signals from all nodes other than the intended transmitter as interference adding to the noise floor. These approaches attempt to convert the wireless network into a wired network but are strictly sub-optimal.

A. Related Work

In the literature, there has been extensive research over the last three decades to characterize the capacity of relay networks. The single-relay channel was first introduced in 1971 by van der Meulen [3] and the most general strategies for this network were developed by Cover and El Gamal [4]. There has also been a significant effort to generalize these ideas to arbitrary multirelay networks with simple channel models. An early attempt was done in the Ph.D. thesis of Aref [5] where a max-flow min-cut result was established to characterize the unicast capacity of a deterministic broadcast relay network *without superposition*. This was an early precursor to network coding which established the multicast capacity of wired networks, a deterministic capacitated graph without broadcast or superposition [6]–[8]. These two ideas were combined in [9], which established a max-flow min-cut characterization for multicast flows for "Aref networks". However, such complete characterizations are not known for arbitrary (even deterministic) networks with both broadcast and superposition. One notable exception is the work [10] which takes a scalar deterministic linear finite-field model and uses probabilistic erasures to model channel failures. For this model using results of erasure broadcast networks [11], they established an asymptotic result on the unicast capacity as the field size grows. However, in all these works, there is no connection between the proposed channel model and the physical wireless channel.

There has also been a rich body of literature in directly tackling the noisy relay network capacity problem. In [12] the "diamond" network of parallel relay channels with no direct link between the source and the destination was examined. Xie and Kumar generalized the decode-forward encoding scheme for a network of multiple relays [13]. Kramer *et al.* [14] also generalized the compress-forward strategy to networks with a single layer of relay nodes. Though there have been many interesting and important ideas developed in these papers, the capacity characterization of Gaussian relay networks is still unresolved. In fact even a performance guarantee, such as establishing how far these schemes are from an upper bound is unknown. In fact, as we will see in Section III, these strategies do not yield an approximation guarantee for general networks.

There are subtle but critical differences between the quantize-map-forward strategy, proposed in this paper, with the natural extension of compress-forward to networks for the following reasons. The compress-forward scheme proposed in [4] quantized the received signal and then mapped the digital bin index onto the transmit sequence. This means that we need to make choices on the binning rates at each relay node. However, the quantize-map-forward scheme proposed in this paper directly maps the the quantized sequence to the transmit sequence, and therefore, does not make such choices on the binning rates. In fact this gives the scheme a "universality" property, which allows the same relay operation to work for multiple destinations (multicast) and network situations (compound networks); a property that could fail to hold if specific choices of binning rates were made. Moreover, our scheme, unlike the classical compress-forward scheme, does not require the quantized values at the relays to be reconstructed at the destination while it is attempting to decode the transmitted message. These are the essential differences between our scheme and the traditional compress-forward, or the natural network generalization of it.

Our results are connected to the concept of network coding in several ways. The most direct connection is that our results on the multicast capacity of deterministic networks are direct generalizations of network coding results [6]–[8], [15], [16] as well as Aref networks [5], [9]. The coding techniques for the deterministic case are inspired by and generalize the random network coding technique of [6] and the linear coding technique of [7], [8], [17]. The quantize-map-and-forward technique proposed in this paper for the Gaussian wireless networks uses the insights from the deterministic framework and is philosophically the network coding technique generalized to noisy wireless networks.

B. Outline of the Paper

We first develop an analytically simple linear finite-field model and motivate it by connecting it to the Gaussian model in the context of several simple multiuser networks. We also discuss its limitations. This is done in Section II. This model also suggests achievable strategies to explore in Gaussian relay networks, as done in Section III, where we illustrate the deterministic approach on several progressively more complex example networks. The deterministic model also makes clear

that several well-known strategies can be in fact arbitrarily far away from optimality in these example networks.

Section IV summarizes the main results of the paper. Section V focuses on the capacity analysis of networks with nodes connected by deterministic channels. We examine arbitrary deterministic channel model (not necessarily linear nor finite-field) and establish an achievable rate for an arbitrary network. For the special case of linear finite-field deterministic models, this achievable rate matches the cut-set bound; therefore, exact characterization is possible. The achievable strategy involves each node randomly mapping the received signal to a transmitted signal, and the final destination solving for the information bits from all the received equations.

The examination of the deterministic relay network motivates the introduction of a simple *quantize-map-and-forward* strategy for general Gaussian relay networks. In this scheme each relay first quantizes the received signal at the noise level, then randomly maps it to a Gaussian codeword and transmits it. In Section VI, we use the insights of the deterministic result to demonstrate that we can achieve a rate that is guaranteed to be within a constant gap from the cut-set upper bound on capacity. As a byproduct, we show in Section VII that a deterministic model formed by quantizing the received signals at noise level at all nodes and then removing the noise is within a constant gap to the capacity of the Gaussian relay network.

In Section VIII, we show that the quantize-map-and-forward scheme has the desirable property that the relay nodes do not need the knowledge of the channel gains. As long as the network can support a given rate, we can achieve it without the relays' knowledge of the channel gains. In Section VIII, we also establish several other extensions to our results, such as relay networks with half-duplex constraints, and relay networks with fading or frequency selective channels.

II. Deterministic Modeling of Wireless Channel

The goal of this section is to introduce the linear deterministic model and illustrate how we can deterministically model three key features of a wireless channel.

A. Modeling Signal Strength

Consider the *real* scalar Gaussian model for a point-to-point link

$$y = hx + z \tag{1}$$

where $z \sim \mathcal{N}(0,1)$. There is also an average power constraint $E[|x|^2] \leq 1$ at the transmitter. The transmit power and noise power are both normalized to be equal to 1 and the channel gain h is related to the signal-to-noise ratio (SNR) by

$$|h| = \sqrt{\mathsf{SNR}}. \tag{2}$$

It is well known that the capacity of this point-to-point channel is

$$C_{\text{AWGN}} = \frac{1}{2}\log(1+\mathsf{SNR}). \tag{3}$$

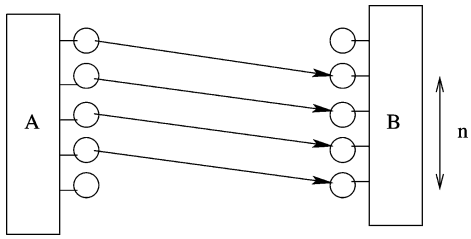

Fig. 1. Pictorial representation of the deterministic model for point-to-point channel.

To get an intuitive understanding of this capacity formula let us write the received signal in (1), y, in terms of the binary expansions of x and z. For simplicity assuming h, x and z are positive real numbers and x has a peak power constraint of 1, we have

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}} \sum_{i=1}^{\infty} x(i)2^{-i} + \sum_{i=-\infty}^{\infty} z(i)2^{-i}. \tag{4}$$

To simplify the effect of background noise assume it has a peak power equal to 1. Then we can write

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}} \sum_{i=1}^{\infty} x(i)2^{-i} + \sum_{i=1}^{\infty} z(i)2^{-i} \tag{5}$$

or,

$$y \approx 2^{n} \sum_{i=1}^{n} x(i)2^{-i} + \sum_{i=1}^{\infty} (x(i+n) + z(i))2^{-i} \tag{6}$$

where $n = \lceil \frac{1}{2}\log \mathsf{SNR} \rceil^{+}$. Therefore, if we just ignore the 1 bit of the carry-over from the second summation ($\sum_{i=1}^{\infty}(x(i+n)+z(i))2^{-i}$) to the first summation ($2^n \sum_{i=1}^{n} x(i)2^{-i}$) we can approximate a point-to-point Gaussian channel as a pipe that truncates the transmitted signal and only passes the bits that are above the noise level. Therefore, think of transmitted signal x as a sequence of bits at different signal levels, with the highest signal level in x being the most significant bit and the lowest level being the least significant bit. In this simplified model the receiver can see the n most significant bits of x without any noise and the rest are not seen at all. There is a correspondence between n and SNR in dB scale

$$n \leftrightarrow \left\lceil \frac{1}{2}\log \mathsf{SNR} \right\rceil^{+}. \tag{7}$$

This simplified model, shown in Fig. 1, is deterministic. Each circle in the figure represents a signal level which holds a binary digit for transmission. The most significant n bits are received at the destination while less significant bits are not.

These signal levels can potentially be created using a multilevel lattice code in the AWGN channel [18]. Then the first n levels in the deterministic model represent those levels (in the lattice chain) that are above noise level, and the remaining are the ones that are below noise level. We can algebraically write this input-output relationship by shifting $\mathbf{x}$ down by $q-n$ elements

$$\mathbf{y} = \mathbf{S}^{\mathbf{q-n}}\mathbf{x} \tag{8}$$

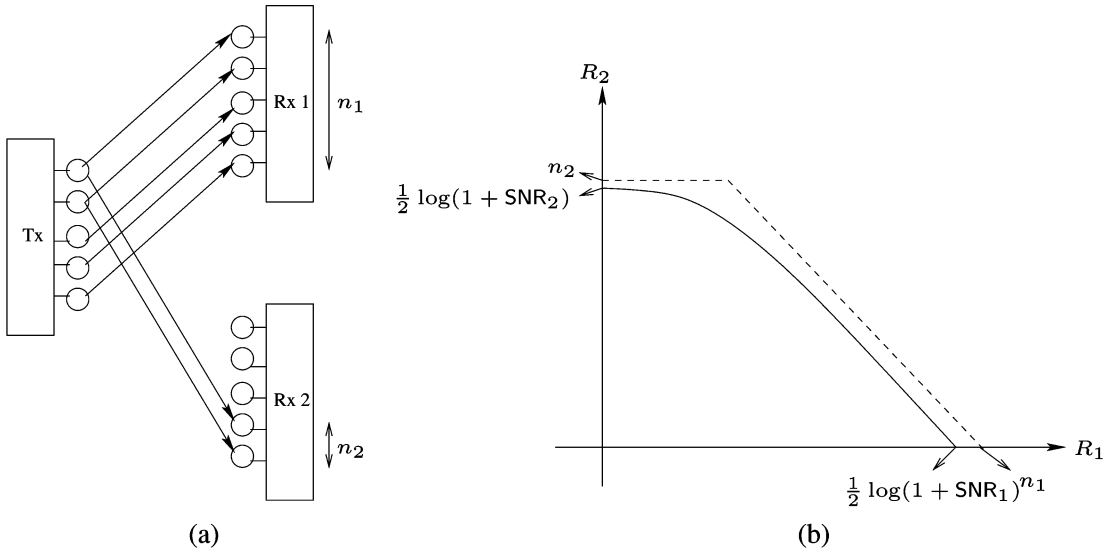

Fig. 2. Pictorial representation of the deterministic model for Gaussian BC is shown in (a). Capacity region of Gaussian and deterministic BC are shown in (b).

where $\mathbf{x}$ and $\mathbf{y}$ are binary vectors of length q denoting transmit and received signals respectively and $\mathbf{S}$ is the $q \times q$ shift matrix

$$\mathbf{S} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \tag{9}$$

The capacity of this deterministic point-to-point channel is n, where $n = \lceil \frac{1}{2} \log \mathsf{SNR} \rceil^+$. This capacity is within $\frac{1}{2}$-bit approximation of the capacity of the AWGN channel. In the case of complex Gaussian channel we set $n = \lceil \log \mathsf{SNR} \rceil^+$ and we get an approximation within 1-bit of the capacity.

B. Modeling Broadcast

Based on the intuition obtained so far, it is straightforward to think of a deterministic model for a broadcast scenario. Consider the real scalar Gaussian broadcast channel (BC). Assume there are only two receivers. The received SNR at receiver i is denoted by SNR_i for $i = 1, 2\,(\mathsf{SNR}_2 \leq \mathsf{SNR}_1)$. Consider the binary expansion of the transmitted signal, x. Then we can deterministically model the Gaussian broadcast channel as follows:

- Receiver 2 (weak user) receives only the most significant n_2 bits in the binary expansion of x. Those bits are the ones that arrive above its noise level.
- Receiver 1 (strong user) receives the most significant n_1 $(n_1 > n_2)$ bits in the binary expansion of x. Clearly these bits contain what receiver 2 gets.

The deterministic model makes explicit the functioning of superposition coding and successive interference cancellation decoding in the Gaussian broadcast channel. The most significant n_2 levels in the deterministic model represent the cloud center that is decoded by both users, and the remaining $n_1 - n_2$ levels represent the cloud detail that is decoded only by the strong user (after decoding the cloud center and canceling it from the received signal).

Pictorially the deterministic model is shown in Fig. 2(a). In this particular example $n_1 = 5$ and $n_2 = 2$; therefore, both users receive the two most significant bits of the transmitted signal. However, user 1 (strong user) receives three additional bits from the next three signal levels of the transmitted signal. There is also the same correspondence between n and channel gains in dB

$$n_i \leftrightarrow \left\lceil \frac{1}{2} \log \mathsf{SNR}_i \right\rceil^+, \quad i = 1, 2. \tag{10}$$

To analytically demonstrate how closely we are modeling the Gaussian BC channel, the capacity region of the Gaussian BC channel and the deterministic BC channel are shown in Fig. 2(b). As it is seen their capacity regions are very close to each other. In fact it is easy to verify that for all SNR's these regions are always within one bit per user of each other, that is, if (R_1, R_2) is in the capacity region of the deterministic BC then there is a rate pair within one bit component-wise of (R_1, R_2) that is in the capacity region of the Gaussian BC. However, this is only the worst-case gap and in the typical case where SNR_1 and SNR_2 are very different, the gap is much smaller than one bit.

C. Modeling Superposition

Consider a superposition scenario in which two users are simultaneously transmitting to a node. In the Gaussian model the received signal can be written as

$$y = h_1 x_1 + h_2 x_2 + z. \tag{11}$$

To intuitively see what happens in superposition in the Gaussian model, we again write the received signal, y, in terms of the binary expansions of x_1, x_2 and z. Assume x_1, x_2 and z are all positive real numbers smaller than one, and also the channel gains are

$$h_i = \sqrt{\mathsf{SNR}_i}, \quad i = 1, 2. \tag{12}$$

Without loss of generality assume $\mathsf{SNR}_2 \leq \mathsf{SNR}_1$. Then

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i)2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i)2^{-i} + \sum_{i=-\infty}^{\infty} z(i)2^{-i}.$$

To simplify the effect of background noise assume it has a peak power equal to 1. Then we can write

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i)2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i)2^{-i} + \sum_{i=1}^{\infty} z(i)2^{-i}$$

or,

$$y \approx 2^{n_1} \sum_{i=1}^{n_1-n_2} x_1(i)2^{-i} + 2^{n_2} \sum_{i=1}^{n_2} (x_1(i+n_1-n_2) + x_2(i))2^{-i} + \sum_{i=1}^{\infty} (x_1(i+n_1) + x_2(i+n_2) + z(i))2^{-i}$$

where $n_i = \lceil \frac{1}{2}\log \mathsf{SNR}_i \rceil^+$ for $i = 1, 2$. Therefore, based on the intuition obtained from the point-to-point and broadcast AWGN channels, we can approximately model this as follows:

- That part of x_1 that is above SNR_2 $(x_1(i), 1 \leq i \leq n_1 - n_2)$ is received clearly without any contribution from x_2.
- The remaining part of x_1 that is above noise level $(x_1(i), n_1 - n_2 < i \leq n_1)$ and that part of x_2 that is above noise level $(x_1(i), 1 \leq i \leq n_2)$ are superposed on each other and are received without any noise.
- Those parts of x_1 and x_2 that are below noise level are truncated and not received at all.

The key point is how to model the superposition of the bits that are received at the same signal level. In our deterministic model we ignore the carry-overs of the real addition and we model the superposition by the modulo 2 sum of the bits that are arrived at the same signal level. Pictorially the deterministic model is shown in Fig. 4(a). Analogous to the deterministic model for the point-to-point channel, as seen in Fig. 3, we can write

$$\mathbf{y} = \mathbf{S}^{\mathbf{q}-\mathbf{n_1}}\mathbf{x_1} \oplus \mathbf{S}^{\mathbf{q}-\mathbf{n_2}}\mathbf{x_2} \tag{13}$$

where the summation is in $\mathbb{F}_2$ (modulo 2). Here $\mathbf{x_i}$ $(i = 1, 2)$ and $\mathbf{y}$ are binary vectors of length q denoting transmitted and received signals respectively and $\mathbf{S}$ is a $q \times q$ shift matrix. The relationship between n_i's and the channel gains is the same as in (10).

Compared to the point-to-point case we now have interaction between the bits that are received at the same signal level at the receiver. We limit the receiver to observe only the modulo 2 summation of those bits that arrive at the same signal level. This way of modeling signal interaction has two advantages over the simplistic collision model. First, if two bits arrive simultaneously at the same signal level, they are not both dropped and the receiver gets their modulo 2 summation. Second, unlike in the collision model where the entire packet is lost when there is collision, the most significant bits of the stronger user remain intact. This is reminiscent of the familiar *capture* phenomenon in CDMA systems: the strongest user can be heard even when multiple users simultaneously transmit.

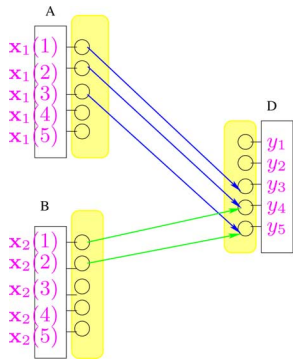

Fig. 3. Algebraic representation of shift matrix deterministic model.

Now we can apply this model to the Gaussian MAC, in which

$$y = h_1 x_1 + h_2 x_2 + z \tag{14}$$

where $z \sim \mathcal{CN}(0, 1)$. There is also an average power constraint equal to 1 at both transmitters. A natural question is how close is the capacity region of the deterministic model to that of the actual Gaussian model. Assume $\mathsf{SNR}_2 < \mathsf{SNR}_1$. The capacity region of this channel is known to be the set of non-negative pairs (R_1, R_2) satisfying

$$R_i \leq \log(1 + \mathsf{SNR}_i), \quad i = 1, 2 \tag{15}$$

$$R_1 + R_2 \leq \log(1 + \mathsf{SNR}_1 + \mathsf{SNR}_2). \tag{16}$$

This region is plotted with solid line in Fig. 4(b).

It is easy to verify that the capacity region of the deterministic MAC is the set of non-negative pairs (R_1, R_2) satisfying

$$R_2 \leq n_2 \tag{17}$$

$$R_1 + R_2 \leq n_1 \tag{18}$$

where $n_i = \lceil \log \mathsf{SNR}_i \rceil^+$ for $i = 1, 2$. This region is plotted with dashed line in Fig. 4(b). In this deterministic model the "carry-over" from one level to the next that would happen with real addition is ignored. However, as we notice still the capacity region is very close to the capacity region of the Gaussian model. In fact it is easy to verify that they are within one bit per user of each other. The intuitive explanation for this is that in real addition once two bounded signals are added together the magnitude can become as large as twice the larger of the two signals. Therefore, the number of bits in the sum is increased by at most one bit. On the other hand in finite-field addition there is no magnitude associated with signals and the summation is still in the same field as the individual signals. So the gap between Gaussian and deterministic model for two user MAC is intuitively this one bit of cardinality increase. Similar to the broadcast example, this is only the worst case gap and when the channel gains are different it is much smaller than one bit.

Now we define the linear finite-field deterministic model for the relay network.

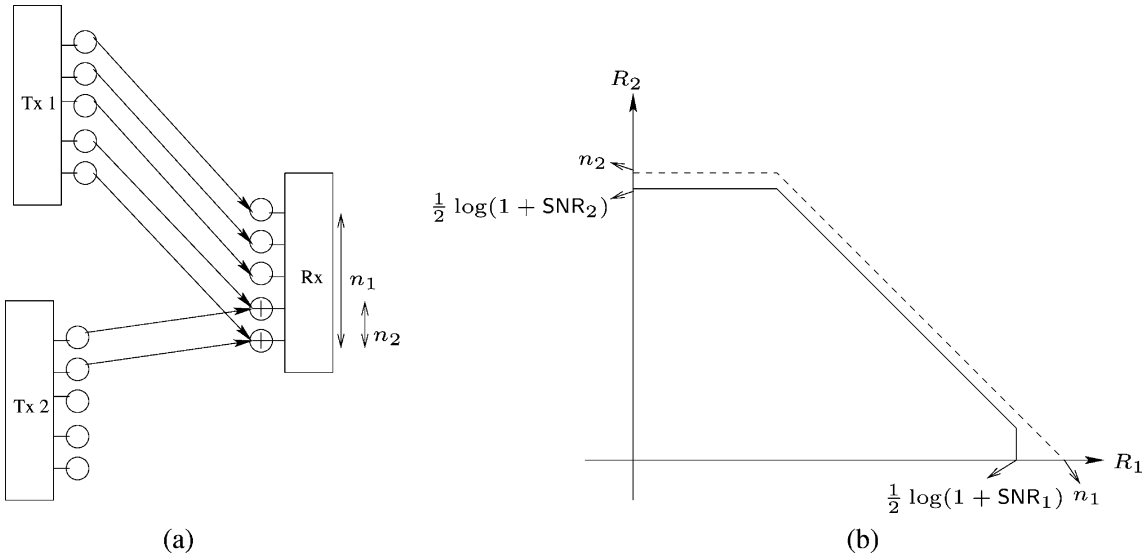

Fig. 4. Pictorial representation of the deterministic MAC is shown in (a). Capacity region of Gaussian and deterministic MACs are shown in (b).

D. Linear Finite-Field Deterministic Model

The relay network is defined using a set of vertices $\mathcal{V}$. The communication link from node i to node j has a non-negative integer gain n_{ij} associated with it. This number models the channel gain in the corresponding Gaussian setting. At each time t, node i transmits a vector $\mathbf{x}_i[t] \in \mathbb{F}_p^q$ and receives a vector $\mathbf{y}_i[t] \in \mathbb{F}_p^q$ where $q = \max_{i,j}(n_{(ij)})$ and p is a positive integer indicating the field size. The received signal at each node is a deterministic function of the transmitted signals at the other nodes, with the following input-output relation: if the nodes in the network transmit $\mathbf{x}_1[t], \mathbf{x}_2[t], \ldots \mathbf{x}_N[t]$ then the received signal at node j, $1 \leq j \leq N$ is

$$\mathbf{y}_j[t] = \sum \mathbf{S}^{q-n_{ij}} \mathbf{x}_i[t] \tag{19}$$

where the summations and the multiplications are in $\mathbb{F}_p$. Throughout this paper the field size, p, is assumed to be 2, unless it is stated otherwise.

E. Limitation: Modeling MIMO

The examples in the previous subsections may give the impression that the capacity of any Gaussian channel is within a constant gap to that of the corresponding linear deterministic model. The following example shows that is not the case.

Consider a 2×2 MIMO real Gaussian channel with channel gain values as shown in Fig. 5(a), where k is an integer larger than 2. The channel matrix is

$$\mathbf{H} = 2^k \begin{pmatrix} \frac{3}{4} & 1 \\ 1 & 1 \end{pmatrix}. \tag{20}$$

The channel gain parameters of the corresponding linear finite-field deterministic model are:

$$\begin{aligned} n_{11} &= \left\lceil \frac{1}{2}\log_2 |h_{11}|^2 \right\rceil^+ = \lceil \log_2(2^k - 2^{k-2}) \rceil^+ = k \\ n_{12} &= n_{21} = n_{22} = \lceil \log_2 2^k \rceil^+ = k \end{aligned} \tag{21}$$

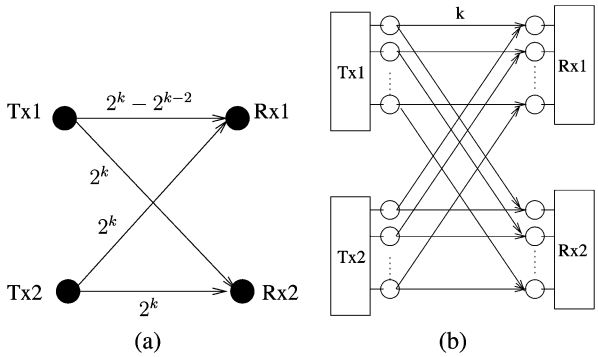

Fig. 5. Example of a 2×2 Gaussian MIMO channel is shown in (*a*). The corresponding linear finite-field deterministic MIMO channel is shown in (*b*).

Now let us compare the capacity of the MIMO channel under these two models for large values of k. For the Gaussian model, both singular values of $\mathbf{H}$ are of the order of 2^k. Hence, the capacity of the real Gaussian MIMO channel is of the order of

$$2 \times \frac{1}{2}\log(1 + |2^k|^2) \approx 2k.$$

However, the capacity of the corresponding linear finite-field deterministic MIMO is simply

$$C_{\text{LFF}} = \text{rank} \begin{pmatrix} I_k & I_k \\ I_k & I_k \end{pmatrix} = k. \tag{22}$$

Hence, the gap between the two capacities goes to infinity as k increases.

Even though the linear deterministic channel model does not approximate the Gaussian channel in all scenarios, it is still useful in providing insights in many cases, as will be seen in the next section. Moreover, its analytic simplicity allows an exact analysis of the relay network capacity. This in turns provides the foundation for the analysis of the Gaussian network.

III. Motivation of Our Approach

In this section we motivate and illustrate our approach. We look at three simple relay networks and illustrate how the analysis of these networks under the simpler linear finite-field deterministic model enables us to conjecture an approximately optimal relaying scheme for the Gaussian case. We progress from the single relay channel where several strategies yield uniform approximation to more complicated networks where progressively we see that several "simple" strategies in the literature fail to achieve a constant gap. Using the deterministic model we can whittle down the potentially successful strategies. This illustrates the power of the deterministic model to provide insights into transmission techniques for noisy networks.

The network is assumed to be synchronized, i.e., all transmissions occur on a common clock. The relays are allowed to do any *causal* processing. Therefore, their current output depends only its past received signals. For any such network, there is a natural information-theoretic cut-set bound [19], which upper bounds the reliable transmission rate R. Applied to the relay network, we have the cut-set upper bound $\bar{C}$ on its capacity

$$\bar{C} = \max_{p(\{\mathbf{x}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(\mathbf{y}_{\Omega^c}; \mathbf{x}_\Omega \,|\, \mathbf{x}_{\Omega^c}) \tag{23}$$

where $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts. In words, the value of a given cut Ω is the information rate achieved when the nodes in Ω fully cooperate to transmit and the nodes in Ω^c fully cooperate to receive. In the case of Gaussian networks, this is simply the mutual information achieved in a MIMO channel, the computation of which is standard. We will use this cut-set bound to assess how good our achievable strategies are.

A. Single-Relay Network

We start by looking at the simplest Gaussian relay network with only one relay as shown in Fig. 6(a). To approximate its capacity uniformly (uniform over all channel gains), we need to find a relaying protocol that achieves a rate close to an upper bound on the capacity for all channel parameters. To find such a scheme we use the linear finite-field deterministic model to gain insight. The corresponding linear finite-field deterministic model of this relay channel with channel gains denoted by n_{SR}, n_{SD} and n_{RD} is shown in Fig. 6(b). It is easy to see that the capacity of this deterministic relay channel, C^d_{relay}, is smaller than both the maximum number of bits that the source can broadcast, and the maximum number of bits that the destination can receive. Therefore

$$\begin{aligned} C^d_{\text{relay}} &\le \min(\max(n_{SR}, n_{SD}), \max(n_{RD}, n_{SD})) \\ &= \begin{cases} n_{SD}, & \text{if } n_{SD} > \min(n_{SR}, n_{RD}) \\ \min(n_{SR}, n_{RD}), & \text{otherwise.} \end{cases} \end{aligned} \tag{24}$$

It is not difficult to see that this is in fact the cut-set upper bound for the linear deterministic network.

Note that (24) naturally implies a capacity-achieving scheme for this deterministic relay network: if the direct link is better than any of the links to/from the relay then the relay is silent, otherwise it helps the source by decoding its message and sending innovations. In the example of Fig. 6, the destination receives two bits directly from the source, and the relay increases the capacity by 1 bit by forwarding the least significant bit it receives on a level that does not overlap with the direct transmission at the destination. This suggests a decode-and-forward scheme for the original Gaussian relay channel. The question is: how does it perform? Although unlike in the deterministic network, the decode-forward protocol cannot achieve exactly the cut-set bound in the Gaussian nettwork, the following theorem shows it is close.

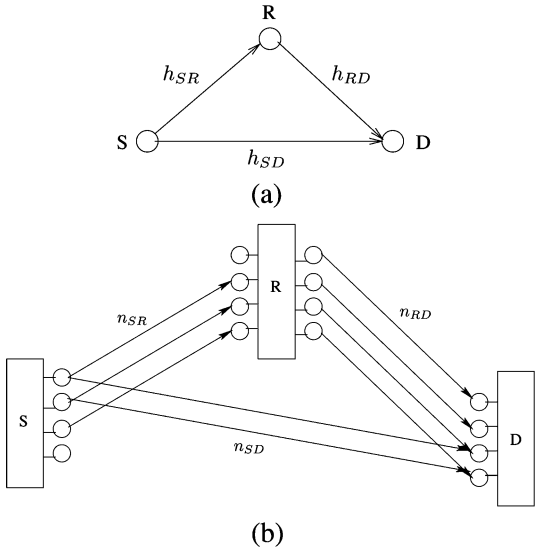

Fig. 6. Relay channel: (a) Gaussian model; (b) linear finite-field deterministic model.

Theorem 3.1: The decode-and-forward relaying protocol achieves within 1 bit/s/Hz of the cut-set bound of the single-relay Gaussian network, for all channel gains.

Proof: See Appendix A. ∎

We should point out that even this 1-bit gap is too conservative for many parameter values. In fact the gap would be at the maximum value only if two of the channel gains are exactly the same. This is rare in wireless scenarios. In Fig. 7 the gap between the achievable rate of decode-forward scheme and the cut-set upper bound is plotted for different channel gains.

The deterministic network in Fig. 6(b) suggests that several other relaying strategies are also optimal. For example, compress–forward [4] will also achieve the cut-set bound. Moreover a "network coding" strategy of sending the sum (or linear combination) of the received bits is also optimal as long as the destination receives linearly independent equations. All these schemes can also be translated to the Gaussian case and can be shown to be uniformly approximate strategies. Therefore, for the simple relay channel there are many successful candidate strategies.

B. Diamond Network

Now consider the diamond Gaussian relay network, with two relays, as shown in Fig. 8(a). Schein introduced this network in his Ph.D. thesis [12] and investigated its capacity. However, the

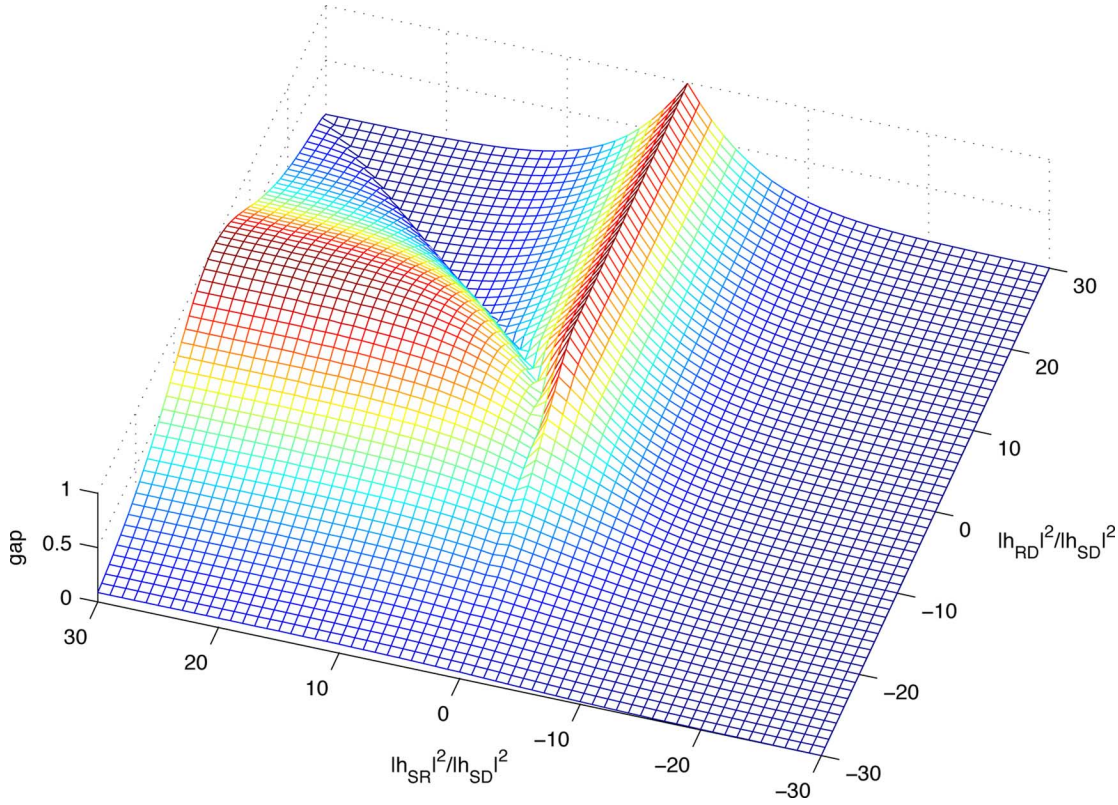

Fig. 7. The x and y axis respectively represent the channel gains from relay to destination and source to relay normalized by the gain of the direct link (source to destination) in dB scale. The z axis shows the value of the gap between the cut-set upper bound and the achievable rate of decode-forward scheme in bits/sec/Hz.

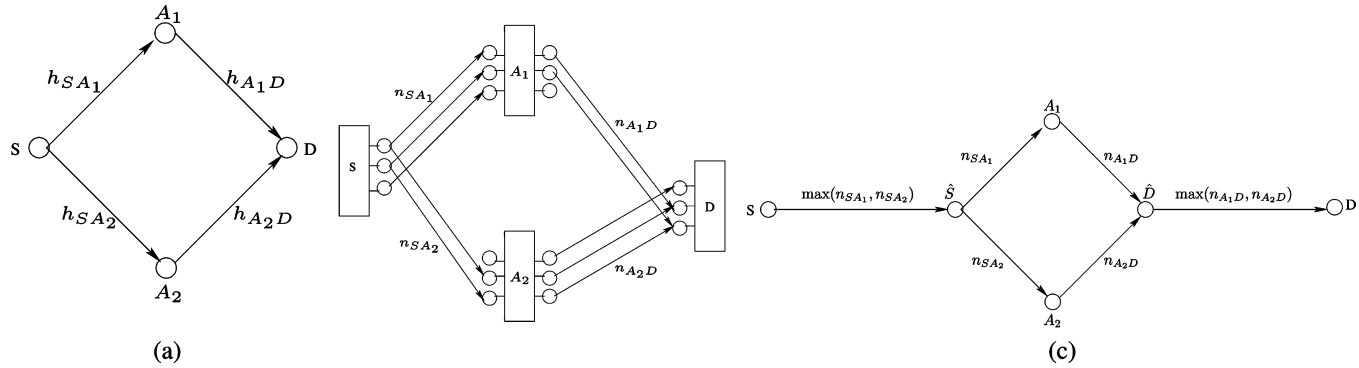

Fig. 8. Diamond network with two relays: (a) Gaussian model, (b) Linear finite-field deterministic model, and (c) wired model.

capacity of this network is still open. We would like to uniformly approximate its capacity.

First we build the corresponding linear finite-field deterministic model for this relay network as shown in Fig. 8(b). To investigate its capacity first we relax the interactions between incoming links at each node and create the wired network shown in Fig. 8(c). In this network there are two other links added, which are from S to $\hat{S}$ and from $\hat{D}$ to D. Since the capacities of these links are respectively equal to the maximum number of bits that can be sent by the source and maximum number of bits that can be received by the destination in the original linear finite-field deterministic network, the capacity of the wired diamond network cannot be smaller than the capacity of the linear finite-field deterministic diamond network. Now by the max-flow min-cut theorem we know that the capacity C^w_{diamond} of the wired diamond network is equal to the value of its minimum cut. Hence

$$C^d_{\text{diamond}} \le C^w_{\text{diamond}} = \min\{\max(n_{SA_1}, n_{SA_2}), \max(n_{A_1D}, n_{A_2D}), n_{SA_1} + n_{A_2D}, n_{SA_2} + n_{A_1D}\}. \quad (25)$$

As we will show in Section V, this upper bound is in fact the cut-set upper bound on the capacity of the deterministic diamond network.

Now, we know that the capacity of a wired network is achieved by a routing solution. We can indeed mimic the wired network routing solution in the linear finite-field deterministic diamond network and send the same amount of information through noninterfering links from source to relays and then

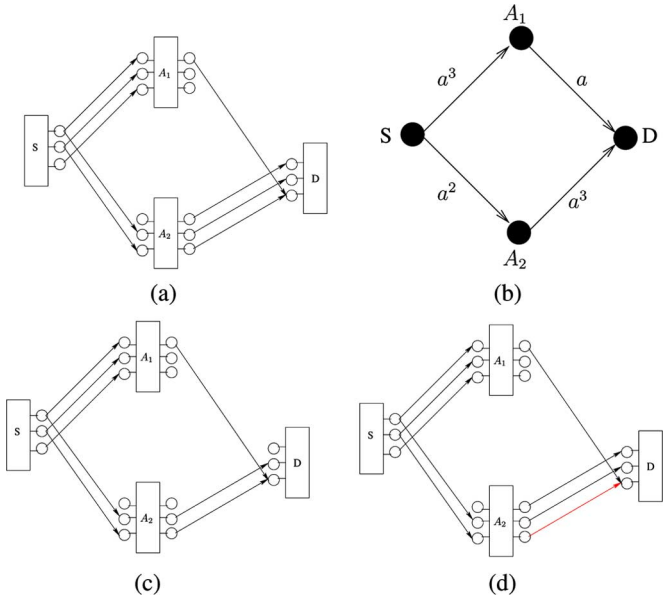

Fig. 9. Example of the linear finite-field deterministic diamond network is shown in (a). The corresponding Gaussian network is shown in (b), with the gains chosen such that the ratio of the gains in dB scale match the ratios of the gains in the deterministic network. The effective network when R_2 just forwards the received signal is shown in (c). The effective network when R_2 amplifies the received signal to shift it up one signal level and then forward the message is shown in (d).

from relays to destination. Therefore, the capacity of the deterministic diamond network is equal to its cut-set upper bound.

A natural analogy of this routing scheme for the Gaussian network is the following partial-decode-and-forward strategy:

1) The source broadcasts two messages, m_1 and m_2, at rate R_1 and R_2 to relays A_1 and A_2, respectively.
2) Each relay A_i decodes message m_i, $i = 1, 2$.
3) Then A_1 and A_2 re-encode the messages and transmit them via the MAC channel to the destination.

Clearly the destination can decode both m_1 and m_2 if (R_1, R_2) is inside the capacity region of the BC from source to relays as well as the capacity region of the MAC from relays to the destination. The following theorem shows how good this scheme is.

Theorem 3.2: Partial-decode-and-forward relaying protocol achieves within 1 bit/s/Hz of the cut-set upper bound of the two-relay diamond Gaussian network, for all channel gains.

Proof: See Appendix B. ∎

We can also use the linear finite-field deterministic model to understand why other simple protocols such as decode-forward and amplify-forward are not universally-approximate strategies for the diamond network.

Consider an example linear finite-field diamond network shown in Fig. 9(a). The cut-set upper bound on the capacity of this network is 3 bits/unit time. In a decode-forward scheme, all participating relays should be able to decode the message. Therefore, the maximum rate of the message broadcasted from the source can at most be 2 bits/unit time. Also, if we ignore relay A_2 and only use the stronger relay, still it is not possible to send information more at a rate more than 1 bit/unit time. As a result we cannot achieve the capacity of this network by using a decode-forward strategy.

We next show that this 1-bit gap can be translated into an unbounded gap in the corresponding Gaussian network, as shown in Fig. 9(b). By looking at the cut between the destination and the rest of the network, it can be seen that for large a, the cut-set upper bound is approximately

$$\bar{C} \approx 3 \log a. \tag{26}$$

The achievable rate of the decode-forward strategy is upper bounded by

$$R_{DF} \leq 2 \log a. \tag{27}$$

Therefore, as a gets larger, the gap between the achievable rate of decode-forward strategy and the cut-set upper bound (26) increases.

Let us look at the amplify-forward scheme. Although this scheme does not require all relays to decode the entire message, it can be quite sub-optimal if relays inject significant noise into the system. We use the deterministic model to intuitively see this effect. In a deterministic network, the amplify-forward operation can be simply modeled by shifting bits up and down at each node. However, once the bits are shifted up, the newly created LSB's represent the amplified bits of the noise and we model them by random bits. Now, consider the example shown

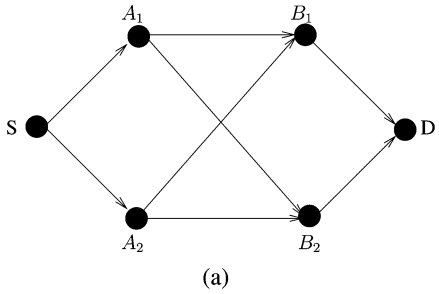

Fig. 10. Two layer relay network with four relays.

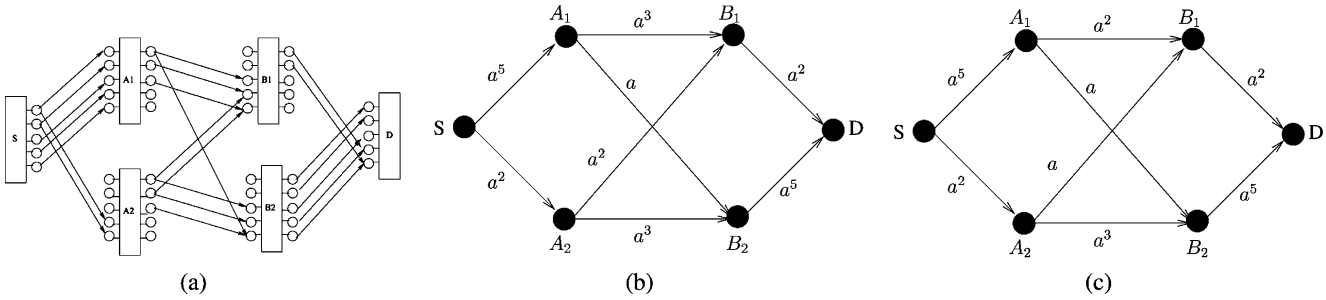

Fig. 11. Example of a four relay linear finite filed deterministic relay network is shown in (a). The corresponding Gaussian relay network is shown in (b). The effective Gaussian network for compress-forward strategy is shown in (c).

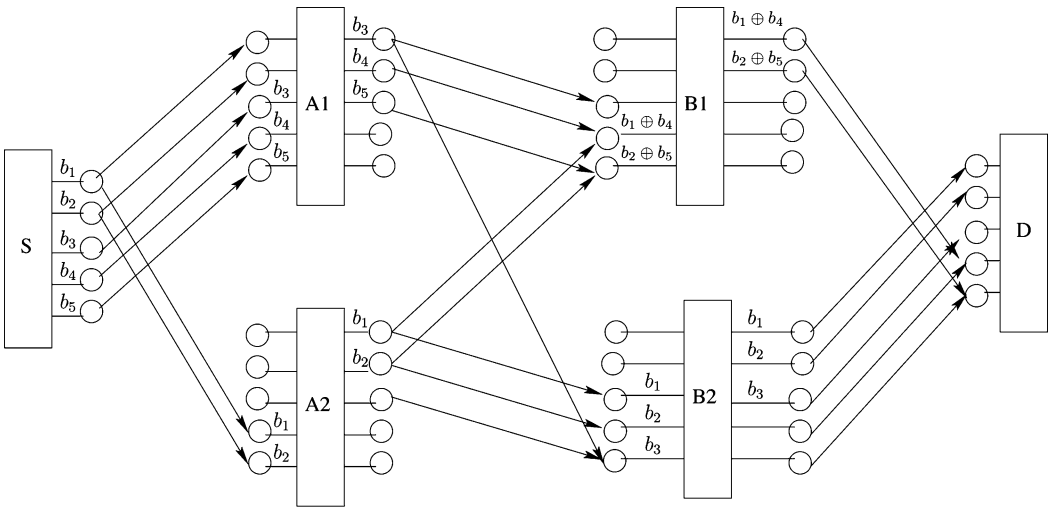

Fig. 12. Demonstration of a capacity achieving strategy.

in Fig. 9(a). We notice that to achieve a rate of 3 from the source to the destination, the least significant bit of the source's signal should go through A_1 while the remaining two bits go through A_2. Now if A_2 is doing amplify-forward, it will have two choices: to either forward the received signal without amplifying it, or to amplify the received signal to have three signal levels in magnitude and forward it.

The effective networks under these two strategies are respectively shown in Fig. 9(c) and (d). In the first case, since the total rate going through the MAC from A_1 and A_2 to D is less than two, the overall achievable rate cannot exceed two. In the second case, however, the inefficiency of amplify-forward strategy comes from the fact that A_2 is transmitting pure noise on its lowest signal level. As a result, it is corrupting the bit transmitted by A_1 and reducing the total achievable rate again to two bits/channel use. Therefore, for this channel realization, the amplify-forward scheme does not achieve the capacity. This intuition can again be translated to the corresponding Gaussian network to show that amplify-and-forward is not a universally-approximate strategy for the diamond network.

C. A Four-Relay Network

We now look at a more complicated relay network with four relays, as shown in Fig. 10. As the first step let us find the optimal relaying strategy for the corresponding linear finite field deterministic model. Consider an example of a linear finite field deterministic relay network shown in Fig. 11(a). Now focus on the relaying strategy that is pictorially shown in Fig. 12. In this scheme:

- Source broadcasts $\mathbf{b} = [b_1, \ldots, b_5]^t$.
- Relay A_1 decodes b_3, b_4, b_5 and relay A_2 decodes b_1, b_2.
- Relay A_1 and A_2 respectively send $\mathbf{x}_{A_1} = [b_3, b_4, b_5, 0, 0]^t$ and $\mathbf{x}_{A_2} = [b_1, b_2, 0, 0, 0]^t$.
- Relay B_2 decodes b_1, b_2, b_3 and sends $\mathbf{x}_{B_2} = [b_1, b_2, b_3, 0, 0]^t$.
- Relay B_1 receives $\mathbf{y}_{B_1} = [0, 0, b_3, b_4 \oplus b_1, b_5 \oplus b_2]^t$ and forwards the last two equations, $\mathbf{x}_{B_1} = [b_4 \oplus b_1, b_5 \oplus b_2, 0, 0, 0]^t$.
- The destination gets $\mathbf{y}_D = [b_1, b_2, b_3, b_4 \oplus b_1, b_5 \oplus b_2]^t$ and is able to decode all five bits.

This scheme can achieve 5 bits per unit time, clearly the best that one can do since the destination only receives 5 bits per unit time. In this optimal scheme the relay B_1 is not decoding or partially decoding the original flows of bits that were broadcasted by the source; it is decoding and forwarding a linear combination of them. One may wonder if this is necessary. To answer this question note that since all transmitted signal levels of A_1 and A_2 are interfering with each other, it is not possible to get a rate of more than 3 bits/unit time by any scheme which does not allow mixing of the flows of information bits originating from the source.

The last stage in the above scheme can actually be interpreted as a compress–forward strategy: relays B_1 and B_2 want to send their 3-bit received vectors to the destination D, but because the link from B_1 to D only supports 2 bits, the dependency between these received vectors must be exploited. However, in the Gaussian network, we *cannot* implement this strategy using a standard compress–forward scheme pretending that the two received signals at B_1 and B_2 are jointly Gaussian. They are not. Relay A_2 sends nothing on its LSB, allowing the MSB of relay A_1 to come through and appear as the LSB of the received signal at B_2. In fact, the statistical correlation between the real-valued received signals at B_1 and B_2 is quite weak since their MSBs are totally independent. Only when one views the received signals as vectors of bits, as guided by the deterministic model, the dependency between them becomes apparent. In fact, it can be shown that a compress–forward strategy assuming jointly Gaussian distributed received signals cannot achieve a constant gap to the cut-set bound.

D. Summary

We learned two key points from the above examples:

- All the schemes that achieve capacity of the deterministic networks in the examples forward the received bits at the various signal levels.
- Using the deterministic model as a guide, it is revealed that commonly used schemes such as decode–forward, amplify–forward, and Gaussian compress–forward can all be very far-away from the cut-set bound.

We devote the rest of the paper to generalizing the steps we took for the examples. As we will show, in the deterministic relay network the optimal strategy for each relay is to simply shuffle and linearly combine the received signals at various levels and forward them. This insight leads to a natural *quantize-map-and-forward* strategy for noisy (Gaussian) relay networks. The strategy for each relay is to quantize the received signal at the distortion of the noise power. This in effect extracts the bits of the received signals above the noise level. These bits are then mapped randomly to a transmit Gaussian codeword. The main result of our paper is to show that such a scheme is indeed universally approximate for arbitrary noisy Gaussian relay networks.

IV. Main Results

In this section we precisely state the main results of the paper and briefly discuss their implications. The capacity of a relay network, C, is defined as the supremum of all achievable rates of reliable communication from the source to the destination. Similarly, the multicast capacity of relay network is defined as the maximum rate at which the source can send the same information simultaneously to all destinations.

A. Deterministic Networks

1) General Deterministic Relay Network: In the general deterministic model the received vector signal $\mathbf{y}_j$ at node $j \in \mathcal{V}$ at time t is given by

$$\mathbf{y}_j[t] = \mathbf{g}_j(\{\mathbf{x}_i[t]\}_{i\in\mathcal{V}}) \tag{28}$$

where $\{\mathbf{x}_i[t]\}_{i\in\mathcal{V}}$ denotes the transmitted signals at all of the nodes in the network. Note that this implies a deterministic multiple access channel for node j and a deterministic broadcast channel for the transmitting nodes, so both broadcast and multiple access is allowed in this model. This is a generalization of Aref networks [5] which only allow broadcast.

The cut-set bound of a general deterministic relay network is

$$\bar{C} = \max_{p(\{\mathbf{x}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(\mathbf{y}_{\Omega^c}; \mathbf{x}_\Omega | \mathbf{x}_{\Omega^c}) \tag{29}$$

$$\stackrel{(a)}{=} \max_{p(\{\mathbf{x}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} H(\mathbf{y}_{\Omega^c} | \mathbf{x}_{\Omega^c}) \tag{30}$$

where $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts. Step (a) follows since we are dealing with deterministic networks.

The following are our main results for arbitrary deterministic networks.

Theorem 4.1: A rate of

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{x}_i)} \min_{\Omega\in\Lambda_D} H(\mathbf{y}_{\Omega^c} \,|\, \mathbf{x}_{\Omega^c}) \tag{31}$$

can be achieved on a deterministic network.

This theorem easily extends to the multicast case, where we want to simultaneously transmit one message from S to all destinations in the set $D \in \mathcal{D}$:

Theorem 4.2: A multicast rate of

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{x}_i)} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} H(\mathbf{y}_{\Omega^c} \,|\, \mathbf{x}_{\Omega^c}) \tag{32}$$

to all the destinations $D \in \mathcal{D}$ can be achieved on a deterministic network.

Note that when we compare (31) to the cut-set upper bound in (30), we see that the difference is in the maximizing set, i.e., we are only able to achieve independent (product) distributions whereas the cut-set optimization is over any arbitrary distribution. In particular, if the network and the deterministic functions are such that the cut-set is optimized by the product distribution, then we would have matching upper and lower bounds. This happens for deterministic networks with broadcast only, specializing to the result in [9]. It also happens when we consider the linear finite-field model, whose results are stated next.

2) Linear Finite-Field Deterministic Relay Network: Applying the cut-set bound to the linear finite-field deterministic relay network defined in Section II-D, (19), and using (30) since we have a deterministic network, we get

$$\bar{C} = \max_{p(\{\mathbf{x}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} H(\mathbf{y}_{\Omega^c} \,|\, \mathbf{x}_{\Omega^c}) \stackrel{(b)}{=} \min_{\Omega\in\Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) \tag{33}$$

where $\mathbf{G}_{\Omega,\Omega^c}$ is the transfer matrix associated with the cut Ω, i.e., the matrix relating the vector of all the inputs at the nodes in Ω to the vector of all the outputs in Ω^c induced by (19). This is illustrated in Fig. 13. Step (b) follows since in the linear finite-field model all cut values (i.e., $H(\mathbf{y}_{\Omega^c} \,|\, \mathbf{x}_{\Omega^c})$) are simultaneously optimized by independent and uniform distribution of $\{\mathbf{x}_i\}_{i\in\mathcal{V}}$ and the optimum value of each cut Ω is logarithm of

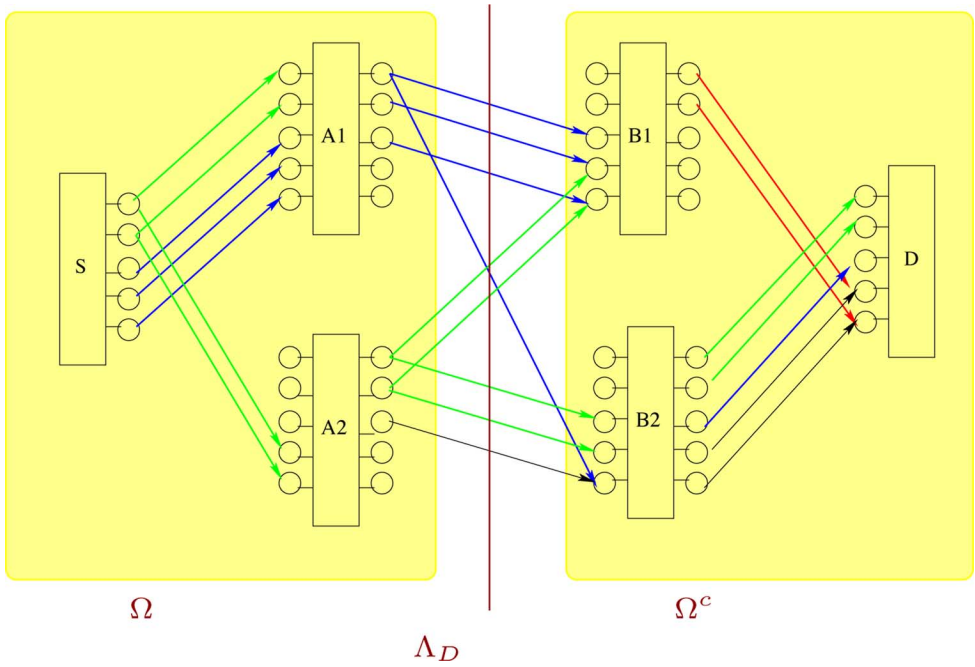

Fig. 13. Illustration of cut-set bound and cut-set transfer matrix $\mathbf{G}_{\Omega,\Omega^c}$.

the size of the range space of the transfer matrix $\mathbf{G}_{\Omega,\Omega^c}$ associated with that cut. Theorems 4.1 and 4.2 immediately imply that this cutset bound is achievable.

Theorem 4.3: The capacity C of a linear finite-field deterministic relay network is given by

$$C = \min_{\Omega \in \Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{34}$$

Theorem 4.4: The multicast capacity C of a linear finite-field deterministic relay network is given by

$$C = \min_{D \in \mathcal{D}} \min_{\Omega \in \Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) \tag{35}$$

where $\mathcal{D}$ is the set of destinations.

Remark: Note that the results in Theorems 4.1, 4.2, 4.3, and 4.4 apply to networks with arbitrary topology, possible including cycles. For a single source-destination pair the result in Theorem 4.3 generalizes the classical max-flow min-cut theorem for wired networks and for multicast, the result in Theorem 4.4 generalizes the network coding result in [6]. As we will see in the proof, the encoding functions at the relay nodes for the linear finite-field model can be restricted to linear functions to obtain the result in Theorem 4.3.

B. Gaussian Relay Networks

In the Gaussian model each node $j \in \mathcal{V}$ has M_j transmit and N_j receive antennas. The received signal $\mathbf{y}_j$ at node j and time t is

$$\mathbf{y}_j[t] = \sum_{i \in \mathcal{V}} \mathbf{H}_{ij}\mathbf{x}_i[t] + \mathbf{z}_j[t] \tag{36}$$

where $\mathbf{H}_{ij}$ is an $M_i \times N_j$ complex matrix whose (k,l) element represents the channel gain from the kth transmit antenna in node i to the lth receive antenna in node j. Furthermore, we assume there is an average power constraint equal to 1 at each transmit antenna. Also $\mathbf{z}_j$, representing the channel noise, is modeled as complex Gaussian random vector. The Gaussian noises at different receivers are assumed to be independent of each other.

The following are our main results for Gaussian relay networks; it is proved in Section VI.

Theorem 4.5: The capacity C of the Gaussian relay network satisfies

$$\bar{C} - \kappa \le C \le \bar{C} \tag{37}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in (23), and κ is a constant and is upper bounded by $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$.

Remark: The gap κ holds for all values of the channel gains and the result is relevant particularly in the high rate regime. It is a stronger result than a degree-of-freedom result, because it is nonasymptotic and provides a uniform guarantee to optimality for all channel SNRs. This is the first constant-gap approximation of the capacity of arbitrary Gaussian relay networks. As shown in Section III, the gap between the achievable rate of well known relaying schemes and the cut-set upper bound in general depends on the channel parameters and can become arbitrarily large. Analogous to the results for deterministic networks, the result in Theorem 4.5 applies to a network with arbitrary topology, possibly with cycles.

The result in Theorem 4.5 easily extends to the multicast case where we want to simultaneously transmit one message from S to all destinations in the set $D \in \mathcal{D}$.

Theorem 4.6: The multicast capacity C_{mult} of the Gaussian relay network satisfies

$$\bar{C}_{\text{mult}} - \kappa \le C_{\text{mult}} \le \bar{C}_{\text{mult}} \tag{38}$$

where $\bar{C}_{\text{mult}}$ is the multicast cut-set upper bound on the capacity of $\mathcal{G}$ given by

$$\bar{C}_{\text{mult}} = \max_{p(\{\mathbf{x}_j\}_{j\in\mathcal{V}})} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} I(\mathbf{y}_{\Omega^c}; \mathbf{x}_\Omega \,|\, \mathbf{x}_{\Omega^c}) \tag{39}$$

and κ is a constant and is upper bounded by $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$.

Remark: The gap κ stated in Theorems 4.5–4.6 hold for scalar quantization scheme explored in detail in Section VI. It is shown in [20] that a vector quantization scheme even with structured lattice codebooks can improve this constant to $2\sum_{i=1}^{|\mathcal{V}|} N_i + 2\sum_{i=1}^{|\mathcal{V}|} M_i$, which means when all nodes have single antennas, the gap is at most $4|\mathcal{V}|$, for complex Gaussian networks (or $2|\mathcal{V}|$ for real Gaussian networks). Also, the results have been extended to the case when there are multiple sources and *all* destinations need to decode *all* the sources, i.e., multisource multicast, in [21], [29].

C. Proof Program

In the following sections we formally prove these main results. The main proof program consists of first proving Theorem 4.3 and the corresponding multicast result for linear finite-field deterministic networks in Section V. Since the proof logic of the achievable rate for general deterministic networks (31), (32) is similar to that for the linear case, Theorems 4.1 and 4.2 are proved in Appendix C. We use the proof ideas for the deterministic analysis to obtain the universally-approximate capacity characterization for Gaussian relay networks in Section VI. In both cases, we illustrate the proof by first going through an example.

V. Deterministic Relay Networks

In this section we characterize the capacity of linear finite-field deterministic relay networks and prove Theorems 4.3 and 4.4.

To characterize the capacity of linear finite-field deterministic relay networks, we first focus on networks that have a layered structure, i.e., all paths from the source to the destination have equal lengths. With this special structure we get a major simplification: a sequence of messages can each be encoded into a block of symbols and the blocks do not interact with each other as they pass through the relay nodes in the network. The proof of the result for layered network is similar in style to the random coding argument in Ahlswede *et al.* [6]. We do this in Section V-A. Next, in Section V-B, we extend the result to an arbitrary network by expanding the network over time.[1] Since the time-expanded network is layered, we can apply our result in the first step to it and complete the proof.

[1]The concept of time-expanded network is also used in [6], but the use there is to handle cycles. Our main use is to handle interaction between messages transmitted at different times, an issue that only arises when there is superposition of signals at nodes.

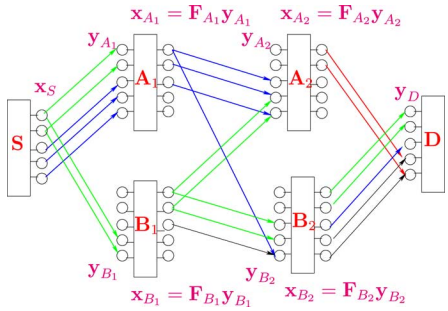

Fig. 14. Illustration of linear encoding strategy.

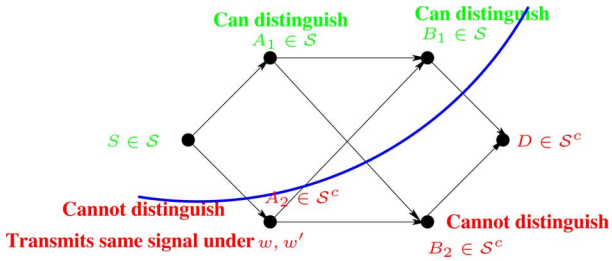

Fig. 15. Example of layered relay network. Nodes on the left hand side of the cut can distinguish between messages w and w', while nodes on the right hand side cannot.

A. Layered Networks

The network given in Fig. 15 is an example of a layered network where the number of hops for each path from S to D is three. We start by describing the encoding scheme.

1) Encoding for Layered Linear Deterministic Relay Network: We have a single source S with a sequence of messages $w_k \in \{1, 2, \ldots, 2^{TR}\}$, $k = 1, 2, \ldots$. Each message is encoded by the source S into a signal over T transmission times (symbols), giving an overall transmission rate of R. Relay j operates over blocks of time T symbols, and uses a mapping $f_j : \mathcal{Y}_j^T \to \mathcal{X}_j^T$ on its received symbols from the previous block of T symbols to transmitted signals in the next block. For the linear deterministic model (19), we use linear mappings $f_j(\cdot)$, i.e.,

$$\mathbf{x}_j = \mathbf{F}_j \mathbf{y}_j \tag{40}$$

where the vectors $\mathbf{x}_j = [\mathbf{x}_j[1], \ldots, \mathbf{x}_j[T]]^t$ and $\mathbf{y}_j = [\mathbf{y}_j[1], \ldots, \mathbf{y}_j[T]]^t$ respectively represent the transmit and received signals over T time units, and the matrix $\mathbf{F}_j$ is chosen uniformly randomly over all matrices in $\mathbb{F}_2^{qT\times qT}$. Each relay does the encoding prescribed by (40). Given the knowledge of all the encoding functions $\mathbf{F}_j$ at the relays, the destination D attempts to decode each message w_k sent by the source. This encoding strategy is illustrated in Fig. 14.

Suppose message w_k is sent by the source in block k. Since each relay j operates only on block of lengths T and the network is layered, the signals received at block k at any relay pertain to only message w_{k-l_j} where l_j is the path length from source to relay j.

2) Proof Illustration: In order to illustrate the proof ideas of Theorem 4.1 we examine the network shown in Fig. 15.

Without loss of generality consider the message $w = w_1$ transmitted by the source at block $k = 1$. At node j the signals pertaining to this message are received by the relays at block l_j.

For notational simplicity we will drop the block numbers associated with the transmitted and received signals for this analysis.

Now, since we have a deterministic network, the message w will be mistaken for another message w' only if the received signal $\mathbf{y}_D(w)$ under w is the same as that would have been received under w'. This leads to a notion of *distinguishability*: messages w, w' are distinguishable at any node j if $\mathbf{y}_j(w) \neq \mathbf{y}_j(w')$.

The probability of error at destination D can be upper bounded using the union bound as

$$P_e \leq 2^{RT}\mathbb{P}\{w \to w'\} = 2^{RT}\mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w')\}. \quad (41)$$

Since channels are deterministic, the randomness is only due to that of the encoder maps. Therefore, the probability of this event depends on the probability that we choose such encoder maps. Now, we can write (42), shown at the bottom of the page, since the events that correspond to occurrence of the distinguishability sets $\Omega \in \Lambda_D$ are disjoint. Let us examine one term in the summation in (42). For example, consider the cut $\Omega = \{S, A_1, B_1\}$ shown in Fig. 15. A necessary condition for this cut to be the distinguishability set is that $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$, along with $\mathbf{y}_{B_2}(w) = \mathbf{y}_{B_2}(w')$ and $\mathbf{y}_D(w) = \mathbf{y}_D(w')$. We first define the following events:

$$\begin{aligned}
\mathcal{A}_i &= \text{the event that } w \text{ and } w' \text{ are} \\
&\quad \text{undistinguished at node} \\
&\quad A_i (\text{i.e.}, \mathbf{y}_{A_i}(w) = \mathbf{y}_{A_i}(w')), \quad i = 1, 2 \\
\mathcal{B}_i &= \text{the event that } w \text{ and } w' \text{ are} \\
&\quad \text{undistinguished at node} \\
&\quad B_i (\text{i.e.}, \mathbf{y}_{B_i}(w) = \mathbf{y}_{B_i}(w')), \quad i = 1, 2 \\
\mathcal{D} &= \text{the event that } w \text{ and } w' \text{ are} \\
&\quad \text{undistinguished at node} \\
&\quad D (\text{i.e.}, \mathbf{y}_D(w) = \mathbf{y}_D(w')). \quad (43)
\end{aligned}$$

We now have

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{A}_2, \mathcal{B}_2, \mathcal{D}, \mathcal{A}_1^c, \mathcal{B}_1^c\} \\
&= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2, \mathcal{A}_1^c | \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D}, \mathcal{B}_1^c | \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} \\
&\leq \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} \\
&= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{B}_2\} \quad (44)
\end{aligned}$$

where the last step is true since there is an independent random mapping at each node and we have the following Markov structure in the network

$$X_S \to (Y_{A_1}, Y_{A_2}) \to (Y_{B_1}, Y_{B_2}) \to Y_D. \quad (45)$$

As the source does a random linear mapping of the message onto $\mathbf{x}_S(w)$, the probability of $\mathcal{A}_2$ is

$$\begin{aligned}
\mathbb{P}\{\mathcal{A}_2\} &= \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{S,A_2})(\mathbf{x}_S(w) - \mathbf{x}_S(w')) = \mathbf{0}\right\} \\
&= 2^{-T\text{rank}(\mathbf{G}_{S,A_2})} \quad (46)
\end{aligned}$$

because the random mapping given in (40) induces independent uniformly distributed $\mathbf{x}_S(w), \mathbf{x}_S(w')$. Here, $\otimes$ is the Kronecker matrix product.[2] Now, in order to analyze the second probability, we see that $\mathcal{A}_2$ implies $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, i.e., the *same* signal is sent under both w, w'. Also if $\mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w')$, then the random mapping given in (40) induces independent uniformly distributed $\mathbf{x}_{A_1}(w), \mathbf{x}_{A_1}(w')$ Therefore, we get

$$\begin{aligned}
&\mathbb{P}\left\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\right\} \\
&\quad = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{A_1,B_2})(\mathbf{x}_{A_1}(w) - \mathbf{x}_{A_1}(w')) = \mathbf{0}\right\} \\
&\quad = 2^{-T\text{rank}(\mathbf{G}_{A_1,B_2})}. \quad (47)
\end{aligned}$$

Similarly, we get

$$\begin{aligned}
&\mathbb{P}\left\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{B}_2\right\} \\
&\quad = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{B_1,D})(\mathbf{x}_{B_1}(w) - \mathbf{x}_{B_1}(w')) = \mathbf{0}\right\} \\
&\quad = 2^{-T\text{rank}(\mathbf{G}_{B_1,D})}. \quad (48)
\end{aligned}$$

Putting these together we see that in (42), for the network in Fig. 15, we have,

$$\begin{aligned}
\mathcal{P} &\leq 2^{-T\text{rank}(\mathbf{G}_{S,A_2})} 2^{-T\text{rank}(\mathbf{G}_{A_1,B_2})} 2^{-T\text{rank}\left(\mathbf{G}_{B_1,D}\right)} \\
&= 2^{-T\{\text{rank}(\mathbf{G}_{S,A_2}) + \text{rank}(\mathbf{G}_{A_1,B_2}) + \text{rank}(\mathbf{G}_{B_1,D})\}}. \quad (49)
\end{aligned}$$

Note that since

$$\mathbf{G}_{\Omega,\Omega^c} = \begin{bmatrix} \mathbf{G}_{S,A_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{A_1,B_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}_{B_1,D} \end{bmatrix}$$

the upper bound for $\mathcal{P}$ in (49) is exactly $2^{-T\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}$. Therefore, by substituting this back into (42) and (41), we get

$$P_e \leq 2^{RT} |\Lambda_D| 2^{-T \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c})} \quad (50)$$

which can be made as small as desired if $R < \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c})$, which is the result claimed in Theorem 4.3.

[2]If A is an m-by-n matrix and B is a p-by-q matrix, then the Kronecker product $A \otimes B$ is the mp-by-nq block matrix $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$

$$\mathbb{P}\{w \to w'\} = \sum_{\Omega \in \Lambda_D} \mathbb{P}\underbrace{\{\text{Nodes in } \Omega \text{ can distinguish } w, w' \text{ and nodes in } \Omega^c \text{cannot}\}}_{\mathcal{P}} \quad (42)$$

3) Proof of Theorems 4.3 and 4.4 for General Layered Networks: Consider the message $w = w_1$ transmitted by the source at block $k = 1$. The message w will be mistaken for another message w' only if the received signal $\mathbf{y}_D(w)$ under w is the same as that would have been received under w'. Hence, the probability of error at destination D can be upper bounded by,

$$P_e \leq 2^{RT}\mathbb{P}\{w \to w'\} = 2^{RT}\mathbb{P}\{\mathbf{y}_D(w) = \mathbf{y}_D(w')\}. \quad (51)$$

Similar to Section V-A2, we can write (52), shown at the bottom of the page. For any such cut Ω, define the following sets:

- $L_l(\Omega)$: the nodes that are in Ω and are at layer l (for example $S \in L_1(\Omega)$).
- $R_l(\Omega)$: the nodes that are in Ω^c and are at layer l (for example $D \in R_{l_D}(\Omega)$).

We now define the following events:

- $\mathcal{L}_l$: Event that the nodes in L_l can distinguish between w and w', i.e., $\mathbf{y}_{L_l}(w) \neq \mathbf{y}_{L_l}(w')$.
- $\mathcal{R}_l$: Event that the nodes in R_l cannot distinguish between w and w', i.e., $\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w')$.

Similar to Section V-A2, we can write

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1}, l = 2, \ldots, l_D\} && (53)\\
&= \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1} \,|\, \mathcal{R}_j, \mathcal{L}_{j-1}, j = 2, \ldots, l-1\} && (54)\\
&\leq \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l \,|\, \mathcal{R}_j, \mathcal{L}_j, j = 2, \ldots, l-1\} && (55)\\
&\overset{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} && (56)
\end{aligned}$$

where (a) is true due to the Markovian nature of the layered network. Note that as in the example, all nodes in R_{l-1} transmit the same signal under both w and w' (i.e., $\mathbf{x}_j(w) = \mathbf{x}_j(w')$, $\forall j \in R_{l-1}$). Therefore, just as in Section V-A2, we see that i.e.,

$$\begin{aligned}
&\mathbb{P}\{\mathcal{R}_l \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\}\\
&\quad= \mathbb{P}\left\{\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w') \,|\, \mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w'),\right.\\
&\qquad\left.\mathbf{y}_{R_{l-1}}(w) = \mathbf{y}_{L_{R-1}}(w')\right\}\\
&\quad= \mathbb{P}\left\{\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w') \,|\, \mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w'),\right.\\
&\qquad\left.\mathbf{x}_{R_{l-1}}(w) = \mathbf{x}_{L_{R-1}}(w')\right\}\\
&\quad= \mathbb{P}\left\{\left(\mathbf{I}_T \otimes \mathbf{G}_{L_{l-1},R_l}\right)\left(\mathbf{x}_{L_{l-1}}(w) - \mathbf{x}_{L_{l-1}}(w')\right) = \mathbf{0} \,|\right.\\
&\qquad\left.\mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w')\right\}\\
&\quad\overset{(a)}{=} 2^{-T\mathrm{rank}(\mathbf{G}_{L_{l-1},R_l})}.
\end{aligned}$$

where $\mathbf{G}_{L_{l-1},R_l}$ is the transfer matrix from transmitted signals in L_{l-1} to the received signals in R_l. Step (a) is true since $\mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w')$, and hence, the random mapping given in (40) induces independent uniformly distributed $\mathbf{x}_{L_{l-1}}(w), \mathbf{x}_{L_{l-1}}(w')$.

Therefore, we get

$$\mathcal{P} \leq \prod_{l=2}^{d} 2^{-T\mathrm{rank}(\mathbf{G}_{L_{l-1},R_l})} = 2^{-T\mathrm{rank}(\mathbf{G}_{\Omega,\Omega^c})}. \quad (57)$$

By substituting this back into (52) and (51), we see that

$$P_e \leq 2^{RT}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D}\mathrm{rank}(\mathbf{G}_{\Omega,\Omega^c})} \quad (58)$$

which can be made as small as desired if $R < \min_{\Omega\in\Lambda_D}\mathrm{rank}(\mathbf{G}_{\Omega,\Omega^c})$, which is the result claimed in Theorem 4.3 for layered networks.

To prove Theorem 4.4 for layered networks, we note that for *any* destination $D \in \mathcal{D}$, the probability of error expression in (58) holds. Therefore, if *all* receivers in $\mathcal{D}$ have to be able to decode the message, then an error occurs if any of them fails to decode. Therefore, using the union bound and (58), we can bound this error probability as

$$\begin{aligned}
P_e &\leq 2^{RT}\sum_{D\in\mathcal{D}}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D}}\\
&\leq 2^{RT}2^{|\mathcal{V}|}2^{-T\min_{D\in\mathcal{D}}\min_{\Omega\in\Lambda_D}\mathrm{rank}(\mathbf{G}_{\Omega,\Omega^c})}, \quad (59)
\end{aligned}$$

which clearly goes to zero as long as $R < \min_{D\in\mathcal{D}}\min_{\Omega\in\Lambda_D}\mathrm{rank}(\mathbf{G}_{\Omega,\Omega^c})$, which is the result claimed in Theorem 4.4 for layered networks.

Therefore, we have proved a special case of Theorem 4.4 for layered networks.

B. Arbitrary Networks (Not Necessarily Layered)

Given the proof for layered networks with equal path lengths, we are ready to tackle the proof of Theorem 4.3 and Theorem 4.4 for general relay networks. The ingredients are developed below.

We first unfold the network $\mathcal{G}$ over time to create a layered network. The idea is to unfold the network to K stages such that ith stage represents what happens in the network during $(i-1)T$ to $iT-1$ symbol times. More concretely, the K time-steps unfolded network, $\mathcal{G}^{(K)}_{\mathrm{unf}} = (\mathcal{V}^{(K)}_{\mathrm{unf}}, \mathcal{E}^{(K)}_{\mathrm{unf}})$, is constructed from $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ as follows:

- The network has $K+2$ stages (numbered from 0 to $K+1$).
- Stage 0 has only node $S[0]$ and stage $K+1$ has only node $D[K+1]$. $S[0]$ and $D[K+1]$ respectively represent the source and the destination in $\mathcal{G}^{(K)}_{\mathrm{unf}}$.

$$\mathbb{P}\{w \to w'\} = \sum_{\Omega\in\Lambda_D} \mathbb{P}\underbrace{\{\text{Nodes in } \Omega \text{ can distinguish } w, w' \text{ and nodes in } \Omega^c \text{ cannot}\}}_{\mathcal{P}} \quad (52)$$

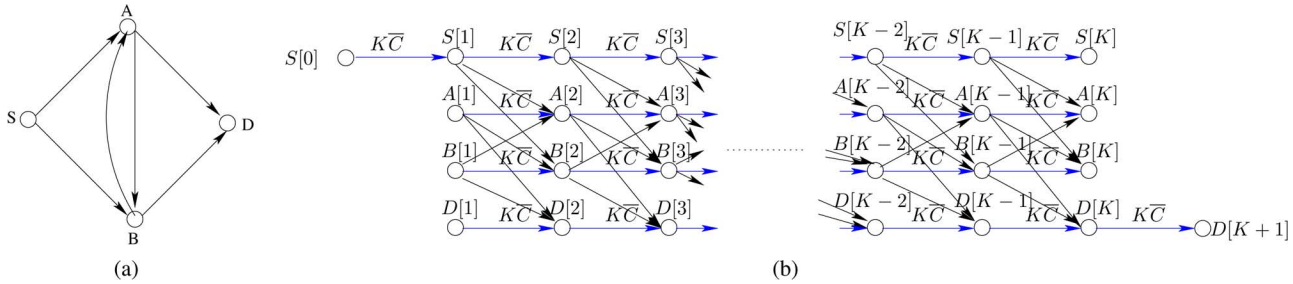

Fig. 16. An example of a general deterministic network with unequal paths from S to D is shown in (*a*). The corresponding unfolded network is shown in (*b*).

- Each node $v \in \mathcal{V}$ appears at stage i as a relay denoted by $v[i]$, $i = 1, \ldots, K$.

Also, the links in $\mathcal{G}_{\text{unf}}^{(K)}$ are as follows:

- There are wired links (i.e., links that are orthogonal to all other transmissions in the network) of capacity $K\bar{C}$, where $\bar{C} = \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c})$ is the min-cut value of $\mathcal{G}$, between
  1) $(S[0], S[1])$ and $(D[K], D[K+1])$.
  2) $(v[i], v[i+1])$, for all $v \in \mathcal{V}$ and $1 \leq i < K$.
- Node $v[i]$ is connected to node $w[i+1]$ with the linear finite-field deterministic channel of the original network $\mathcal{G}$, for all $(v, w) \in \mathcal{E}$, $v \neq w$.

The transmit vector of node $v[i] \in \mathcal{V}_{\text{unf}}^{(K)}$ is denoted by the pair $(\mathbf{x}_{v[i]}^{(1)}, \mathbf{x}_{v[i]}^{(2)})$, where $\mathbf{x}_{v[i]}^{(1)} \in \mathbb{F}_2^{K\bar{C}}$ and $\mathbf{x}_{v[i]}^{(2)} \in \mathbb{F}_2^{q}$ (q is the size of the vectors in the linear finite field network) are respectively the inputs of the wired and the linear finite-field channels. Intuitively, the wired channels represent the memory at each node. Furthermore, the cut-set bound on the capacity of $\mathcal{G}_{\text{unf}}^{(K)}$ is denoted by $\bar{C}_{\text{unf}}^{(K)}$, i.e.,

$$\bar{C}_{\text{unf}}^{(K)} = \min_{\Omega_{\text{unf}} \in \Lambda_{\text{unf}}} \text{rank}\left(\mathbf{G}_{\Omega_{\text{unf}}, \Omega_{\text{unf}}^c}\right) \tag{60}$$

where the minimum is taken over all cuts Ω_{unf} in $\mathcal{G}_{\text{unf}}^{(K)}$.

For example in Fig. 16(a) a network with unequal paths from S to D is shown. Fig. 16(b) shows the unfolded form of this network.

We now prove the following lemma.

Lemma 5.1: Any communication rate $R < \frac{1}{K}\bar{C}_{\text{unf}}^{(K)}$ is achievable in $\mathcal{G}$, where $\bar{C}_{\text{unf}}^{(K)}$ is defined in (60).

Proof: Note that $\mathcal{G}_{\text{unf}}^{(K)}$ is a layered linear finite-field network. Therefore, by our result of Section V-A3, we can achieve any rate $R_{\text{unf}} < \bar{C}_{\text{unf}}^{(K)}$ in $\mathcal{G}_{\text{unf}}^{(K)}$. In particular, it is achieved by the encoding strategy described in Section V-A1, in which each node $v[i] \in \mathcal{V}_{\text{unf}}^{(K)}$, $i = 1, \ldots, K$, operates over blocks of size T symbols and transmits $\mathbf{x}_{v[i]}^{(1)} = \mathbf{F}_{v[i]}^{(1)} \mathbf{y}_{v[i]}$ and $\mathbf{x}_{v[i]}^{(2)} = \mathbf{F}_{v[i]}^{(2)} \mathbf{y}_{v[i]}$ respectively over the wired and the linear finite-field channels.

Now, we can implement the scheme in $\mathcal{G}$ by using K blocks of size T symbols. The construction is as follows:

- The source S transmits $\mathbf{x}_{S[i]}^{(2)}$ at block i, $i = 1, \ldots, K$.
- Each node $v \in \mathcal{V}$, $v \notin \{S, D\}$, transmits $\mathbf{x}_{v[i]}^{(2)}$ and puts $\mathbf{x}_{v[i]}^{(1)}$ in its memory at block i, $i = 1, \ldots, K$ (note that this is possible, because $\mathbf{x}_{v[i]}^{(1)}$ and $\mathbf{x}_{v[i]}^{(2)}$ are only a function of the received signal at node v in the previous block and the the signal stored in the memory of node v at the beginning of block i).

Finally, the destination decodes based on $\mathbf{x}_{D[K]}^{(1)}$, which is a function of the received signal at the destination during the K blocks. Therefore, the rate $\frac{1}{K}R_{\text{unf}}$ is achievable in $\mathcal{G}$ and the proof is complete. ■

Now, if we show that $\lim_{K\to\infty} \frac{1}{K}\bar{C}_{\text{unf}}^{(K)} = \bar{C}$, then by using Lemma 5.1, the proof of Theorem 4.3 will be complete. We will show this next.

Lemma 5.2:

$$\lim_{K\to\infty} \frac{1}{K}\bar{C}_{\text{unf}}^{(K)} = \bar{C} \tag{61}$$

where $\bar{C} = \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c})$ and $\bar{C}_{\text{unf}}^{(K)}$ is defined in (60).

Proof: Any cut $\Omega_{\text{unf}} \in \Lambda_{\text{unf}}$ is a subset of nodes in $\mathcal{G}_{\text{unf}}^{(K)}$ such that $S[0] \in \Omega_{\text{unf}}$ and $D[K+1] \in \Omega_{\text{unf}}^c$. Now for any cut Ω_{unf} we define

$$\mathcal{V}[i] = \{v \in \mathcal{V} | v[i] \in \Omega_{\text{unf}}\}, \quad i = 0, \ldots K+1. \tag{62}$$

In other words, $\mathcal{V}[i]$ is the set of nodes of $\mathcal{G}$ such that at stage i they appear in Ω_{unf}.

Every cut $\Omega \in \Lambda_D$ in the original network $\mathcal{G}$ corresponds to a cut in the unfolded network $\mathcal{G}_{\text{unf}}^{(K)}$, by choosing $\mathcal{V}[1] = \cdots = \mathcal{V}[K] = \Omega$. Also, the value of such a "steady" cut is $K\text{rank}(\mathbf{G}_{\Omega,\Omega^c})$, thereby

$$\bar{C}_{\text{unf}}^{(K)} \leq K\bar{C}. \tag{63}$$

Therefore, we need to only focus on cuts whose values are smaller than $K\bar{C}$. We will next identify other cuts which have value larger than $K\bar{C}$, in order to reduce the set of cuts to consider for $\bar{C}_{\text{unf}}^{(K)}$.

We claim that the value of any cut $\Omega_{\text{unf}} \in \Lambda_{\text{unf}}$ is at least $K\bar{C}$, if the following is *not* satisfied:

$$\mathcal{V}[1] \subseteq \mathcal{V}[2] \subseteq \cdots \subseteq \mathcal{V}[K] \tag{64}$$

The reason is that if $\mathcal{V}[1] \subseteq \mathcal{V}[2] \subseteq \cdots \subseteq \mathcal{V}[K]$ is *not* true, then there exists a node $v \in \mathcal{V}$ and a stage j $(1 \leq j < K)$ such that

$$v[j] \in \mathcal{V}[j] \quad \text{and} \quad v[j+1] \notin \mathcal{V}[j+1]. \tag{65}$$

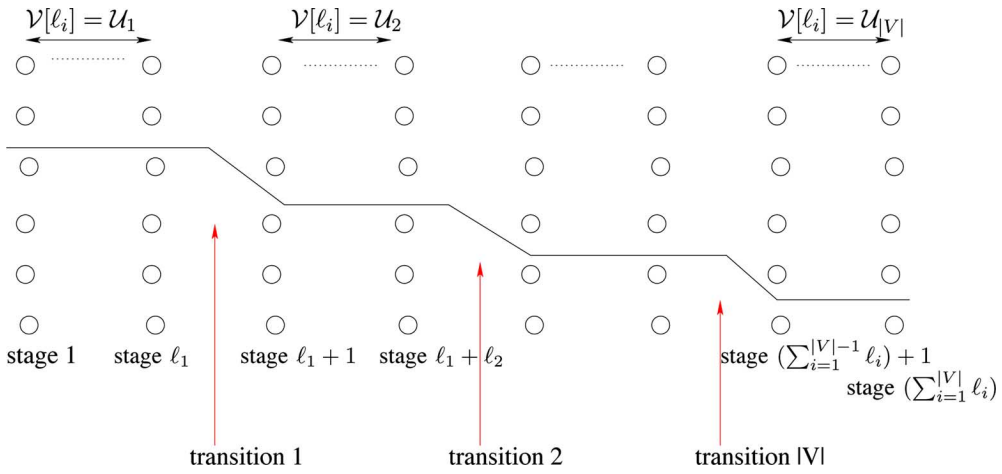

Fig. 17. Illustration of cuts in $\mathcal{G}_{\text{unf}}^{(K)}$ which can have a value smaller than $K\bar{C}$.

If this happens, then the edge $(v[j], v[j+1])$, which has capacity $K\bar{C}$, traverses from Ω_{unf} to Ω_{unf}^c; hence, the cut-value (i.e., $\text{rank}(\mathbf{G}_{\Omega_{\text{unf}},\Omega_{\text{unf}}^c})$) becomes at least $K\bar{C}$.

Hence, we only need to focus on cuts, Ω_{unf} that satisfy (64), i.e., contain an increasing set of nodes at the stages. Since there are total of $|V|$ nodes in $\mathcal{G}$, we can have at most $|V|$ transitions in the size of $\mathcal{V}[i]$s. Now, using the notation in Fig. 17 and the fact that the network is layered, for any cut $\Omega_{\text{unf}} \in \Lambda_{\text{unf}}$ satisfying (64) we can write

$$\begin{aligned}
&\text{rank}\left(\mathbf{G}_{\Omega_{\text{unf}},\Omega_{\text{unf}}^c}\right)\\
&\quad=\sum_{i=0}^{K}\text{rank}\left(\mathbf{G}_{\mathcal{V}[i],\mathcal{V}[i+1]^c}\right)\\
&\quad=\sum_{i=1}^{|V|}(\ell_i-1)\text{rank}\left(\mathbf{G}_{\mathcal{U}_i,\mathcal{U}_i^c}\right)+\sum_{i=1}^{|V|-1}\text{rank}\left(\mathbf{G}_{\mathcal{U}_i,\mathcal{U}_{i+1}^c}\right)\\
&\qquad+\text{rank}\left(\mathbf{G}_{\mathcal{V}[0],\mathcal{U}_1^c}\right)+\text{rank}(\mathbf{G}_{\mathcal{U}_K,\mathcal{V}[K+1]^c})\\
&\quad\geq\sum_{i=1}^{|V|}(\ell_i-1)\text{rank}\left(\mathbf{G}_{\mathcal{U}_i,\mathcal{U}_i^c}\right)\\
&\quad\overset{(a)}{\geq}\left(\sum_{i=1}^{|V|}(\ell_i-1)\right)\bar{C}\\
&\quad\overset{(b)}{=}(K-|V|)\bar{C}
\end{aligned}$$

where (a) follows because $\text{rank}(\mathbf{G}_{\Omega,\Omega^c})\geq\bar{C}$, for any cut Ω in $\mathcal{G}$; and (b) is because there are at most $|V|$ transitions implying that $\sum_{i=1}^{|V|}(\ell_i-1))=(K-|V|)$. As a result

$$\bar{C}_{\text{unf}}^{(K)}\geq(K-|V|)\bar{C}. \tag{66}$$

Combining (63) and (64), we get

$$\lim_{K\to\infty}\frac{1}{K}\bar{C}_{\text{unf}}^{(K)}=\bar{C}. \tag{67}$$

■

This combined with Lemma 5.1 completes the proof of Theorem 4.3.[3]

VI. Gaussian Relay Networks

So far, we have focused on deterministic relay networks. As we illustrated in Sections II and III, linear finite-field deterministic model captures some (but not all) aspects of the high SNR behavior of the Gaussian model. Therefore, we have some hope to be able to translate the intuition and the techniques used in the deterministic analysis to obtain approximate results for Gaussian relay networks. This is what we will accomplish in this section.

Theorem 4.5 is the main result for Gaussian relay networks and this section is devoted to proving it. The proof of the result for layered network is done in Section V. We extend the result to an arbitrary network by expanding the network over time, as done in Section V. We first prove the theorem for the single antenna case, then at the end we extend it to the multiple antenna scenario.

A. Layered Gaussian Relay Networks

In this section we prove Theorem 4.5 for the special case of layered networks, where all paths from the source to the destination in $\mathcal{G}$ have equal length.

1) Proof Illustration: Our proof has two steps. In the first step we propose a relaying strategy, which is similar to our strategy for deterministic networks, and show that by operating over a large block, it is possible to achieve an end-to-end mutual information which is within a constant gap to the cut-set upper bound. Therefore, the relaying strategy creates an inner code which provides certain end-to-end mutual information between the transmit signal at the source and the received signal at the destination. Each symbol of this inner code is a block. In the next step, we use an outer code to map the message to multiple inner code symbols and send them to the destination. By coding over many such symbols, it is possible to achieve a reliable communication rate arbitrarily close to the mutual information of the

[3]An alternate proof of the same result was given in [22]. In that proof, only the previous received block was used by the relays, instead of the larger number of blocks used above. However, we needed to use the sub-modularity properties of entropy to demonstrate the performance of that scheme [22].

inner code, and hence, the proof is complete. The system diagram of our coding strategy is illustrated in Fig. 18.

We now explicitly describe our encoding strategy

2) Encoding for Layered Gaussian Relay Networks: We first define a quantization operation.

Definition 6.1: The quantization operation $[\cdot]: \mathbb{C} \to \mathbb{Z} \times \mathbb{Z}$ maps a complex number $c = x + iy$ to $[c] = ([x], [y])$, where $[x]$ and $[y]$ are the closest integers to x and y, respectively. Since the Gaussian noise at all receive antennas has variance 1, this operation is basically scalar quantization at noise-level.

As shown in Fig. 18, the encoding consists of an inner code and an outer code:

a) Inner Code: Each symbol of the inner code is represented by $u \in \{1, \ldots, 2^{R_{\text{in}}T}\}$, where T and R_{in} are respectively the block length and the rate of the inner code. The source node S generates a set of $2^{R_{\text{in}}T}$ independent complex Gaussian codewords of length T with components distributed as i.i.d. $\mathcal{CN}(0,1)$, denoted by $\mathcal{T}_{x_S}$. At relay node i, there is also a random mapping $F_i: (\mathbb{Z}^T, \mathbb{Z}^T) \to \mathcal{T}_{x_i}$ which maps each quantized received signal vector of length T independently into an i.i.d. $\mathcal{CN}(\mathbf{0}, 1)$ random vector of length T. A particular realization of F_i is denoted by f_i. Summarizing:

- Source: maps each inner code symbol $u \in \{1, \ldots, 2^{R_{\text{in}}T}\}$ to $F_S(u) \in \mathcal{T}_{x_S}$.
- Relay i: receives $\mathbf{y}_i$ of length T. Quantizes it to $[\mathbf{y}_i]$. Then maps it to $F_i([\mathbf{y}_i]) \in \mathcal{T}_{x_i}$.

b) Outer Code: The message is encoded by the source into N inner code symbols, $u_1, \ldots, u_N$. Each inner code symbol is then sent via the inner code over T transmission times, giving an overall transmission rate of R. The received signal at the destination, corresponding to inner code symbol u_i, is denoted by $\mathbf{y}_{D,i}$, $i = 1, \ldots, N$.

Now, given the knowledge of all the encoding functions F_i's at the relays and quantized received signals $[\mathbf{y}_{D,1}], \ldots, [\mathbf{y}_{D,N}]$, the destination attempts to decode the message sent by the source.

3) Proof of Theorem 4.5 for Layered Networks: Our first goal is to lower bound the average end-to-end mutual information, averaged over the random mappings $F_{\mathcal{V}} = \{F_i : i \in \mathcal{V}\}$, achieved by the inner code defined in Subsection VI-A2.

Note that

$$\frac{1}{T} I(u; [\mathbf{y}_D] | F_{\mathcal{V}}) \geq \frac{1}{T} I(u; [\mathbf{y}_D] \,|\, \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) - \frac{1}{T} H([\mathbf{y}_D] | u, F_{\mathcal{V}}) \quad (68)$$

where $\mathbf{z}_{\mathcal{V}}$ is the vector of the channel noises at all nodes in the network. The first term on the right hand side of (68) is the average end-to-end mutual information conditioned on the noise vector. Once we condition on a noise vector, the network turns into a deterministic network. We use an analysis technique similar to the one we used for linear deterministic relay networks to upper bound the probability that the destination will confuse an inner code symbol with another and then use Fano's inequality to lower bound the end-to-end mutual information. This is done in Lemma 6.3. The second term on the RHS of (68) is the average entropy of the received signal conditioned on the source's transmit signal, and is upper bounded in Lemma 6.5. This term represents roughly the penalty due to noise-forwarding at the relay, and is proportional to the number of relay nodes.

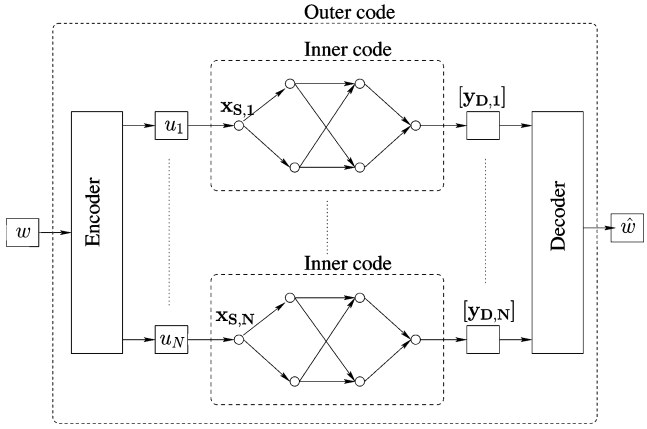

Fig. 18. System diagram.

Definition 6.2: We define

$$\bar{C}_{i.i.d.} \triangleq \min_{\Omega} I(x_\Omega; y_{\Omega^c} | x_{\Omega^c}) \quad (69)$$

where x_i, $i \in \mathcal{V}$, are i.i.d. $\mathcal{CN}(0,1)$ random variables.

Lemma 6.3: Assume all nodes perform the operation described in Subsection VI-A2 (a) and the inner code symbol U is distributed uniformly over $\{1, \ldots, 2^{R_{\text{in}}T}\}$. Then

$$I(u; [\mathbf{y}_D] \,|\, \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) \geq R_{\text{in}}T - \left(1 + \min\left\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C}_{iid} - |\mathcal{V}| - R_{\text{in}})}\right\} R_{\text{in}}T\right)$$

where $\bar{C}_{iid}$ is defined in Definition 6.2.

Proof: Consider a fixed noise realization in the network $\mathbf{z}_{\mathcal{V}} = \mathbf{a}$. Suppose the destination attempts to detect the transmitted symbol u at the source given the received signal, all the mappings, channel gains, and $\mathbf{a}$. A symbol value u will be mistaken for another value u' only if the received signal $[\mathbf{y}_D(u)]$ under u is the same as what would have been received under u'. This leads to a notion of *distinguishability* for a fixed $\mathbf{a}$, which is that symbol values u, u' are distinguishable at any node j if $[\mathbf{y}_j(u)] \neq [\mathbf{y}_j(u')]$. Hence [see (70), shown at the bottom of the page]. For any cut $\Omega \in \Lambda_D$, define the following sets:

$$\mathbb{P}\{u \to u' \,|\, \mathbf{z}_{\mathcal{V}} = \mathbf{a}\} = \mathbb{P} \sum_{\Omega \in \Lambda_D} \underbrace{\{\text{Nodes in } \Omega \text{ can distinguish } u, u' \text{ and nodes in } \Omega^c \text{ cannot} | \mathbf{z}_{\mathcal{V}} = \mathbf{a}\}}_{\mathcal{P}} \quad (70)$$

- $L_l(\Omega)$: the nodes that are in Ω and are at layer l (for example $S \in L_1(\Omega)$).
- $R_l(\Omega)$: the nodes that are in Ω^c and are at layer l (for example $D \in R_{l_D}(\Omega)$).

We also define the following events:

- $\mathcal{L}_l$: Event that the nodes in L_l can distinguish between u and u', i.e., $[\mathbf{y}_{L_l}(u)] \neq [\mathbf{y}_{L_l}(u')]$.
- $\mathcal{R}_l$: Event that the nodes in R_l can not distinguish between u and u', i.e., $[\mathbf{y}_{R_l}(u)] = [\mathbf{y}_{R_l}(u')]$.

Note that the source node by definition distinguishes between the two distinct messages u, u', i.e., $\mathbb{P}\{\mathcal{L}_1\} = 1$

$$\begin{aligned}\mathcal{P} &= \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1}, l = 2, \ldots, l_D \,|\, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \\ &= \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1} \,|\, \mathcal{R}_j, \mathcal{L}_{j-1}, j = 2, \ldots, l-1, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \\ &\leq \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l \,|\, \mathcal{R}_j, \mathcal{L}_j, j = 2, \ldots, l-1, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \\ &\overset{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \\ &= \prod_{l=2}^{l_D} \mathbb{P}\{[\mathbf{y}_{R_l}(u)] = [\mathbf{y}_{R_l}(u')] \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \quad (71)\end{aligned}$$

where (a) is true due to the Markov structure in the layered network.

Note that if $\mathbf{A}$ and $\mathbf{B}$ are complex $m \times n$ matrices, then

$$[\mathbf{A}_{i,j}] = [\mathbf{B}_{i,j}], \forall i, j \;\Rightarrow ||\mathbf{A} - \mathbf{B}||_\infty \leq \sqrt{2}n. \quad (72)$$

Therefore, by (71) and (72), we have

$$\begin{aligned}\mathcal{P} &\leq \prod_{l=2}^{l_D} \mathbb{P}\{||\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')||_\infty \\ &\leq \sqrt{2} \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \\ &\overset{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{||\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')||_\infty \leq \sqrt{2} \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} \quad (73)\end{aligned}$$

where (a) is true since conditioned on $\mathcal{R}_{l-1}, \mathcal{L}_{l-1}$ the distribution of $\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')$ does not depend on the noise (due to the random mapping).

By defining $\mathbf{H}_l$ to be the transfer matrix from the left side of the cut at stage $l-1$ to the right side of the cut at stage l (i.e., the MIMO channel from L_{l-1} to R_l), we have

$$\begin{aligned}\mathcal{P} &\overset{(73)}{\leq} \prod_{l=2}^{l_D} \mathbb{P}\{||\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')||_\infty \leq \sqrt{2} \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} \\ &= \prod_{l=2}^{l_D} \mathbb{P}\{\forall 1 \leq j \leq T : ||\mathbf{H}_l(\mathbf{x}_{L_{l-1},j}(u) - \mathbf{x}_{L_{l-1},j}(u'))||_\infty \\ &\leq \sqrt{2} \,|\, \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} \\ &\overset{(b)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\forall 1 \leq j \leq T : ||\mathbf{H}_l(\mathbf{x}_{L_{l-1},j}(u) - \mathbf{x}_{L_{l-1},j}(u'))||_\infty \\ &\leq \sqrt{2} \,|\, \mathcal{L}_{l-1}\} \quad (74)\end{aligned}$$

where (b) is true since the nodes in $R_{l-1}(\Omega)$ transmit the same codeword under both u and u'.

Since $\mathbf{x}_{L_{l-1}}(u) \neq \mathbf{x}_{L_{l-1}}(u')$, due to the random mapping, $\mathbf{x}_{L_{l-1}}(u)$ and $\mathbf{x}_{L_{l-1}}(u')$ are two independent random vectors with i.i.d. $\mathcal{CN}(0,1)$ elements. Therefore, their difference is a random vector with i.i.d. $\mathcal{CN}(0,2)$ elements. Now, we state the following Lemma which is proved in Appendix D.

Lemma 6.4: Assume $[\tilde{x}_{i,1}, \ldots, \tilde{x}_{i,T}], i = 1, \ldots, m$, are i.i.d. vectors of length T with i.i.d. $\mathcal{CN}(0,2)$ elements, and $\mathbf{H} \in \mathbb{C}^{n \times m}$ is an $n \times m$ matrix. Then

$$\begin{aligned}&\mathbb{P}\{\forall 1 \leq j \leq T : ||\mathbf{H}[\tilde{x}_{1,j}, \ldots, \tilde{x}_{m,j}]^t||_\infty \leq \sqrt{2}\} \\ &\qquad \leq 2^{-T(I(\mathbf{x};\mathbf{Hx}+\mathbf{z}) - \min(m,n))} \quad (75)\end{aligned}$$

where $\mathbf{x}$ and $\mathbf{z}$ are i.i.d. complex unit variance Gaussian vectors of length m and n respectively.

By applying Lemma 6.4 to (74), we get

$$\begin{aligned}&\mathbb{P}\{\forall 1 \leq j \leq T : ||\mathbf{H}_l(\mathbf{x}_{L_{l-1},j}(u) - \mathbf{x}_{L_{l-1},j}(u'))||_\infty \\ &\quad \leq \sqrt{2} \,|\, \mathcal{L}_{l-1}\} \\ &\qquad \leq 2^{-T(I(\mathbf{x}_{L_{l-1}};\mathbf{y}_{R_l} \,|\, \mathbf{x}_{R_{l-1}}) - \min(|L_{l-1}|, |R_l|))} \quad (76)\end{aligned}$$

where $x_i, i \in \mathcal{V}$, are i.i.d. with Gaussian distribution. Hence

$$\begin{aligned}\mathcal{P} &\leq \prod_{l=2}^{l_D} 2^{-T(I(\mathbf{x}_{L_{l-1}};\mathbf{y}_{R_l} \,|\, \mathbf{x}_{R_{l-1}}) - \min(|L_{l-1}|, |R_l|))} \\ &\leq 2^{-T(\bar{C}_{iid} - |\mathcal{V}|)} \quad (77)\end{aligned}$$

where $\bar{C}_{iid}$ is defined in Definition 6.2.

The average probability of symbol detection error at the destination can be upper bounded as

$$P_e = \mathbb{P}\{\hat{u} \neq u \,|\, \mathbf{z}_\mathcal{V} = \mathbf{a}\} \leq 2^{R_{\text{in}}T} \mathbb{P}\{u \to u' | \mathbf{z}_\mathcal{V} = \mathbf{a}\}. \quad (78)$$

By the union bound, we have

$$P_e \leq \sum_{\Omega} 2^{-T(\bar{C}_{iid} - |\mathcal{V}| - R_{\text{in}})} \leq 2^{|\mathcal{V}|} 2^{-T(\bar{C}_{iid} - |\mathcal{V}| - R_{\text{in}})}. \quad (79)$$

Now, using Fano's inequality, we get

$$\begin{aligned}&I(u; [\mathbf{y}_D] \,|\, \mathbf{z}_\mathcal{V} = \mathbf{a}, F_\mathcal{V}) \\ &\quad = H(u) - H(u \,|\, [\mathbf{y}_D], \mathbf{z}_\mathcal{V} = \mathbf{a}, F_\mathcal{V}) \\ &\quad = R_{\text{in}}T - H(u \,|\, [\mathbf{y}_D], \mathbf{z}_\mathcal{V} = \mathbf{a}, F_\mathcal{V}) \\ &\quad = R_{\text{in}}T - \mathbb{E}_{F_\mathcal{V}}[H(u \,|\, [\mathbf{y}_D], \mathbf{z}_\mathcal{V} = \mathbf{a}, F_\mathcal{V} = f_\mathcal{V})] \\ &\quad \overset{\text{Fano}}{\geq} R_{\text{in}}T - (1 + \mathbb{E}_{F_\mathcal{V}} \\ &\qquad [\mathbb{P}\{\hat{u} \neq u | \mathbf{z}_\mathcal{V} = \mathbf{a}, F_\mathcal{V} = f_\mathcal{V})\}] R_{\text{in}}T) \\ &\quad = R_{\text{in}}T - (1 + P_e R_{\text{in}}T) \\ &\quad \geq R_{\text{in}}T - \left(1 + \min\left\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C}_{iid} - |\mathcal{V}|) - R_{\text{in}})}\right\} R_{\text{in}}T\right)\end{aligned}$$

Hence, the proof is complete. ∎

The following lemma, which is proved in Appendix E, bounds the second term on the RHS of (68).

Lemma 6.5: Assume all nodes perform the operation described in Subsection VI-A2 (a). Then

$$H([\mathbf{y}_D]|u, F_{\mathcal{V}})] \leq 12T|\mathcal{V}| \tag{80}$$

The next lemma, which is proved in Appendix F, bounds the gap between $\bar{C}$ and $\bar{C}_{iid}$.

Lemma 6.6: For a Gaussian relay network $\mathcal{G}$

$$\bar{C} - \bar{C}_{iid} < 2|\mathcal{V}| \tag{81}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ and $\bar{C}_{iid}$ is defined in Definition 6.2.

Finally, using Lemmas 6.3, 6.5, and 6.6, we have

Lemma 6.7: Assume all nodes perform the operation described in Subsection VI-A2 (a) and the inner code symbol U is distributed uniformly over $\{1, \ldots, 2^{R_{\text{in}}T}\}$. Then

$$\frac{1}{T} I(u; [\mathbf{y}_D] \,|\, F_{\mathcal{V}}) \geq R_{\text{in}} - 12|\mathcal{V}| - \left(\frac{1}{T} + \min\left\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C} - 3|\mathcal{V}| - R_{\text{in}})}\right\} R_{\text{in}}\right) \tag{82}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$.

Proof: By using (68) and Lemmas 6.3, 6.5, and 6.6, we have

$$\begin{aligned}
&\frac{1}{T} I(u; [\mathbf{y}_D] \,|\, F_{\mathcal{V}}) \\
&\quad \geq \frac{1}{T} I(u; [\mathbf{y}_D] \,|\, \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) - \frac{1}{T} H([\mathbf{y}_D] \,|\, u, F_{\mathcal{V}}) \\
&\quad \overset{\text{Lemma 6.3 and 6.5}}{\geq} \frac{1}{T} \Big(R_{\text{in}} T \\
&\qquad - \Big[1 + \min\Big\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C}_{iid} - |\mathcal{V}|) - R_{\text{in}})}\Big\} \\
&\qquad R_{\text{in}} T\Big] - 12T|\mathcal{V}|\Big) \\
&\quad \overset{\text{Lemma 6.6}}{\geq} R_{\text{in}} - 12|\mathcal{V}| \\
&\qquad - \left(\frac{1}{T} + \min\left\{1, 2^{|\mathcal{V}|} 2^{-T(\bar{C} - 3|\mathcal{V}| - R_{\text{in}})}\right\} R_{\text{in}}\right).
\end{aligned}$$

■

An immediate corollary of this lemma is that by choosing R_{in} arbitrarily close to $\bar{C} - 2|\mathcal{V}|$, and letting T be arbitrary large, for any $\delta > 0$ we get

$$\frac{1}{T} I(u; [\mathbf{y}_D] \,|\, F_{\mathcal{V}}) \geq \bar{C} - 15|\mathcal{V}| - \delta. \tag{83}$$

Therefore, there exists a choice of mappings that provides an end-to-end mutual information close to $\bar{C} - 15|\mathcal{V}|$. Hence, we have created a point-to-point channel from u to $[\mathbf{y}_D]$ with at least this mutual information. We can now use a good outer code to reliably send a message over N uses of this channel (as illustrated in Fig. 18) at any rate up to $\bar{C} - 15|\mathcal{V}|$.

Hence, we get an intermediate proof of Theorem 4.5 for the special case of layered Gaussian relay networks, with single antennas in the network. This is stated below for convenience, and its generalization to arbitrary networks with multiple antennas is given in Section VI-B.

Theorem 6.8: Given a Gaussian relay network $\mathcal{G}$ with a layered structure and single antenna at each node, all rates R satisfying the following condition are achievable

$$R < \bar{C} - \kappa_{\text{Lay}} \tag{84}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in (23), $\kappa_{\text{Lay}} = 15|\mathcal{V}|$ is a constant not depending on the channel gains.

4) Vector Quantization and Network Operation: The network operation can easily be generalized to include vector quantization at each node. Each node in the network generates a transmission Gaussian codebook of length T with components distributed as i.i.d. $\mathcal{CN}(0,1)$. The source operation is as before, it produces a random mapping from messages $w \in \{1, \ldots, 2^{RT}\}$ to its transmit codebook $\mathcal{T}_{x_S}$. We denote this codebook by $\mathbf{x}_S^{(w)}, w \in \{1, \ldots, 2^{TR}\}$. Each received sequence $\mathbf{y}_i$ at node i is quantized to $\hat{\mathbf{y}}_i$ through a Gaussian vector quantizer, with quadratic distortion set to the noise-level. This quantized sequence is randomly mapped onto a transmit sequence $\mathbf{x}_i$ using a random function $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$. This mapping as before is chosen such that each quantized sequence is mapped uniformly at random to a transmit sequence. These transmit sequences are chosen to be in $\mathcal{T}_{x_i}$, which are i.i.d. Gaussian $\mathcal{CN}(0,1)$. We denote the 2^{TR_i} sequences of $\hat{\mathbf{y}}_i$ as $\hat{\mathbf{y}}_i^{(k_i)}, k_i \in \{1, \ldots, 2^{TR_i}\}$. Standard rate-distortion theory tells us that we need $R_i > I(Y_i; \hat{Y}_i)$ for this quantization to be successful, where the reconstruction is chosen such that the quadratic distortion is at the noise-level.[4] Since the uniform random mapping produces $\mathbf{x}_i = f_i(\hat{\mathbf{y}}_i)$, for a quantized value of index k_i, we will denote it by $\hat{\mathbf{y}}_i^{(k_i)}$ and the sequence it is mapped to by $\mathbf{x}_i^{(k_i)} = f_i(\hat{\mathbf{y}}_i^{(k_i)})$. At the destination, we can either employ a maximum-likelihood decoder (for which the mutual information is evaluated), or a typicality decoder (see [20] for more details).

B. General Gaussian Relay Networks (Not Necessarily Layered)

Given the proof for layered networks, we are ready to tackle the proof of Theorem 4.5 for general Gaussian relay networks.

Similar to the deterministic case, we first unfold the network $\mathcal{G}$ over K stages to create a layered network $\mathcal{G}_{\text{unf}}^{(K)}$. The details of the construction are described in Section V-B, except now the linear finite-field channels are replaced by Gaussian channels and the wired links of capacity $K\bar{C}$ are replaced by orthogonal point-to-point Gaussian links of capacity $K\bar{C}$ that do not interfere with the other links in the network, where $\bar{C}$ is defined in (23). We now state the following lemma which is a corollary of Theorem 6.8.

[4]Note that we can be conservative and assume the maximal received power, depending on the maximal channel gains. Since we do not directly convey this quantization index, but just map it forward, this conservative quantization suffices.

Lemma 6.9: All rates R satisfying the following condition are achievable in $\mathcal{G}$

$$R < \frac{1}{K}\bar{C}_{\text{unf}}^{(K)} - \kappa \tag{85}$$

where $\bar{C}_{\text{unf}}^{(K)}$ is the cut-set upper bound on the capacity of $\mathcal{G}_{\text{unf}}^{(K)}$, and $\kappa = 15(|\mathcal{V}| + \frac{2}{K})$.

Proof: $\mathcal{G}_{\text{unf}}^{(K)}$ is a layered network. Therefore, by Theorem 6.8, all rates R_{unf}, satisfying the following condition are achievable in $\mathcal{G}_{\text{unf}}^{(K)}$:

$$R_{\text{unf}} < \bar{C}_{\text{unf}}^{(K)} - \kappa_{\text{unf}} \tag{86}$$

where $\kappa_{\text{unf}} = 15|\mathcal{V}_{\text{unf}}^{(K)}|$. But the number of nodes at each stage of $\mathcal{G}_{\text{unf}}^{(K)}$ is exactly $|\mathcal{V}|$ (other than stage 0 and $K+1$ which respectively contain the source, $S[0]$, and the destination, $D[K+1]$). Hence, $\kappa_{\text{unf}} = 15(K|\mathcal{V}| + 2)$. Now, similar to the proof of Lemma 5.1, our achievability scheme (described in Section VI-A2) can be implemented in $\mathcal{G}$ by using K blocks of size T symbols. Therefore, we can achieve $\frac{1}{K}R_{\text{unf}}$ in $\mathcal{G}$ and the proof is complete. ∎

Similar to the deterministic case, it is easy to see that

$$\bar{C}_{\text{unf}}^{(K)} \geq (K - |\mathcal{V}|)\bar{C}. \tag{87}$$

Hence, by Lemma 6.9 and (87), we can achieve all rates up to

$$R < \frac{K - |\mathcal{V}|}{K}\bar{C} - \kappa \tag{88}$$

where $\kappa = 15(|\mathcal{V}| + \frac{2}{K})$. By letting $K \to \infty$ the proof of Theorem 4.5 is complete.

To prove Theorem 4.6, i.e., the multicast scenario, we just need to note that if all relays will perform exactly the same strategy then by our theorem, each destination, $D \in \mathcal{D}$, will be able to decode the message with low error probability as long as the rate of the message satisfies

$$R < \min_{D \in \mathcal{D}} \bar{C}_{i.i.d.,D} - \kappa' \tag{89}$$

where $\kappa' < 15|\mathcal{V}|$ is a constant and as in Definition 6.2 we have $\bar{C}_{i.i.d.,D} = \min_{\Omega \in \Lambda_D} \log|\mathbf{I} + P\mathbf{G}_\Omega\mathbf{G}_\Omega^*|$ is the cut-set bound evaluated for i.i.d. input distributions. Therefore, as long as $R < \bar{C}_{\text{mult}} - \kappa$, where $\kappa < 15|\mathcal{V}|$, all destinations can decode the message, and hence, the theorem is proved.

In the case that we have multiple antennas at each node, the achievability strategy remains the same, except now each node receives a vector of observations from different antennas. We first quantize the received signal of each antenna at the noise level and then map it to another transmit codeword, which is joint across all antennas. The error probability analysis is exactly the same as before. However, the gap between the achievable rate and the cut-set bound will be larger. We can upper bound the gap between $\bar{C}$ and $\bar{C}_{iid}$ by twice the maximum number of degrees of freedom of the cuts, which due to (152) is at most $2\sum_{i=1}^{|\mathcal{V}|} M_i$ (see the last paragraph in Appendix F). Also, by treating each receive antenna as a separate node and applying Lemma 6.5, we get that $H([\mathbf{y}_D]\,|\,u, F_{\mathcal{V}})] \leq 12T\sum_{i=1}^{|\mathcal{V}|} N_i$. Therefore, from our previous analysis we know that the gap is at most $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$ and the theorem is proved when we have multiple antennas at each node.

VII. Connections Between Models

In Section II, we showed that while the linear finite-field channel model captures certain high SNR behaviors of the Gaussian model, it does not capture all aspects. In particular, its capacity is not within a constant gap to the Gaussian capacity for all MIMO channels. A natural question is: is there a deterministic channel model which approximates the Gaussian relay network capacity to within a constant gap?

The proof of the approximation theorem for the Gaussian network capacity in the previous section already provides a partial answer to this question. We showed that, after quantizing all the output at the relays as well as the destination, the end-to-end mutual information achieved by the relaying strategy in the noisy network is close to that achieved when the noise sequences are known at the destination, uniform over all realizations of the noise sequences. In particular, this holds true when the noise sequences are all zero. Since the former has been proved to be close to the capacity of the Gaussian network, this implies that the capacity of the *quantized* deterministic model with

$$\mathbf{y}_j[t] = \left[\sum_{i \in \mathcal{V}} \mathbf{H}_{ij} x_i[t]\right], \quad j = 1, \ldots, |\mathcal{V}| \tag{90}$$

must be *at least* within a constant gap to the capacity of the Gaussian network. It is not too difficult to show that the deterministic model capacity cannot be much larger. We establish all this more formally in the next section, where we call the model in (90) as the *truncated deterministic model*.

A. Connection Between the Truncated Deterministic Model and the Gaussian Model

Theorem 7.1: The capacity of any Gaussian relay network, C_{Gaussian}, and the capacity of the corresponding truncated deterministic model, $C_{\text{Truncated}}$, satisfy the following relationship:

$$|C_{\text{Gaussian}} - C_{\text{Truncated}}| \leq 33|\mathcal{V}|. \tag{91}$$

To prove this theorem we need the following lemma which is proved in Appendix G.

Lemma 7.2: Let G be the channel gains matrix of a $m \times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{x}}$

$$|I(\mathbf{x}; G\mathbf{x} + Z) - I(\mathbf{x}; [G\mathbf{x}])| \leq 19n \tag{92}$$

where $Z = [z_1, \ldots, z_n]$ is a vector of n i.i.d. $\mathcal{CN}(0,1)$ random variables.

Proof (proof of Theorem 7.1): First note that the value of any cut in the network is the same as the mutual information of a MIMO system. Therefore, from Lemma 7.2, we have

$$|\bar{C}_{\text{Gaussian}} - \bar{C}_{\text{Truncated}}| \leq 19|\mathcal{V}|. \tag{93}$$

Now pick i.i.d. normal $\mathcal{CN}(0,1)$ distribution for $\{x_i\}_{i\in\mathcal{V}}$. By applying Theorem 4.1 to the truncated deterministic relay network, we find

$$\begin{aligned} C_{\text{Truncated}} &\geq \min_{\Omega\in\Lambda_D} I\left(y_{\Omega^c}^{\text{truncated}}; x_\Omega \,|\, x_{\Omega^c}\right) \\ &\overset{(a)}{=} H\left(y_{\Omega^c}^{\text{truncated}} \,|\, x_{\Omega^c}\right) \end{aligned} \tag{94}$$

where (a) is because we have a deterministic network. By Lemma 6.6 and Lemma 7.2, we have

$$\begin{aligned} &\min_{\Omega\in\Lambda_D} I\left(y_{\Omega^c}^{\text{truncated}}; x_\Omega \,|\, x_{\Omega^c}\right) \\ &\quad\geq I\left(y_{\Omega^c}^{\text{Gaussian}}; x_\Omega \,|\, x_{\Omega^c}\right) - 19|\mathcal{V}| \\ &\quad\geq \bar{C}_{\text{Gaussian}} - 20|\mathcal{V}|. \end{aligned} \tag{95}$$

Then from (93) and (95) we have

$$\bar{C}_{\text{Gaussian}} - 20|\mathcal{V}| \leq C_{\text{Truncated}} \leq \bar{C}_{\text{Gaussian}} + 19|\mathcal{V}|. \tag{96}$$

Also from Theorem 4.5 we know that

$$\bar{C}_{\text{Gaussian}} - 15|\mathcal{V}| \leq C_{\text{Gaussian}} \leq \bar{C}_{\text{Gaussian}}. \tag{97}$$

Therefore

$$|C_{\text{Gaussian}} - C_{\text{Truncated}}| \leq 34|\mathcal{V}|. \tag{98}$$

■

VIII. Extensions

In this section, we extend our main result for Gaussian relay networks (Theorem 4.5) to the following scenarios:

1) Compound relay network.
2) Frequency selective relay network.
3) Half-duplex relay network.
4) Quasi-static fading relay network (underspread regime).
5) Low rate capacity approximation of Gaussian relay network.

A. Compound Relay Network

The relaying strategy we proposed for general Gaussian relay networks does not require any channel information at the relays; relays just quantize at noise level and forward through a random mapping. The approximation gap also does not depend on the channel gain values. As a result our main result for Gaussian relay networks (Theorem 4.5) can be extended to compound relay networks where we allow each channel gain $h_{i,j}$ to be from a set $\mathcal{H}_{i,j}$, and the particular chosen values are unknown to the source node S, the relays, and the destination node D. A communication rate R is achievable if there exists a scheme such that for any channel gain realizations, the source can communicate to the destination at rate R.

Theorem 8.1: The capacity C_{cn} of the compound Gaussian relay network satisfies

$$\bar{C}_{cn} - \kappa \leq C_{cn} \leq \bar{C}_{cn} \tag{99}$$

where $\bar{C}_{cn}$ is the cut-set upper bound on the compound capacity of $\mathcal{G}$, i.e.,

$$\bar{C}_{cn} = \max_{p(\{\mathbf{x}_i\}_{j\in\mathcal{V}})} \inf_{h\in\mathcal{H}} \min_{\Omega\in\Lambda_D} I(\mathbf{y}_{\Omega^c}; \mathbf{x}_\Omega | \mathbf{x}_{\Omega^c}) \tag{100}$$

and κ is a constant and is upper bounded by $13\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof outline: We sketch the proof for the case that nodes have single antenna; its extension to the multiple antenna scenario is straightforward. As we mentioned earlier, the relaying strategy that we used in Theorem 4.5 does not require any channel information. However, if all channel gains are known at the final destination, all rates within a constant gap to the cut-set upper bound are achievable. We first evaluate how much we lose if the final destination only knows a quantized version of the channel gains. In particular assume that each channel gain is bounded $|h_{ij}| \in [h_{\min}, h_{\max}]$, and final destination only knows the channel gain values quantized at level $\frac{1}{\sqrt{d_{\max}}}$, where $d_{\max}$ is the maximum degree of nodes in $\mathcal{G}$. Then since there is a transmit power constraint equal to one at each node, the effect of this channel uncertainty can be mimicked by adding a Gaussian noise of variance $d_{\max} \times (\frac{1}{\sqrt{d_{\max}}})^2 = 1$ at each relay node (i.e., doubling the noise variance at each node), which will result in a reduction of at most $|\mathcal{V}|$ bits from the cut-set upper bound. Therefore, with access to only quantized channel gains, we will lose at most $|\mathcal{V}|$ more bits, which means the gap between the achievable rate and the cut-set bound is at most $16|\mathcal{V}|$.

Furthermore, as shown in [23] there exists a universal decoder for this finite set of channel sets. Hence, we can use this decoder at the final destination and decode the message as if we knew the channel gains quantized at the noise level, for all rates up to

$$R < \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(y_{\Omega^c}; x_\Omega | x_{\Omega^c}) \tag{101}$$

where $\hat{\mathcal{H}}$ is representing the quantized state space. Now as we showed earlier, if we restrict the channels to be quantized at noise level the cut-set upper bound changes at most by $|\mathcal{V}|$, therefore

$$\bar{C}_{cn} - |\mathcal{V}| \leq \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(y_{\Omega^c}; x_\Omega \,|\, x_{\Omega^c}). \tag{102}$$

Therefore, from (101) and (102) all rates up to $\bar{C}_{cn} - 16|\mathcal{V}|$ are achievable and the proof can be completed.

Now by using the ideas in [24] and [25], we believe that an infinite state universal decoder can also be analysed to give "completely oblivious to channel" results. ■

B. Frequency Selective Gaussian Relay Network

In this section we generalize our main result to the case that the channels are frequency selective. Since one can present a frequency selective channel as a MIMO link, where each antenna

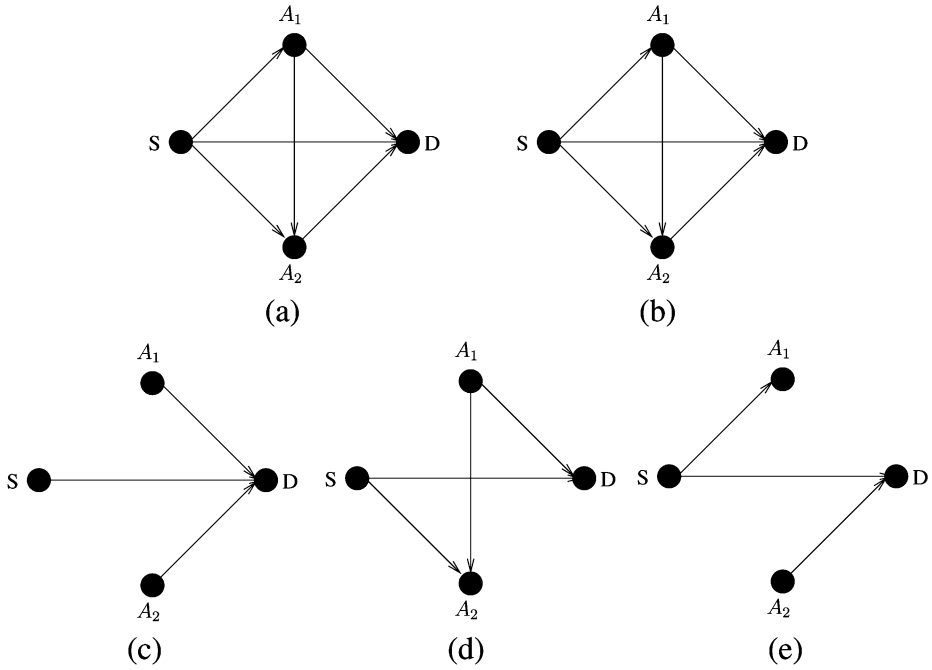

Fig. 19. Example of a relay network with two relays is shown in (*a*). All four modes of half-duplex operation of the relays are shown in (*b*)–(*e*). (a) A two-relay network, (b) Mode 1, (c) Mode 2, (d) Mode 3 (e) Mode 4.

is operating at a different frequency band,[5] this extension is just a straightforward corollary of the case that nodes have multiple antennas.

Theorem 8.2: The capacity C of the frequency selective Gaussian relay network with F different frequency bands satisfies

$$\bar{C} - \kappa \leq C \leq \bar{C} \tag{103}$$

where $\bar{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in (23), and κ is a constant and is upper bounded by $12F\sum_{i=1}^{|\mathcal{V}|} N_i + 3F\sum_{i=1}^{|\mathcal{V}|} M_i$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

C. Half Duplex Relay Network (Fixed Transmission Scheduling)

One of the practical constraints on wireless networks is that the transceivers cannot transmit and receive at the same time on the same frequency band, known as the half-duplex constraint. As a result of this constraint, the achievable rate of the network will in general be lower. The model that we use to study this problem is the same as [26]. In this model the network has finite modes of operation. Each mode of operation (or state of the network), denoted by $m \in \{1, 2, \ldots, M\}$, is defined as a valid partitioning of the nodes of the network into two sets of "sender" nodes and "receiver" nodes such that there is no active link that arrives at a sender node.[6] For each node i, the transmit and the receive signal at mode m are respectively shown by x_i^m and y_i^m. Also t_m defines the fraction of the time that network will operate in state m, as the network use goes to infinity. The cut-set upper bound on the capacity of the Gaussian relay network with half-duplex constraint, C_{hd}, is shown to be [26]

$$\begin{aligned} C_{hd} &\leq \bar{C}_{hd} \\ &= \max_{\substack{p(\{x_j^m\}_{j\in\mathcal{V}, m\in\{1,\ldots,M\}}) \\ t_m:0\leq t_m\leq 1, \sum_{m=1}^{M} t_m = 1}} \min_{\Omega\in\Lambda_D} \sum_{m=1}^{M} t_m \\ &\qquad I\left(y_{\Omega^c}^m; x_{\Omega}^m \,|\, x_{\Omega^c}^m\right). \end{aligned} \tag{104}$$

Theorem 8.3: The capacity C_{hd} of the Gaussian relay network with half-duplex constraint satisfies

$$\bar{C}_{hd} - \kappa \leq C_{hd} \leq \bar{C}_{hd} \tag{105}$$

where $\bar{C}_{hd}$ is the cut-set upper bound on the capacity as described in (104) and κ is a constant and is upper bounded by $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof: We prove the result for the case that nodes have single antenna; its extension to the multiple antenna scenario is straightforward. Since each relay can be either in a transmit or receive mode, we have a total of $M = 2^{|\mathcal{V}|-2}$ number of modes. An example of a network with two relay and all four modes of half-duplex operation of the relays are shown in Fig. 19.

Consider the t_i's that maximize $\bar{C}_{hd}$ in (104). Assume that they are rational numbers (otherwise look at the sequence of rational numbers approaching them) and set W to be the LCM (least common divisor) of the denominators. Now increase the bandwidth of system by W and allocate Wt_i of bandwidth to mode i, $i = 1, \ldots, M$. Each mode is running at a different frequency band. Therefore, as shown in Fig. 20, we can combine all these modes and create a frequency selective relay network.

[5]This can be implemented in particular by using OFDM and appropriate spectrum shaping or allocation.

[6]Active link is defined as a link which is departing from the set of sender nodes

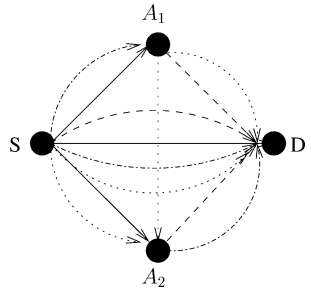

Fig. 20. Combination of all half-duplex modes of the network shown in Fig. 19. Each mode operates at a different frequency band.

Since the links are orthogonal to each other, the cut-set upper bound on the capacity of this frequency selective relay network (in bits/sec/Hz) is the same as (104). By theorem 8.2 we know that our quantize-map-and-forward scheme achieves, within a constant gap, κ, of $\bar{C}_{hd}$ for all channel gains. In this relaying scheme, at each block, each relay transmits a signal that is only a function of its received signal in the previous block and, hence, does not have memory over different blocks. We will translate this scheme to a scheme in the original network that modes are just at different times (not different frequency bands). The idea is that we can expand exactly communication block of the frequency selective network into W blocks of the original network and allocating Wt_i of these blocks to mode i. In the Wt_i blocks that are allocated to mode i, all relays do exactly what they do in frequency band i. This is described in Fig. 21 for the network of Fig. 20. This figure shows how one communication block of the frequency selective network (a) is expanded over W blocks of the the original half-duplex network (b). Since the transmitted signal at each frequency band is only a function of the data received in the previous block of the frequency selective network, the ordering of the modes inside the W blocks of the original network is not important at all. Therefore, with this strategy we can achieve within a constant gap, κ, of the cut-set bound of the half-duplex relay network and the proof is complete.

One of the differences between this strategy and our original strategy for full duplex networks is that now the relays might be required to have a much larger memory. In the full duplex scenario, in the layered case the relays had only memory over one block[7] (what they sent was only a function of the previous block). However, for the half-duplex scenario the relays are required to have a memory over W blocks and W can be arbitrarily large. ■

D. Quasi-Static Fading Relay Network (Underspread Regime)

In a wireless environment channel gains are not fixed and can change. In this section we consider a typical scenario in which although the channel gains change, they can be considered time invariant over a long time scale (for example during the transmission of a block). This happens when the coherence time of the channel (T_c) is much larger than the delay spread (T_d). Here the delay spread is the largest extent of the unequal path lengths, which is in some sense corresponding to intersymbol interference. Now, depending on how fast the channel gains are changing compared to the delay requirements, we have two different regimes: fast fading or slow fading scenarios. We consider each case separately.

1) Fast Fading: In the fast fading scenario the channel gains are changing much faster compared to the delay requirement of the application (i.e., coherence time of the channel, T_c, is much smaller than the delay requirements). Therefore, we can interleave data and encode it over different coherence time periods. In this scenario, ergodic capacity of the network is the relevant capacity measure to look at.

Theorem 8.4: The ergodic capacity C_{ergodic} of the quasi-static fast fading Gaussian relay network satisfies

$$\mathcal{E}_{h_{ij}}[\bar{C}(\{h_{ij}\})] - \kappa \le C_{\text{ergodic}} \le \mathcal{E}_{h_{ij}}[\bar{C}(\{h_{ij}\})] \tag{106}$$

where $\bar{C}$ is the cut-set upper bound on the capacity as described in (23) and the expectation is taken over the channel gain distribution. Also, the constant κ is upper bounded by $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof: We prove the result for the case that nodes have single antenna. Its extension to the multiple antenna scenario is straightforward. An upper bound is just the cut-set upper bound. For the achievability note that the relaying strategy we proposed for general relay networks does not depend on the channel realization, relays just quantize at noise level and forward through a random mapping. The approximation gap also does not depend on the channel parameters. As a result by coding data over L different channel realizations the following rate is achievable

$$\frac{1}{L}\sum_{l=1}^{L}(\bar{C}(\{h_{ij}\}^l) - \kappa). \tag{107}$$

Now as $L \to \infty$

$$\frac{1}{L}\sum_{l=1}^{L}\bar{C}(\{h_{ij}\}^l) \to \mathcal{E}_{h_{ij}}[\bar{C}] \tag{108}$$

and the theorem is proved. ■

2) Slow Fading: In a slow fading scenario the delay requirement does not allow us to interleave data and encode it over different coherence time periods. We assume that there is no channel gain information available at the source; therefore, there is no definite capacity and for a fixed target rate R we should look at the outage probability

$$\mathcal{P}_{\text{out}}(R) = \mathbb{P}\{C(\{h_{ij}\}) < R\} \tag{109}$$

where the probability is calculated over the distribution of the channel gains and the ϵ-outage capacity is defined as

$$C_\epsilon = \mathcal{P}_{\text{out}}^{-1}(\epsilon). \tag{110}$$

Here is our main result to approximate the outage probability.

Theorem 8.5: The outage probability $\mathcal{P}_{\text{out}}(R)$ of the quasi-static slow fading Gaussian relay network satisfies

$$\begin{aligned}\mathbb{P}\{\bar{C}(\{h_{ij}\}) < R\} &\le \mathcal{P}_{\text{out}}(R)\\ &\le \mathbb{P}\{\bar{C}(\{h_{ij}\}) < R+\kappa\}\end{aligned} \tag{111}$$

[7]This could be also done in the arbitrary networks but requires an alternative analysis. See footnote in Section V-B.

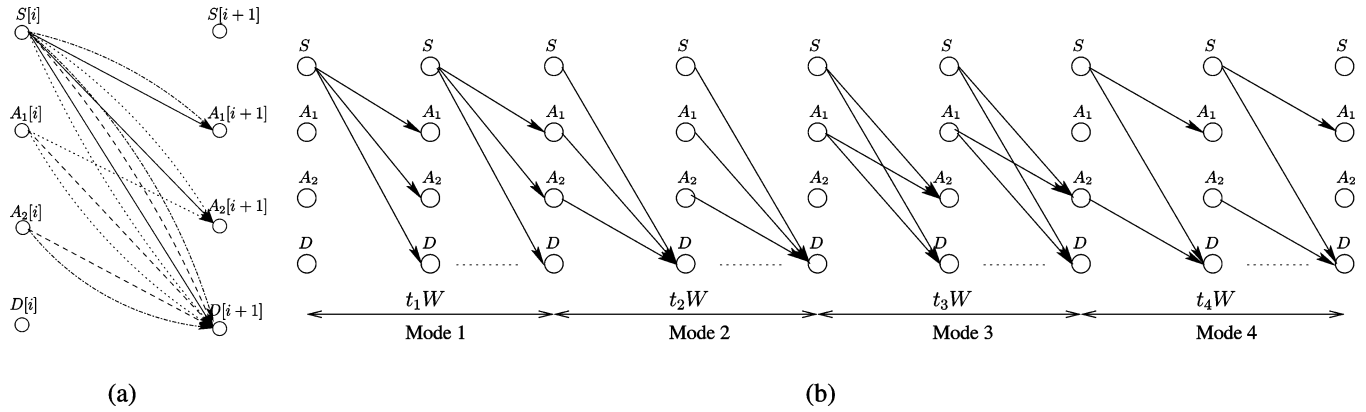

Fig. 21. One communication block of the frequency selective network (a), and its expansion over W blocks of the original half-duplex network (b).

where $\bar{C}$ is the cut-set upper bound on the capacity as described in (23) and the probability is calculated over the distribution of the channel gains. The constant κ is upper bounded by $12\sum_{i=1}^{|\mathcal{V}|} N_i + 3\sum_{i=1}^{|\mathcal{V}|} M_i$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

Proof: Lower bound is just based on the cut-set upper bound on the capacity. For the upper bound we use the compound network result. Therefore, based on Theorem 8.1, we know that as long as $\bar{C}(\{h_{ij}\}) - \kappa < R$ there will not be an outage. ∎

E. Low Rate Capacity Approximation of Gaussian Relay Network

In the low data rate regime, a constant-gap approximation of the capacity may not be useful any more. A more useful kind of approximation in this regime would be a universal multiplicative approximation, where the multiplicative factor does not depend on the channel gains in the network.

Theorem 8.6: The capacity C of the Gaussian relay network satisfies

$$\lambda \bar{C} \leq C \leq \bar{C} \tag{112}$$

where $\bar{C}$ is the cut-set upper bound on the capacity, as described in (23), and λ is a constant and is lower bounded by $\frac{1}{2d(d+1)}$, where d is the maximum degree of nodes in $\mathcal{G}$.

Proof: First we use a time-division scheme and make all links in the network orthogonal to each other. By Vizing's theorem (e.g., see [27] p.153) any simple undirected graph can be edge colored with at most $d+1$ colors, where d is the maximum degree of nodes in $\mathcal{G}$. Since our graph $\mathcal{G}$ is a directed graph we need at most $2(d+1)$ colors. Therefore, we can generate $2(d+1)$ time slots and assign the slots to directed graphs such that at any node all the links are orthogonal to each other. Therefore, each link is used a $\frac{1}{2(d+1)}$ fraction of the time. We further impose the constraint that each of these links uses a total $\frac{1}{2d(d+1)}$ of the time, but with a factor of d more power. By coding we can convert each links $h_{i,j}$ into a noise free link with capacity

$$c_{i,j} = \frac{1}{2d(d+1)} \log(1 + d|h_{i,j}|^2). \tag{113}$$

By Ford–Fulkerson theorem we know that the capacity of this network is

$$C_{\text{orthogonal}} = \min_{\Omega} \sum_{i,j:i\in\Omega, j\in\Omega^c} c_{i,j} \tag{114}$$

and this rate is achievable in the original Gaussian relay network. Now we will prove that

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)} \bar{C}. \tag{115}$$

To show this, assume that in the orthogonal network each node transmits the same signal on its outgoing links. Furthermore, each node j takes the summation of all incoming signals (normalized by $\frac{1}{\sqrt{d}}$) and denotes it as its received signal y_j, i.e.,

$$y_j[t] = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} (h_{ij}\sqrt{d}x_i[t] + z_{ij}[t]) \tag{116}$$

$$= \sum_{i=1}^{d} h_{ij}x_i[t] + \tilde{z}_j[t] \tag{117}$$

where

$$\tilde{z}_j[t] = \frac{\sum_{i=1}^{d} z_{ij}[t]}{\sqrt{d}} \sim \mathcal{CN}(0,1). \tag{118}$$

Therefore, we get a network which is statically similar to the original nonorthogonal network; however, each time-slot is only a $\frac{1}{d(d+1)}$ fraction of the time slots in the original network. Therefore, without this restriction the cut-set of the orthogonal network can only increase. Hence

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)} \bar{C}. \tag{119}$$

∎

IX. Conclusion

In this paper we presented a new approach to analyze the capacity of Gaussian relay networks. We start with deterministic models to build insights and use them as the foundation to analyze Gaussian models. The main results are a new scheme for

general Gaussian relay networks called quantize-map-and-forward and a proof that it can achieve to within a constant gap to the cutset bound. The gap does not depend on the SNR or the channel values of the network. No other scheme in the literature has this property.

One limitation of these results is that the gap grows with the number of nodes in the network. This is due to the noise accumulation property of the quantize-map-and-forward scheme. It is an interesting question whether there is another scheme that can circumvent this to achieve a universal constant gap to the cutset bound, independent of the number of nodes, or if this is an inherent feature of any scheme for arbitrary networks.[8] In this case a better upper bound than the cutset bound is needed.

Appendix A
Proof of Theorem 3.1

If $|h_{SR}| < |h_{SD}|$ then the relay is ignored and a communication rate equal to $R = \log(1+|h_{SD}|^2)$ is achievable. If $|h_{SR}| > |h_{SD}|$ the problem becomes more interesting. In this case by using the decode-forward scheme described in [4] we can achieve

$$R = \min(\log(1+|h_{SR}|^2), \log(1+|h_{SD}|^2+|h_{RD}|^2)).$$

Therefore, overall the following rate is always achievable

$$R_{\rm DF} = \max\{\log(1+|h_{SD}|^2), \min[\log(1+|h_{SR}|^2), \log(1+|h_{SD}|^2+|h_{RD}|^2)]\}.$$

Now we compare this achievable rate with the cut-set upper bound on the capacity of the Gaussian relay network

$$C \le \bar{C} = \max_{|\rho|\le 1} \min\{\log(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)), \log(1+|h_{SD}|^2+|h_{RD}|^2+2\rho|h_{SD}||h_{RD}|)\}.$$

Note that if $|h_{SR}| \ge |h_{SD}|$ then

$$R_{DF} = \min(\log(1+|h_{SR}|^2),\ \log(1+|h_{SD}|^2+|h_{RD}|^2))$$

and for all $|\rho| \le 1$ we have

$$\begin{aligned}&\log(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)) \\ &\quad \le \log(1+|h_{SR}|^2)+1 \\ &\log(1+|h_{SD}|^2+|h_{RD}|^2+2\rho|h_{SD}||h_{RD}|) \\ &\quad \le \log(1+|h_{SD}|^2+|h_{RD}|^2)+1\end{aligned}$$

[8]For a special class of parallel relay networks, such a universal constant independent of the number of parallel relays have been found recently in [30].

Hence

$$R_{\rm DF} \ge \bar{C}_{\rm relay} - 1.$$

Also, if $|h_{SD}| > |h_{SR}|$

$$R_{DF} = \log(1+|h_{SD}|^2)$$

and

$$\log(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)) \le \log(1+|h_{SD}|^2)+1$$

therefore again

$$R_{\rm DF} \ge \bar{C}_{\rm relay} - 1.$$

Appendix B
Proof of Theorem 3.2

The cut-set upper bound on the capacity of diamond network is shown in the equation at the bottom of the page. Without loss of generality assume $|h_{SA_1}| \ge |h_{SA_2}|$. Then we have the following cases:

1) $|h_{SA_1}| \le |h_{A_1D}|$: In this case

$$R_{PDF} \ge \log\left(1+|h_{SA_1}|^2\right) \ge \bar{C} - 1.$$

2) $|h_{SA_1}| > |h_{A_1D}|$:
Let $\alpha = \frac{|h_{A_1D}|^2}{|h_{SA_1}|^2}$, then

$$\begin{aligned}R_{PDF} &= \log\left(1+|h_{A_1D}|^2\right) \\ &\quad + \min\left\{\log\left(1+\frac{(1-\alpha)|h_{SA_2}|^2}{\alpha|h_{SA_2}|^2+1}\right), \log\left(1+\frac{|h_{A_2D}|^2}{1+|h_{A_1D}|^2}\right)\right\}\end{aligned}$$

or

$$R_{PDF} = \min\left\{\log\left(\frac{\left(1+|h_{SA_2}|^2\right)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right), \log(1+|h_{A_1D}|^2+|h_{A_2D}|^2)\right\}. \quad (120)$$

Now if

$$\log\left(\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{\alpha|h_{SA_2}|^2+1}\right) \ge \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right)$$

$$C_{\rm diamond} \le \bar{C} \le \min\left\{\log\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right), \log\left(1+(|h_{A_1D}|+|h_{A_2D}|)^2\right), \log\left(1+|h_{SA_1}|^2\right)+\log\left(1+|h_{A_2D}|^2\right) \log\left(1+|h_{SA_2}|^2\right)+\log\left(1+|h_{A_1D}|^2\right)\right\}.$$

we have

$$R_{PDF} = \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right) \geq \log\left(1+(|h_{A_1D}|+|h_{A_2D}|)^2\right)-1 \geq \bar{C}-1.$$

Therefore, the achievable rate of partial decode-forward scheme is within one bit of the cut-set bound. So we just need to look at the case that

$$R_{PDF} = \log\left(\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{\alpha|h_{SA_2}|^2+1}\right).$$

In this case, consider two possibilities:

- $\alpha|h_{SA_2}|^2 \leq 1$: Here, we have

$$\begin{aligned} R_{PDF} &= \log\left(\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{\alpha|h_{SA_2}|^2+1}\right) \\ &\geq \log\left(\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{2}\right) \\ &= \log\left(1+|h_{SA_2}|^2\right)+\log\left(1+|h_{A_1D}|^2\right)-1 \\ &\geq \bar{C}-1. \end{aligned}$$

- $\alpha|h_{SA_2}|^2 \geq 1$:
  In this case, we will show that

$$\begin{aligned} R_{PDF} &= \log\left(\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{\alpha|h_{SA_2}|^2+1}\right) \\ &\geq \log\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)-1 \\ &\geq \bar{C}-1. \end{aligned}$$

To show this, we just need to prove

$$\frac{\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right)}{\alpha|h_{SA_2}|^2+1} \geq \frac{1}{2}\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right).$$

By replacing $\alpha = \frac{|h_{A_1D}|^2}{|h_{SA_1}|^2}$, we get

$$\begin{aligned} &2|h_{SA_1}|^2\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right) \\ &\quad\geq \left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \\ &\qquad\times\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right). \end{aligned}$$

But note that

$$\begin{aligned} &2|h_{SA_1}|^2\left(1+|h_{SA_2}|^2\right)\left(1+|h_{A_1D}|^2\right) \\ &-\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \\ &\times\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right) \\ &\quad= |h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2 \\ &\qquad+\left(|h_{SA_1}|^2|h_{SA_2}|^2-|h_{SA_2}|^4|h_{A_1D}|^2\right) \\ &\qquad+\left(|h_{SA_1}|^2|h_{A_1D}|^2-|h_{SA_2}|^2|h_{A_1D}|^2\right) \\ &\qquad+\left(|h_{SA_1}|^2|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^4\right) \\ &\quad= |h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2+|h_{SA_2}|^2 \\ &\qquad\times\left(|h_{SA_1}|^2-|h_{SA_2}|^2|h_{A_1D}|^2\right) \\ &\qquad+|h_{A_1D}|^2\left(|h_{SA_1}|^2-|h_{SA_2}|^2\right) \\ &\qquad+|h_{SA_1}|^2\left(|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^2\right) \\ &\quad= |h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2+\left(|h_{SA_1}|^2-|h_{SA_2}|^2\right) \\ &\qquad\times\left(|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^2+|h_{A_1D}|^2\right)\geq 0 \end{aligned}$$

where the last step is true since

$$\begin{aligned} |h_{SA_1}|^2 &\geq |h_{SA_2}|^2 \\ |h_{SA_2}|^2|h_{A_1D}|^2 &\geq |h_{SA_1}|^2 \quad \left(\text{since } \alpha|h_{SA_2}|^2 \geq 1\right). \end{aligned}$$

APPENDIX C
PROOF OF THEOREMS 4.1 AND 4.2

In this Appendix we prove Theorems 4.1 and 4.2. We first generalize the encoding scheme to accommodate arbitrary deterministic functions of (28) in Section C.A. We then illustrate the ingredients of the proof using the same example as in Section V-A2. The complete proof of our result for layered networks is proved in Section C.C. The extension to the non-layered case is very similar to the proof for linear finite-field model discussed in Section V-B, hence, is omitted.

A. Encoding for Layered General Deterministic Relay Network

We have a single source S with a sequence of messages $w_k \in \{1,2,\ldots,2^{TR}\}$, $k=1,2,\ldots$. Each message is encoded by the source S into a signal over T transmission times (symbols), giving an overall transmission rate of R. We will use strong (robust) typicality as defined in [28]. The notion of joint typicality is naturally extended from Definition C.1.

Definition C.1: We define $\underline{x}$ as δ-typical with respect to distribution p, and denote it by $\underline{x} \in T_\delta$, if

$$|\nu_{\underline{x}}(x)-p(x)| \leq \delta p(x), \quad \forall x$$

where $\delta \in \mathbb{R}^+$ and $\nu_{\underline{x}}(x) = \frac{1}{T}|\{t : x_t = x\}|$, is the empirical frequency.

Each relay operates over blocks of time T symbols, and uses a mapping $f_j : \mathcal{Y}_j^T \to \mathcal{X}_j^T$ from its previous block of received T symbols to transmit signals in the next block. In particular, block k of T received symbols is denoted by $\mathbf{y}_j^{(k)} = \{y[(k-1)T+1],\ldots,y[kT]\}$ and the transmit symbols by $\mathbf{x}_j^{(k)}$. Choose some product distribution $\prod_{i\in\mathcal{V}} p(x_i)$. At the source S, map each of the indices in $w_k \in \{1,2,\ldots,2^{TR}\}$, choose $f_S(w_k)$ onto a sequence uniformly drawn from $T_\delta(x_S)$, which is the typical set of sequences in $\mathcal{X}_S^T$. At any relay node j choose f_j

to map each typical sequence in $T_\delta(y_j)$ onto the typical set of transmit sequences $T_\delta(x_j)$, as

$$\mathbf{x}_j^{(k)} = f_j\left(\mathbf{y}_j^{(k-1)}\right) \tag{121}$$

where f_j is chosen to map uniformly randomly each sequence in $T_\delta(y_j)$ onto $T_\delta(x_j)$. Each relay does the encoding prescribed by (121).

B. Proof Illustration

Now, we illustrate the ideas behind the proof of Theorem 4.1 for layered networks using the same example as in Section V-A2, which was done for the linear deterministic model. Since we are dealing with deterministic networks, the logic up to (42) in Section V-A2 remains the same. We will again illustrate the ideas using the cut $\Omega = \{S, A_1, B_1\}$. As in Section V-A2, we can write

$$\begin{aligned}\mathcal{P} &= \mathbb{P}\left\{\mathcal{A}_2, \mathcal{B}_2, \mathcal{D}, \mathcal{A}_1^c, \mathcal{B}_1^c\right\}\\ &= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\left\{\mathcal{B}_2, \mathcal{A}_1^c|\mathcal{A}_2\right\} \times \mathbb{P}\left\{\mathcal{D}, \mathcal{B}_1^c|\mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\right\}\\ &\le \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\left\{\mathcal{B}_2|\mathcal{A}_1^c, \mathcal{A}_2\right\} \times \mathbb{P}\left\{\mathcal{D}|\mathcal{B}_1^c, \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\right\}\\ &= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\left\{\mathcal{B}_2|\mathcal{A}_1^c, \mathcal{A}_2\right\} \times \mathbb{P}\left\{\mathcal{D}|\mathcal{B}_1^c, \mathcal{B}_2\right\}\end{aligned}$$

where the events $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2, \mathcal{D}\}$ are defined in (43), and the last step is true since there is an independent random mapping at each node and we have a Markovian layered structure in the network.

Note that since $\mathbf{y}_j \in T_\delta(y_j)$ with high probability, we can focus only on the typical received signals. Let us first examine the probability that $\mathbf{y}_{A_2}(w) = \mathbf{y}_{A_2}(w')$. Since S can distinguish between w, w', it maps these messages independently to two transmitted signals $\mathbf{x}_S(w), \mathbf{x}_S(w') \in T_\delta(x_S)$, hence, we can see that

$$\begin{aligned}\mathbb{P}\{\mathcal{A}_2\} &= \mathbb{P}\{(\mathbf{x}_S(w'), \mathbf{y}_{A_2}(w)) \in T_\delta(x_S, y_{A_2})\}\\ &\doteq 2^{-TI(x_S; y_{A_2})}\end{aligned} \tag{122}$$

where $\doteq$ indicates exponential equality (where we neglect subexponential constants).

Now, in order to analyze the second probability, as seen in the linear model analysis, $\mathcal{A}_2$ implies $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, i.e., the *same* signal is sent under both w, w'. Therefore, since $(\mathbf{x}_{A_2}(w), \mathbf{y}_{B_2}(w)) \in T_\delta(x_{A_2}, y_{B_2})$, obviously, $(\mathbf{x}_{A_2}(w'), \mathbf{y}_{B_2}(w)) \in T_\delta(x_{A_2}, y_{B_2})$ as well. Therefore, under w', we already have $\mathbf{x}_{A_2}(w')$ to be jointly typical with the signal that is received under w. However, since A_1 can distinguish between w, w', it will map the transmit sequence $\mathbf{x}_{A_1}(w')$ to a sequence which is independent of $\mathbf{x}_{A_1}(w)$ transmitted under w. Since an error occurs when $(\mathbf{x}_{A_1}(w'), \mathbf{x}_{A_2}(w'), \mathbf{y}_{B_2}(w)) \in T_\delta(x_{A_1}, x_{A_2}, y_{B_2})$, and since A_2 cannot distinguish between w, w', we also have $\mathbf{x}_{A_2}(w) = \mathbf{x}_{A_2}(w')$, we require that $(\mathbf{x}_{A_1}, \mathbf{x}_{A_2}, \mathbf{y}_{B_2})$ generated like $p(\mathbf{x}_{A_1})p(\mathbf{x}_{A_2}, \mathbf{y}_{B_2})$ behaves like a jointly typical sequence. Therefore, this probability is given by

$$\begin{aligned}&\mathbb{P}\{\mathcal{B}_2|\mathcal{A}_1^c, \mathcal{A}_2\}\\ &\quad= \mathbb{P}\{(\mathbf{x}_{A_1}(w'), \mathbf{x}_{A_2}(w), \mathbf{y}_{B_2}(w)) \in T_\delta(x_{A_1}, x_{A_2} y_{B_2})\}\\ &\quad\doteq 2^{-TI(x_{A_1}; y_{B_2}, x_{A_2})} \overset{(a)}{=} 2^{-TI(x_{A_1}; y_{B_2}|x_{A_2})}\end{aligned} \tag{123}$$

where (a) follows since we have generated the mappings f_j independently, it induces an independent distribution on x_{A_1}, x_{A_2}. Another way to see this is that the probability (123) is $\frac{|T_\delta(\mathbf{x}_{A_1}|\mathbf{x}_{A_2}, \mathbf{y}_{B_2})|}{|T_\delta(\mathbf{x}_{A_1})|}$, which by using properties of (robustly) typical sequences [28] yields the same expression as in (123). Note that the calculation in (123) is similar to one of the error event calculations in a multiple access channel.

Using a similar logic we can write

$$\begin{aligned}&\mathbb{P}\{\mathcal{D}|\mathcal{B}_1^c, \mathcal{B}_2\}\\ &\quad= \mathbb{P}\{(\mathbf{x}_{B_1}(w'), \mathbf{x}_{B_2}(w), \mathbf{y}_D(w)) \in T_\delta(x_{B_1}, x_{B_2} y_D)\}\\ &\quad\doteq 2^{-TI(x_{B_1}; y_D, x_{B_2})} \overset{(a)}{=} 2^{-TI(x_{B_1}; y_D|x_{B_2})}.\end{aligned} \tag{124}$$

Therefore, putting (122)–(124) together as done in (49) we get

$$\mathcal{P} \le 2^{-T\{I(x_S; y_{A_2}) + I(x_{A_1}; y_{B_2}|x_{A_2}) + I(x_{B_1}; y_D|x_{B_2})\}}.$$

Note that, for this example, due to the Markovian structure of the network we can see that[9] $I(y_{\Omega^c}; x_\Omega|x_{\Omega^c}) = I(x_S; y_{A_2}) + I(x_{A_1}; y_{B_2}|x_{A_2}) + I(x_{B_1}; y_D|x_{B_2})$, hence, as in (50), we get

$$P_e \le 2^{RT}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D} I(y_{\Omega^c}; x_\Omega|x_{\Omega^c})} \tag{125}$$

and hence, the error probability can be made as small as desired if $R < \min_{\Omega\in\Lambda_D} H(y_{\Omega^c}|x_{\Omega^c})$.

C. Proof of Theorems 4.1 and 4.2 for Layered Networks

As in the example illustrating the proof in Section C-B, the logic of the proof in the general deterministic functions follows that of the linear model quite closely.

For any such cut Ω, define the following sets:

- $L_l(\Omega)$: the nodes that are in Ω and are at layer l, (for example $S \in L_1(\Omega)$).
- $R_l(\Omega)$: the nodes that are in Ω^c and are at layer l, (for example $D \in R_{l_D}(\Omega)$).

As in Section V-A we can define the bi-partite network associated with a cut Ω. Instead of a transfer matrix $\mathbf{G}_{\Omega,\Omega^c}(\cdot)$ associated with the cut, we have a transfer function $\tilde{\mathbf{G}}_\Omega$. Since we are still dealing with a layered network, as in the linear model case, this transfer function breaks up into components corresponding to each of the l_D layers of the network. More precisely, we can create $d = l_D$ disjoint sub-networks of nodes corresponding to each layer of the network, with the set of nodes $L_{l-1}(\Omega)$, which

[9]Though, in the encoding scheme, there is a dependence between $x_{A_1}, x_{A_2}, x_{B_1}, x_{B_2}$ and x_S, in the single-letter form of the mutual information, under a product distribution, $x_{A_1}, x_{A_2}, x_{B_1}, x_{B_2}, x_S$ are independent of each other. Therefore, for example, y_{B_2} is independent of x_{B_2} leading to $H(y_{B_2}|x_{A_2}, x_{B_2}) = H(y_{B_2}|x_{A_2})$. Using this argument for the cut-set expression $I(y_{\Omega^c}; x_\Omega|x_{\Omega^c})$, we get the expansion.

are at distance $l-1$ from S and are in Ω, on one side and the set of nodes $R_l(\Omega)$, which are at distance l from S that are in Ω^c, on the other side, for $l=2,\ldots,l_D$. Each of these clusters have a transfer function $\mathbf{G}_l(\cdot), l=1,\ldots,l_D$ associated with them.

As in the linear model, each node i sees a signal related to $w=w_1$ in block $l_i=l-1$, and therefore, waits to receive this block and then does a mapping using the general encoding function given in (121) as

$$\mathbf{x}_j^{(k)}(w)=f_j^{(k)}\left(\mathbf{y}_j^{(k-1)}(w)\right). \tag{126}$$

The received signals in the nodes $j\in R_l(\Omega)$ are deterministic transformations of the transmitted signals from nodes $\mathcal{T}_l=\{u:(u,v)\in\mathcal{E},v\in R_l(\Omega)\}$. As in the linear model analysis of Section V-A, the dependence is on all the transmitting signals at distance $l-1$ from the source, not just the ones in $L_l(\Omega)$. Since all the receivers in $R_l(\Omega)$ are at distance l from S, they form the receivers of the layer l.

We now define the following events:

- $\mathcal{L}_l$: Event that the nodes in L_l can distinguish between w and w', i.e., $\mathbf{y}_{L_l}(w)\neq\mathbf{y}_{L_l}(w')$,
- $\mathcal{R}_l$: Event that the nodes in R_l can not distinguish between w and w', i.e., $\mathbf{y}_{R_l}(w)=\mathbf{y}_{R_l}(w')$.

Similar to Appendix C-B we can write

$$\begin{aligned}\mathcal{P}&=\mathbb{P}\{\mathcal{R}_l,\mathcal{L}_{l-1},l=2,\ldots,l_D\}\\&=\prod_{l=2}^{l_D}\mathbb{P}\{\mathcal{R}_l,\mathcal{L}_{l-1}\,|\,\mathcal{R}_j,\mathcal{L}_{j-1},j=2,\ldots,l-1\}\\&\leq\prod_{l=2}^{l_D}\mathbb{P}\{\mathcal{R}_l|\mathcal{R}_j,\mathcal{L}_j,j=2,\ldots,l-1\}\\&=\prod_{l=2}^{l_D}\mathbb{P}\{\mathcal{R}_l\,|\,\mathcal{R}_{l-1},\mathcal{L}_{l-1}\}.\end{aligned}$$

Note that for all the transmitting nodes in R_{l-1} which cannot distinguish between w,w' the transmitted signal would be the same under both w and w', i.e.,

$$\mathbf{x}_j(w)=\mathbf{x}_j(w'),\ \ j\in R_{l-1}.$$

Therefore, since $(\{\mathbf{x}_j(w)\}_{j\in R_{l-1}},\mathbf{y}_{R_l}(w))\in T_\delta$, we have that

$$(\{\mathbf{x}_j(w')\}_{j\in R_{l-1}},\mathbf{y}_{R_l}(w))\in T_\delta.$$

Therefore, just as in Appendix C-B, we see that

$$\begin{aligned}&\mathbb{P}\{\mathcal{R}_l|\mathcal{R}_{l-1},\mathcal{L}_{l-1}\}\\&\quad=\mathbb{P}\{(\mathbf{x}_{L_{l-1}}(w'),\mathbf{x}_{R_{l-1}}(w),\mathbf{y}_{R_l}(w))\\&\qquad\in T_\delta(x_{L_{l-1}},x_{R_l-1},y_{R_l})\}\doteq 2^{-TI(x_{L_{l-1}};y_{R_l}|x_{R_{l-1}})}.\end{aligned} \tag{127}$$

Therefore

$$\mathcal{P}\leq\prod_{l=2}^{d}2^{-TI(x_{L_{l-1}};y_{R_l}|x_{R_{l-1}})}=2^{-T\sum_{l=2}^{d}H(y_{R_l}|x_{R_{l-1}})}. \tag{128}$$

Due to the Markovian nature of the layered network, $\sum_{l=2}^{d}H(y_{R_l}|x_{R_{l-1}})=H(y_{\Omega^c}|x_{\Omega^c})$. From this point the proof closely follows the steps from (125) onwards. Similarly, in a multicast scenario we declare an error if *any* receiver $D\in\mathcal{D}$ makes an error. Since we have 2^{RT} messages, from the union bound we can drive the error probability to zero if we have

$$R<\max_{\prod_{i\in\mathcal{V}}p(x_i)}\min_{D\in\mathcal{D}}\min_{\Omega\in\Lambda_D}H(y_{\Omega^c}\,|\,x_{\Omega^c}). \tag{129}$$

We can use an argument similar to Section V-B in the linear deterministic case, to show that the layered proof for the general deterministic relay network can be extended to arbitrary (non-layered) deterministic networks. We also had an alternate proof for this conversion in [22], which used submodularity properties of entropy to show the same result.

Appendix D
Proof of Lemma 6.4

Consider the SVD decomposition of $\mathbf{H}$: $\mathbf{H}=\mathbf{U}\Sigma\mathbf{V}^\dagger$, with singular values $\sigma_1,\ldots,\sigma_{\min(m,n)}$. Let us define $K=\min\{m,n\}$ and $\tilde{\mathbf{x}}_j=[\tilde{x}_{1,j},\ldots,\tilde{x}_{m,j}]$, which is i.i.d. (over $1\leq j\leq T$) $\mathcal{CN}(0,\mathbf{I}_m)$.

Therefore, if $\|\mathbf{H}\tilde{\mathbf{x}}_j\|_\infty\leq\sqrt{2}$, then $\|\Sigma\mathbf{V}\tilde{\mathbf{x}}_j\|_2\leq\sqrt{2K}$, which means

$$\begin{aligned}\mathbb{P}\{\|\mathbf{H}\tilde{\mathbf{x}}_j\|_\infty\leq\sqrt{2}\}&\leq\mathbb{P}\{\|\Sigma\mathbf{V}\tilde{\mathbf{x}}_j\|_2\leq\sqrt{2K}\}\\&=\mathbb{P}\{\|\Sigma\tilde{\mathbf{x}}_j\|_2\leq\sqrt{2K}\}\end{aligned} \tag{130}$$

where the last step is true since the distribution of $\tilde{\mathbf{x}}$ and $\mathbf{V}\tilde{\mathbf{x}}$ are the same.

Now by using (130), we get

$$\begin{aligned}&\mathbb{P}\{\forall 1\leq j\leq T:\|\mathbf{H}[\tilde{x}_{1,j},\ldots,\tilde{x}_{m,j}]^t\|_\infty\leq\sqrt{2}\}\\&\quad\leq\mathbb{P}\{\forall 1\leq j\leq T:\|\Sigma[\tilde{x}_{1,j},\ldots,\tilde{x}_{m,j}]^t\|_2\leq\sqrt{2K}\}\\&\quad\leq\mathbb{P}\left\{\forall 1\leq j\leq T:\sum_{i=1}^{\min\{m,n\}}\sigma_i^2|\tilde{x}_{i,j}|^2\leq 2K\right\}\\&\quad=\prod_{j=1}^{T}\mathbb{P}\left\{\sum_{i=1}^{\min\{m,n\}}\sigma_i^2|\tilde{x}_{i,j}|^2\leq 2K\right\}\\&\quad\overset{(a)}{\leq}\prod_{j=1}^{T}e^{-\left(\sum_{i=1}^{\min\{m,n\}}\log(1+\frac{1}{2}2\sigma_i^2)-K\right)}\\&\quad=e^{-T\left(\sum_{i=1}^{\min\{m,n\}}\log(1+\sigma_i^2)-K\right)}\end{aligned}$$

where (a) follows from the Chernoff bound.[10]

Since, $\sum_{i=1}^{\min\{m,n\}}\log(1+\sigma_i^2)=\log\det(\mathbf{I}+\mathbf{H}\mathbf{H}^*)=I(\mathbf{x};\mathbf{H}\mathbf{x}+\mathbf{z})$, for $\mathbf{x}\sim\mathcal{CN}(0,\mathbf{I}_m),\mathbf{z}\sim\mathcal{CN}(0,\mathbf{I}_n)$, we get the desired result.

[10]We would like to acknowledge useful discussions with A. Ozgur on sharpening the proof of this result. It is also related to the proof technique in [20].

APPENDIX E
PROOF OF LEMMA 6.5

We first prove the following lemmas.

Lemma E.1: Consider integer-valued random variables x, r and s such that

$$\begin{aligned} &x \perp r \\ &s \in \{-L, \ldots, 0, \ldots, L\} \\ &\mathbb{P}\{|r| \geq k\} \leq e^{-f(k)}, \quad \text{for all } k \in \mathbb{Z}^+ \end{aligned}$$

for some integer L and a function $f(\cdot)$. Let

$$y = x + r + s.$$

Then

$$\begin{aligned} H(y|x) &\leq 2\log_2 e \left(\sum_{k=1}^{\infty} f(k) e^{-f(k)} \right) \\ &\quad + \frac{2L+1}{2} + N_f \\ H(x|y) &\leq \log(2L+1) + 2\log_2 e \left(\sum_{k=1}^{\infty} f(k) e^{-f(k)} \right) \\ &\quad + \frac{2L+1}{2} + N_f \end{aligned}$$

where

$$N_f = \left| \left\{ n \in \mathcal{Z}^+ \,|\, e^{-f(n)} > \frac{1}{2} \right\} \right|. \tag{131}$$

Proof: By definition we have

$$\begin{aligned} &H(y|x) = H(x + r + s|x) = H(r + s|x) \\ &\leq H(r+s) = -\sum_{k} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\}. \end{aligned}$$

Now since $-p \log p \leq \frac{1}{2}$ for $0 \leq p \leq 1$, we have

$$-\sum_{k=-L}^{L} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \leq \frac{2L+1}{2}. \tag{132}$$

For $|k| > L$ we have

$$\mathbb{P}\{r+s=k\} \leq \mathbb{P}\{|r| \geq |k| - L\} \leq e^{-f(|k|-L)}. \tag{133}$$

Since $p \log p$ is decreasing in p for $p < \frac{1}{2}$ we have

$$\begin{aligned} &-\sum_{k=L+1}^{\infty} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &= -\sum_{\substack{k>L \\ k-L \in N_f}} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &\quad - \sum_{\substack{k>L \\ k-L \notin N_f}} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &\leq \frac{N_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L) \log e \end{aligned} \tag{134}$$

and similarly

$$\begin{aligned} &-\sum_{k=-\infty}^{-L} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &= -\sum_{\substack{k<-L \\ |k|-L \in N_f}} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &\quad - \sum_{\substack{k<-L \\ |k|-L \notin N_f}} \mathbb{P}\{r+s=k\} \log \mathbb{P}\{r+s=k\} \\ &\leq \frac{N_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L) \log e. \end{aligned} \tag{135}$$

By combining (132), (134) and (135) we get

$$H(y|x) \leq 2\log_2 e \left(\sum_{k=1}^{\infty} f(k) e^{-f(k)} \right) + \frac{2L+1}{2} + N_f. \tag{136}$$

Now we prove the second inequality

$$\begin{aligned} H(x|y) &= H(x|x+r+s) = H(x) - I(x; x+r+s) \\ &= H(x) - H(x+r+s) + H(x+r+s|x) \\ &\leq H(x) - H(x+r+s|s) + H(y|x) \\ &= H(x) - H(x+r|s) + H(y|x) \\ &= H(x) - H(x+r) + I(x+r; s) + H(y|x) \\ &\leq H(x) - H(x+r) + H(s) + H(y|x) \\ &\leq H(x) - H(x+r|r) + \log(2L+1) + H(y|x) \\ &= H(x) - H(x) + \log(2L+1) + H(y|x) \\ &= \log(2L+1) + H(y|x). \end{aligned}$$

Therefore

$$\begin{aligned} H(x|y) &\leq \log(2L+1) \\ &\quad + 2\log_2 e \left(\sum_{k=1}^{\infty} f(k) e^{-f(k)} \right) + \frac{2L+1}{2} + N_f. \end{aligned} \tag{137}$$

■

Corollary E.2: Assume v is a continuous complex random variable, then

$$\begin{aligned} H([v+z]|[v]) &\leq 12 \\ H([v]|[v+z]) &\leq 12 \end{aligned}$$

where z is a $\mathcal{CN}(0,1)$ random variable independent of v and $[\cdot]$ is defined in Definition 6.1.

Proof: We use lemma E.1 with variables

$$\begin{aligned} x &= [\mathrm{Re}(v)] \\ r &= [\mathrm{Re}(z)] \\ s &= [\{\mathrm{Re}(v)\} + \{\mathrm{Re}(z)\}]. \end{aligned}$$

Then $L = 1$ and since

$$\begin{aligned}&\mathbb{P}\{|[\mathrm{Re}(z)]| \geq k\}\\ &\quad\leq \mathbb{P}\left\{|[\mathrm{Re}(z)]| - \frac{1}{2} \geq k\right\} = 2Q\left(k - \frac{1}{2}\right)\\ &\quad\leq e^{-\frac{(k-\frac{1}{2})^2}{2}}.\end{aligned}$$

We can use

$$f(k) = \frac{\left(k - \frac{1}{2}\right)^2}{2}.$$

Also since

$$e^{-\frac{(k-\frac{1}{2})^2}{2}} < \frac{1}{2}, \quad \text{for } k \geq 2$$

we have $N_f = 1$. Hence

$$\begin{aligned}&\log(2L+1) + 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + N_f\\ &\quad= 2\log_2 e\left(\sum_{k=1}^{\infty} \frac{(k-\frac{1}{2})^2}{2} e^{-\frac{(k-\frac{1}{2})^2}{2}}\right) + 2.5 + \log_2 3\\ &\quad\approx 5.89 < 6.\end{aligned}$$

As a result

$$\begin{aligned}H([\mathrm{Re}(v+z)]|[\mathrm{Re}(v)]) &\leq 6\\ H([\mathrm{Re}(v)]|[\mathrm{Re}(v+z)]) &\leq 6.\end{aligned}$$

Similarly

$$\begin{aligned}H([\mathrm{Im}(v+z)]|[\mathrm{Im}(v)]) &\leq 6\\ H([\mathrm{Im}(v)]|[\mathrm{Im}(v+z)]) &\leq 6.\end{aligned}$$

Therefore

$$\begin{aligned}H([v+z]|[v]) &\leq H([\mathrm{Re}(v+z)]|[\mathrm{Re}(v)])\\ &\quad + H([\mathrm{Im}(v+z)]|[\mathrm{Im}(v)]) \leq 12\\ H([v]|[v+z]) &\leq H([\mathrm{Re}(v)]|[\mathrm{Re}(v+z)])\\ &\quad + H([\mathrm{Im}(v)]|[\mathrm{Im}(v+z)]) \leq 12.\end{aligned}$$

■

$$\begin{aligned}H([\mathbf{y}_D]|u, F_{\mathcal{V}})] &\leq H([\mathbf{y}_{\mathcal{V}}]|w', F_{\mathcal{V}})\\ &= \sum_{l=2}^{l_D} H([\mathbf{y}_{\mathcal{V}_l}]|[\mathbf{y}_{\mathcal{V}_{l-1}}, F_{\mathcal{V}}])\\ &= \sum_{l=2}^{l_D} H([\mathbf{y}_{\mathcal{V}_l}]|\mathbf{x}_{\mathcal{V}_{l-1}}, F_{\mathcal{V}})\\ &= \sum_{l=2}^{l_D} H([\mathrm{Re}(\mathbf{y}_{\mathcal{V}_l})]|\mathbf{x}_{\mathcal{V}_{l-1}}, F_{\mathcal{V}})\\ &\quad + H([\mathrm{Im}(\mathbf{y}_{\mathcal{V}_l})]|\mathbf{x}_{\mathcal{V}_{l-1}}, F_{\mathcal{V}})\\ &\overset{\text{Corollary E.2}}{\leq} \sum_{l=2}^{l_D} 12T|\mathcal{V}_l|\\ &= 12T|\mathcal{V}|.\end{aligned}$$

Appendix F
Proof of Lemma 6.6

First note that $\bar{C}_\Omega$ is the capacity of the channel that the cut Ω creates. Therefore, intuitively we want to prove that the gap between the capacity of a MIMO channel and its capacity when it is restricted to have equal power allocation at the transmitting antennas is upper bounded by a constant. Therefore, without loss of generality we just focus an $n \times m$ MIMO channel, with $K = \min\{m, n\}$,

$$\mathbf{y}^n = \mathbf{G}\mathbf{x}^m + \mathbf{z}^n \tag{138}$$

with average transmit power per antenna equal to P and i.i.d complex normal noise. We know that the capacity of this MIMO channel is achieved with water filling, and

$$C = C_{wf} = \sum_{i=1}^{K} \log(1 + \tilde{Q}_{ii}\lambda_i) \tag{139}$$

where λ_i's are the singular values of $\mathbf{G}$ and $\tilde{Q}_{ii}$ is given by water filling solution satisfying

$$\sum_{i=1}^{K} \tilde{Q}_{ii} = mP. \tag{140}$$

Now with equal power allocation we have

$$C_{ep} = \sum_{i=1}^{K} \log(1 + P\lambda_i). \tag{141}$$

Now note that

$$\begin{aligned}&C_{wf} - C_{ep}\\ &\quad= \log\left(\frac{\prod_{i=1}^{K}(1 + \tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{K}(1 + P\lambda_i)}\right)\\ &\quad\leq \log\left(\frac{\prod_{i=1}^{K}(1 + \tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{K}\max(1, P\lambda_i)}\right)\\ &\quad= \log\left(\prod_{i=1}^{K}\frac{1 + \tilde{Q}_{ii}\lambda_i}{\max(1, P\lambda_i)}\right)\\ &\quad= \log\left(\prod_{i=1}^{K}\left(\frac{1}{\max(1, P\lambda_i)} + \frac{\tilde{Q}_{ii}\lambda_i}{\max(1, P\lambda_i)}\right)\right)\\ &\quad\leq \log\left(\prod_{i=1}^{K}\left(1 + \frac{\tilde{Q}_{ii}\lambda_i}{P\lambda_i}\right)\right)\\ &\quad= \log\left(\prod_{i=1}^{K}\left(1 + \frac{\tilde{Q}_{ii}}{P}\right)\right).\end{aligned}$$

Now note that

$$\sum_{i=1}^{K}\left(1 + \frac{\tilde{Q}_{ii}}{P}\right) = K + m \tag{142}$$

and therefore, by arithmetic mean-geometric mean inequality we have

$$\prod_{i=1}^{K}\left(1+\frac{\tilde{Q}_{ii}}{P}\right)\leq\left(\frac{\sum_{i=1}^{K}\left(1+\frac{\tilde{Q}_{ii}}{P}\right)}{K}\right)^{K}=\left(1+\frac{m}{K}\right)^{K} \tag{143}$$

and hence

$$\begin{aligned} C_{wf}-C_{ep} &\leq K\log\left(1+\frac{m}{K}\right) && (144)\\ &= K\log\left(\frac{m}{K}\right)+K\log\left(1+\frac{K}{m}\right) && (145)\\ &\leq \underbrace{K\log\left(\frac{m}{K}\right)}_{\log\left(\frac{m}{K}\right)^{K}}+K \overset{(a)}{\leq} \frac{m}{e}+K && (146)\end{aligned}$$

where $K=\min\{m,n\}$, and (a) follows because $\max_K(\frac{m}{K})^K \leq e^{m/e}$ and we also take natural logarithms. Therefore, the loss from restricting ourselves to use equal transmit powers at each antenna of an $m\times n$ MIMO channel is at most $\frac{m}{e}+\min\{m,n\}$ bits.

Now, let us apply (144) to prove Lemma 6.6. Note that the cut-set upper bound of (23) when applied to the Gaussian network yields

$$\begin{aligned}\bar{C} &= \max_{\substack{p(\{\mathbf{X}_i\})\\ \mathbf{Q}:\mathbf{Q}_{ii}\leq P,\forall i}}\ \min_{\Omega\in\Lambda_D}\{h(\mathbf{Y}_{\Omega^c}\,|\,\mathbf{X}_{\Omega^c})\\ &\quad -h(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c},\mathbf{X}_{\Omega})\} && (147)\\ &= \max_{\substack{p(\{\mathbf{X}_i\}),\\ \mathbf{Q}:\mathbf{Q}_{ii}\leq P,\forall i}}\ \min_{\Omega\in\Lambda_D}\{h(\mathbf{Y}_{\Omega^c}-\breve{\mathbf{G}}_{\Omega^c,\Omega^c}\mathbf{X}_{\Omega^c}\,|\,\mathbf{X}_{\Omega^c})\\ &\quad -h(\mathbf{Z}_{\Omega^c})\}\\ &\leq \max_{\mathbf{Q}:\mathbf{Q}_{ii}\leq P,\forall i}\ \min_{\Omega\in\Lambda_D}\log|\mathbf{I}+\mathbf{G}_{\Omega}\mathbf{Q}\mathbf{G}_{\Omega}^{*}| && (148)\end{aligned}$$

where $\mathbf{G}_\Omega$ represents the network transfer matrix from transmitting set Ω to receiving set Ω^c and $\breve{\mathbf{G}}_{\Omega^c,\Omega^c}$ represents the transfer matrix from set Ω^c to Ω^c. The maximization in (23) can be restricted to jointly Gaussian inputs represented by covariance matrix $\mathbf{Q}$ with individual power constraints. Now, clearly these constraints can be relaxed to the sum-power constraints yielding

$$\begin{aligned}\bar{C} &\leq \max_{\mathbf{Q}:\mathbf{Q}_{ii}\leq P,\forall i}\ \min_{\Omega\in\Lambda_D}\log|\mathbf{I}+\mathbf{G}_{\Omega}\mathbf{Q}\mathbf{G}_{\Omega}^{*}|\\ &\leq \min_{\Omega\in\Lambda_D}\ \max_{\mathbf{Q}:\mathrm{tr}(\mathbf{Q})\leq|\Omega|P}\log|\mathbf{I}+\mathbf{G}_{\Omega}\mathbf{Q}\mathbf{G}_{\Omega}^{*}|\\ &= \min_{\Omega\in\Lambda_D}\bar{C}_{\Omega}. && (149)\end{aligned}$$

Now, let us define $\bar{C}_{\Omega}^{iid}$, to be the cut value for *i.i.d.* Gaussian inputs, i.e., $\mathbf{Q}=\mathbf{I}$. More precisely, from Definition 6.2 we have for $p(\{\mathbf{X}_i\})=\prod_i p(\mathbf{X}_i)$, and $\mathbf{X}_i\sim\mathcal{CN}(0,1)$, i.e., i.i.d., unit variance Gaussian variables, the cut value evaluated as

$$\begin{aligned}\bar{C}_{i.i.d.} &= \min_{\Omega\in\Lambda_D}\{h(\mathbf{Y}_{\Omega^c}\,|\,\mathbf{X}_{\Omega^c})-h(\mathbf{Y}_{\Omega^c}\,|\,\mathbf{X}_{\Omega^c},\mathbf{X}_{\Omega})\}\\ &= \min_{\Omega\in\Lambda_D}\{h(\mathbf{Y}_{\Omega^c}-\breve{\mathbf{G}}_{\Omega^c,\Omega^c}\mathbf{X}_{\Omega^c}\,|\,\mathbf{X}_{\Omega^c})-h(\mathbf{Z}_{\Omega^c})\}\\ &\overset{(a)}{=} \min_{\Omega}\underbrace{\log|\mathbf{I}+P\mathbf{G}_{\Omega}\mathbf{G}_{\Omega}^{*}|}_{\bar{C}_{\Omega}^{iid}}=\min_{\Omega}\bar{C}_{\Omega}^{iid} && (150)\end{aligned}$$

where (a) follows because $\mathbf{Y}_{\Omega^c}-\breve{\mathbf{G}}_{\Omega^c,\Omega^c}\mathbf{X}_{\Omega^c}=\mathbf{G}_{\Omega}\mathbf{X}_{\Omega}+\mathbf{Z}_{\Omega^c}$ is independent of $\mathbf{X}_{\Omega^c}$ due to i.i.d. choice of input distrbutions.

By using (144), we get

$$\begin{aligned}&\bar{C}_{\Omega}-\bar{C}_{\Omega}^{iid}\leq\frac{|\Omega|}{e}+\min\{|\Omega|,|\Omega^c|\}\leq 2|\mathcal{V}|,\quad\forall\Omega\\ &\text{or } \bar{C}_{\Omega}\leq\bar{C}_{\Omega}^{iid}+2|\mathcal{V}|,\quad\forall\Omega. && (151)\end{aligned}$$

Since $\min_{\Omega}\bar{C}_{\Omega}\leq\min_{\Omega}\bar{C}_{\Omega}^{iid}+2|\mathcal{V}|$, we get the claimed result in Lemma 6.6, for the scalar case.

For the case with multiple antennas, we see that for any cut Ω, the number of degrees of freedom is $\min\{\sum_{i\in\Omega}M_i,\sum_{i\in\Omega^c}N_i\}$. Note that, $\max_\Omega\min\{\sum_{i\in\Omega}M_i,\sum_{i\in\Omega^c}N_i\}\leq\sum_{i=1}^{|\mathcal{V}|}M_i$ and $\max_\Omega\min\{\sum_{i\in\Omega}M_i,\sum_{i\in\Omega^c}N_i\}\leq\sum_{i=1}^{|\mathcal{V}|}N_i$ and hence $\max_\Omega\min\{\sum_{i\in\Omega}M_i,\sum_{i\in\Omega^c}N_i\}\leq\min\{\sum_{i=1}^{|\mathcal{V}|}M_i,\sum_{i=1}^{|\mathcal{V}|}N_i\}$ yielding

$$\min\left\{\sum_{i\in\Omega}M_i,\sum_{i\in\Omega^c}N_i\right\}\leq\min\left\{\sum_{i=1}^{|\mathcal{V}|}M_i,\sum_{i=1}^{|\mathcal{V}|}N_i\right\},\quad\forall\Omega. \tag{152}$$

For a trivial upper bound to use in an argument analogous to (150), we can use (151) to see that

$$\bar{C}_{\Omega}\leq\bar{C}_{\Omega}^{iid}+2\sum_{i=1}^{|\mathcal{V}|}M_i. \tag{153}$$

Appendix G
Proof of Lemma 7.2

We first prove the following two lemmas:

Lemma G.1: Let G be the channel gains matrix of a $m\times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{x}}$,

$$|I(\mathbf{x};[G\mathbf{x}+\mathbf{z}])-I(\mathbf{x};[G\mathbf{x}])|\leq 12n \tag{154}$$

where $\mathbf{z}=[z_1,\ldots,z_n]$ is a vector of n i.i.d. $\mathcal{CN}(0,1)$ random variables.

Lemma G.2: Let G be the channel gains matrix of a $m\times n$ MIMO system. Assume that there is an average power constraint equal to one at each node. Then for any input distribution $P_{\mathbf{x}}$,

$$|I(\mathbf{x};G\mathbf{x}+\mathbf{z})-I(\mathbf{x};[G\mathbf{x}+\mathbf{z}])|\leq 7n \tag{155}$$

where $\mathbf{z} = [z_1, \ldots, z_n]$ is a vector of n i.i.d. $\mathcal{CN}(0,1)$ random variables.

Note that Lemma 7.2 is just a corollary of these two lemmas, which are proved next.

Proof: (Proof of Lemma G.1): First note that

$$\begin{aligned} &I(\mathbf{x};[G\mathbf{x}]) \\ &\quad \le I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]) + I(\mathbf{x};[G\mathbf{x}]|[G\mathbf{x}+\mathbf{z}]) \\ &\quad = I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]) + H([G\mathbf{x}]|[G\mathbf{x}+\mathbf{z}]) \\ &\quad \overset{\text{(Corollary E.2)}}{\le} I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]) + 12n. \end{aligned} \tag{156}$$

$$\begin{aligned} &I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]) \\ &\quad \le I(\mathbf{x};[G\mathbf{x}]) + I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]|[G\mathbf{x}]) \\ &\quad \le I(\mathbf{x};[G\mathbf{x}]) + H([G\mathbf{x}+\mathbf{z}]|[G\mathbf{x}]) \\ &\quad \overset{\text{(Corollary E.2)}}{\le} I(\mathbf{x};[G\mathbf{x}]) + 12n. \end{aligned} \tag{157}$$

Now from (154) and (155), we have

$$|I(\mathbf{x};[G\mathbf{x}+\mathbf{z}]) - I(\mathbf{x};[G\mathbf{x}])| \le 12n. \tag{158}$$

■

Proof: (Proof of Lemma G.2): Define the following random variables:

$$\begin{aligned} \mathbf{y} &= G\mathbf{x}+\mathbf{z} \\ \hat{\mathbf{y}} &= [G\mathbf{x}+\mathbf{z}] \\ \tilde{\mathbf{y}} &= \hat{\mathbf{y}} + \mathbf{u} \end{aligned}$$

where $\mathbf{u} = [u_1, \ldots, u_n]$ is a vector of n i.i.d. complex variables with distribution uniform [0,1] on both real and complex components, independent of $\mathbf{x}$ and $\mathbf{z}$.

By the data processing inequality we have $I(\mathbf{x};\mathbf{y}) \ge I(\mathbf{x};\hat{\mathbf{y}}) \ge I(\mathbf{x};\tilde{\mathbf{y}})$. Now, note that

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) - I(\mathbf{x};\tilde{\mathbf{y}}) &= h(\mathbf{y}) - h(\tilde{\mathbf{y}}) + h(\tilde{\mathbf{y}}\,|\,\mathbf{x}) - h(\mathbf{y}\,|\,\mathbf{x}) \\ &= h(\mathbf{y}) - h(\tilde{\mathbf{y}}) + h(\tilde{\mathbf{y}}\,|\,\mathbf{x}) - n\log(\pi e) \\ &= h(\mathbf{y}\,|\,\tilde{\mathbf{y}}) - h(\tilde{\mathbf{y}}\,|\,\mathbf{y}) + h(\tilde{\mathbf{y}}\,|\,\mathbf{x}) - n\log(\pi e) \\ &= h(\mathbf{y}\,|\,\tilde{\mathbf{y}}) - h(\mathbf{u}) + h(\tilde{\mathbf{y}}\,|\,\mathbf{x}) - n\log(\pi e) \\ &= h(\mathbf{y}\,|\,\tilde{\mathbf{y}}) + h(\tilde{\mathbf{y}}\,|\,\mathbf{x}) - n\log(\pi e) \end{aligned} \tag{159}$$

where the last step is true since $h(\mathbf{u}) = nh(u_1) = 2n\log 1 = 0$. Now

$$\begin{aligned} &|\mathrm{Re}(y) - \mathrm{Re}(\tilde{y})| \\ &\quad \le \max_{x\in\mathbb{C}}(|[\mathrm{Re}(x)] - \mathrm{Re}(x)|) + \max|\mathrm{Re}(u)| = \frac{3}{2} \end{aligned} \tag{160}$$

and similarly

$$\begin{aligned} &|\mathrm{Im}(y) - \mathrm{Im}(\tilde{y})| \\ &\quad \le \max_{x\in\mathbb{C}}(|[\mathrm{Im}(x)] - \mathrm{Im}(x)|) + \max|\mathrm{Im}(u)| = \frac{3}{2} \end{aligned} \tag{161}$$

Therefore

$$\begin{aligned} &h(\mathbf{y}|\tilde{\mathbf{y}}) \\ &\quad = h(\mathbf{y}-\tilde{\mathbf{y}}|\tilde{\mathbf{y}}) \\ &\quad \le n\log(2\pi e \\ &\qquad \times \sqrt{\max(|\mathrm{Re}(y) - \mathrm{Re}(\tilde{y})|)\max(|\mathrm{Im}(y) - \mathrm{Im}(\tilde{y})|)}) \\ &\quad = n\log 3\pi e. \end{aligned} \tag{162}$$

For the second term, lets look at the ith element of $\tilde{y}$

$$\tilde{y}_i = [\mathbf{g}_i\mathbf{x} + z_i] + u_i = \mathbf{g}_i\mathbf{x} + z_i + \delta(\mathbf{g}_i\mathbf{x} + z_i) + u_i \tag{163}$$

where $\tilde{y}_i$ is the ith component of $\tilde{y}$, $\mathbf{g}_i$ is the ith row of G, and $\delta(x) = x - [x]$. Clearly $|\mathrm{Re}(\delta(x))|, |\mathrm{Im}(\delta(x))| \le \frac{1}{2}$ for all $x \in \mathbb{C}$. Therefore, given x the variance of $\tilde{y}_i$ is bounded by

$$\begin{aligned} &\mathrm{Var}[\mathrm{Re}(\tilde{y}_i)|\mathbf{x}] \\ &\quad = \mathrm{Var}[\mathrm{Re}(z_i) + \mathrm{Re}(\delta(\mathbf{g}_i\mathbf{x}+z_i)) + \mathrm{Re}(u_i)] \\ &\quad \le \mathrm{Var}[\mathrm{Re}(z_i)] + \mathrm{Var}[\mathrm{Re}(\delta(\mathbf{g}_I\mathbf{x}+z_i))|\mathbf{x}] \\ &\qquad + 2\mathrm{Cov}[\mathrm{Re}(z_i), \mathrm{Re}(\delta(\mathbf{g}_i\mathbf{x}+z_i))|\mathbf{x}] + \mathrm{Var}[\mathrm{Re}(u)] \\ &\quad \le \mathrm{Var}[\mathrm{Re}(z_i)] + |\max \mathrm{Re}(\delta(\cdot))|^2 \\ &\qquad + 2\sqrt{\mathrm{Var}[\mathrm{Re}(z_i)] \times |\max \mathrm{Re}(\delta(\cdot))|} \\ &\qquad + \mathrm{Var}[\mathrm{Re}(u_i)] \\ &\quad = \frac{1}{2} + \frac{1}{4} + 1 + \frac{1}{12} = \frac{11}{6}. \end{aligned} \tag{164}$$

Similarly

$$\mathrm{Var}[\mathrm{Im}(\tilde{y}_i)|\mathbf{x}] \le \frac{11}{6}. \tag{165}$$

Therefore

$$\begin{aligned} h(\tilde{\mathbf{y}}|\mathbf{x}) &\le \sum_{i=1}^{n} h(\tilde{y}_i|\mathbf{x}) \le \sum_{i=1}^{n} \log 2\pi e\sqrt{|K_{\tilde{y}_i|X}|} \\ &\overset{(163)}{\le} n\log\frac{11}{3}\pi e. \end{aligned} \tag{166}$$

Now from (158), (161), and (165), we have

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) - I(\mathbf{x};\tilde{\mathbf{y}}) &\le h(\mathbf{y}|\tilde{\mathbf{y}}) + h(\tilde{\mathbf{y}}|\mathbf{x}) - \frac{n}{2}\log(2\pi e) \\ &\le n\log 11\pi e \approx 6.55n < 7n. \end{aligned}$$

■

ACKNOWLEDGMENT

The authors would like to thank A. Sahai for his insightful comments on an earlier draft of this work. In particular, they motivated the simpler proof of the approximation theorem presented in this manuscript for Gaussian relay networks. We would also like to thank several others for stimulating discussions on the topic of this paper including C. Fragouli, S. Mohajer, A. Ozgur, and R. Yeung.

REFERENCES

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

[2] M. Effros, M. Medard, T. Ho, S. Ray, D. Karger, and R. Koetter, "Linear network codes: A unified framework for source channel, and network coding," presented at the DIMACS Workshop on Network Information Theory, 2003.

[3] E. C. van der Meulen, "Three-terminal communication channels," *Ad. Appl. Pmb.*, vol. 3, pp. 120–154, Sep. 1971.

[4] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[5] M. Aref, "Information Flow in Relay Networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Oct. 1980.

[6] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. IT-46, no. 4, pp. 1204–1216, Jul. 2000.

[7] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 371–381, Feb. 2003.

[8] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, 2003.

[9] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2425–2432, Jun. 2006.

[10] P. Gupta, S. Bhadra, and S. Shakkottai, "On network coding for interference networks," presented at the IEEE Int. Symp. Information Theory (ISIT), Seattle, WA, Jul. 2006.

[11] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, Mar. 2006.

[12] B. Schein, *Distributed Coordination in Network Information Theory*. Cambridge, MA: Mass. Inst. Technol., 2001.

[13] L. L. Xie and P. R. Kumar, "A network information theory for wireless communication: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. IT-50, no. 5, pp. 748–767, May 2004.

[14] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[15] R. W. Yeung, S.-Y. Li, and N. Cai, *Network Coding Theory (Foundations and Trends(R) in Communications and Information Theory)*. Hanover, MA: Now Publishers, 2006.

[16] C. Fragouli and E. Soljanin, "Network coding fundamentals," *Found. Trends Netw.*, vol. 2, no. 1, pp. 1–133, 2007.

[17] T. Ho, R. Koetter, M. Medard, M. Effros, J. Shi, and D. Karger, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[18] G. D. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications and Signal Processing, 2nd ed. Hoboken, NJ: Wiley, 2006.

[20] A. Ozgur and S. N. Diggavi, "Approximately achieving Gaussian relay network capacity with lattice codes," presented at the IEEE Int., Symp. Information Theory, Austin, TX, Jun. 2010 [Online]. Available: http://arxiv.org/abs/1005.1284

[21] E. Perron, S. N. Diggavi, and I. E. Telatar, "On noise insertion strategies for wireless network secrecy," in *Proc. IEEE Information Theory and Applications Workshop (ITA)*, Feb. 2009, pp. 77–84.

[22] S. Avestimehr, S. N. Diggavi, and D. Tse, "Wireless network information flow," presented at the 45th Allerton Conf. Comm., Control, and Computing, Monticello, IL, Sep. 26–28, 2007.

[23] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *IEEE Trans. Inf. Theory*, vol. 44, no. 9, pp. 1726–1745, Sep. 1998.

[24] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[25] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM J. Appl. Math.*, vol. 16, no. 6, pp. 1350–1393, Nov. 1968.

[26] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "Bounds on achievable rates for general multi-terminal networks with practical constraints," in *Proc. 2nd International Workshop on Information Processing (IPSN)*, 2003, pp. 146–161.

[27] B. Bollobas, "Modern graph theory," in *Graduate Texts in Mathematics*. New York: Springer, 1998, vol. 184.

[28] A. Orlitsky and J. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[29] E. Perron, "Information-Theoretic Secrecy for Wireless Networks," Ph.D. dissertation, no 4476, EPFL, Aug. 2009.

[30] U. Nieson and S. N. Diggavi, The approximate capacity of the Gaussian N-relay diamond network. [Online]. Available: http://arxiv.org/abs/1008.2813

A. Salman Avestimehr (S'04–M'09) received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2003, and the M.S. degree in 2005 and the Ph.D. degree in 2008, both in electrical engineering and computer science, from the University of California (UC), Berkeley.

He is currently an Assistant Professor at the School of Electrical and Computer Engineering, Cornell University, Troy NY, where he has co-founded the Foundations of Information Engineering (FoIE) Center. He was also a postdoctoral scholar at the Center for the Mathematics of Information (CMI) at Caltech in 2008. His research interests include information theory, communications, and networking.

Dr. Avestimehr has received a number of awards including the NSF CAREER award (2010), the David J. Sakrison Memorial Prize from the UC Berkeley EECS Department (2008), and the Vodafone U.S. Foundation Fellows Initiative Research Merit Award (2005).

Suhas N. Diggavi (S'93–M'98) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1998.

After completing the Ph.D. degree, he was a Principal Member Technical Staff in the Information Sciences Center, AT&T Shannon Laboratories, Florham Park, NJ. After that, he was on the faculty at the School of Computer and Communication Sciences, EPFL, where he directed the Laboratory for Information and Communication Systems (LICOS). He is currently a Professor in the Department of Electrical Engineering, University of California, Los Angeles. His research interests include wireless communications networks, information theory, network data compression, and network algorithms.

Dr. Diggavi is a recipient of the 2006 IEEE Donald Fink prize paper award, the 2005 IEEE Vehicular Technology Conference best paper award, and the Okawa foundation research award. He is currently an editor for ACM/IEEE TRANSACTIONS ON NETWORKING and the IEEE TRANSACTIONS ON INFORMATION THEORY. He has eight issued patents.

David N. C. Tse (M'96–SM'907–F'09) received the B.A.Sc. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1989, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1991 and 1994, respectively.

From 1994 to 1995, he was a postdoctoral member of technical staff at AT & T Bell Laboratories. Since 1995, he has been with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, where he is currently a Professor.

Dr. Tse received a 1967 NSERC 4-year graduate fellowship from the government of Canada in 1989, a NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, the Information Theory Society Paper Award in 2003, and the 2009 Frederick Emmons Terman Award from the American Society for Engineering Education. He has given plenary talks at international conferences such as ICASSP in 2006, MobiCom in 2007, CISS in 2008, and ISIT in 2009. He was the Technical Program co-chair of the International Symposium on Information Theory in 2004 and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003. He is a coauthor, with P. Viswanath, of the text *Fundamentals of Wireless Communication*, which has been used in over 60 institutions around the world.