Wireless Network Information Flow: A Deterministic Approach

A. Salman Avestimehr[1], Suhas N. Diggavi[2], and David N. C. Tse[1]

[1]Wireless Foundations, University of California, Berkeley, CA, USA.
[2]School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland.

Abstract

In contrast to wireline networks, not much is known about the flow of information over wireless networks. The main barrier is the complexity of the signal interaction in wireless channels in addition to the noise in the channel. A widely accepted model is the the additive Gaussian channel model, and for this model, the capacity of even a network with a single relay node is open for 30 years.

In this paper, we present a *deterministic* approach to this problem by focusing on the signal interaction rather than the noise. To this end, we propose a deterministic channel model which is analytically simpler than the Gaussian model but still captures two key wireless channel properties of *broadcast* and *superposition*. We consider a model for a wireless relay network with nodes connected by such deterministic channels, and present an exact characterization of the end-to-end capacity when there is a single source and one or more destinations (all interested in the same information) and an arbitrary number of relay nodes. This result is a natural generalization of the celebrated max-flow min-cut theorem for wireline networks. We then use the insights obtained from the analysis of the deterministic model to study information flow for the Gaussian wireless relay network. We present an achievable rate for general Gaussian relay networks and show that it is within a constant number of bits from the cut-set bound on the capacity of these networks. This constant depends on the number of nodes in the network, but not the values of the channel gains or the signal-to-noise ratios. We show that existing strategies cannot achieve such a constant-gap approximation for arbitrary networks and propose a new *quantize-map-and-forward* scheme that does. We also give several extensions of the approximation framework including robustness results (through compound channels), half-duplex constraints and ergodic channel variations.

I. INTRODUCTION

Two main distinguishing features of wireless communication are:

- *broadcast*: wireless users communicate over the air and signals from any one transmitter is 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 creating signal interference among users. This is in direct contrast to wireline networks, where transmitter-receiver pairs can be thought of as isolated point-to-point links. While there has been significant progress in understanding network flow over wired networks [1], [2], [3], [4], [5], not much 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 the two key features of wireless communication of 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 received noise. Like the Gaussian model, our deterministic model is linear, but unlike the Gaussian model, operations are on a finite field.[1] We exploit these properties to provide a complete characterization of the capacity of a network of nodes connected

[1] The simplicity of scalar finite-field channel models has also been noted in [6].

by such deterministic channels. It is a natural generalization of the max-flow min-cut theorem for wireline networks.

The second step is to utilize the insights from the deterministic approach to find ”near-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 the noisy Gaussian network, an *approximate* max-flow min-cut result can be shown, where the approximation is within an additive constant which is universal over channel parameters and the SNR of operation but depends 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-and-forward*, provides such a universal approximation.

In this paper we focus on the unicast and multicast communication scenarios. In the unicast scenario, one source wants to communicate to a single destination. In the multicast scenario the 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 no interference between different information streams in the network. There is only one information stream. 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-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 wireline 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 [7] and the most general strategies for this network were developed by Cover and El Gamal [8]. There has also been a significant effort to generalize these ideas to arbitrary multi-relay networks with simple interaction models. An early attempt was done in the Ph.D. thesis of Aref [9] where a max-flow min-cut result was established to characterize the unicast capacity of a deterministic broadcast relay network which has *no superposition*. This was an early precursor to network coding which established the multicast capacity of wireline networks, a deterministic capacitated graph without broadcast or superposition [1], [2], [3]. These two ideas were combined in [10], which established a max-flow min-cut characterization for multicast flows for ”Aref networks” which had general (deterministic) broadcast with no superposition. However, such complete characterizations are not known for arbitrary (even deterministic) networks with both broadcast and superposition. One notable exception is the work [11] 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 [12], 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 signal interaction 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 [13] the ”diamond” network of parallel relay channel 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 [14]. Kramer et. al. [15] 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, and hence the approximation guarantees for these schemes is unclear. As we will see in section III, these strategies do not yield an approximation guarantee for general networks.

B. Outline of the paper

We first develop a analytically simple linear deterministic 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 the noisy (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 signal interaction model (not necessarily linear nor finite-field) and establish an achievable rate for an arbitrary network with such interaction. For the special case of *linear finite-field* deterministic models, this achievable rate matches an upper bound to the capacity, therefore the complete characterization is possible. The achievable strategy involves each node randomly mapping the received signal to a transmit 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-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 information-theoretic cut-set upper bound on capacity. This constant depends on the topological parameters of the network (number of nodes in the network), but not on the values of the channel gains. 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 remove the noise is within a constant gap to the capacity of the Gaussian relay network.

In Section VIII, we show that this scheme has the desirable property that the relay nodes do not need the knowledge of the channel gains, in the sense it achieves approximately the capacity of a compound relay network where the gains come from a class of channels. Therefore, 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: signal strength, broadcast and superposition.

A. Modeling signal strength

Consider the *real* scalar Gaussian model for 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 signal-to-noise ratio (SNR) is captured in terms of channel gains. So h is a *fixed* real number representing the channel gain (signal strength), and

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

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

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

To get an intuitive understanding of this capacity formula let us write the received signal in equation (1), y, in terms of the binary expansions of x and z. For simplicity assume h, x and z are positive real numbers, then 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,

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

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 ($\sum_{i=1}^{\infty} \left(x(i+n) + z(i)\right) 2^{-i}$) to the first summation ($2^n \sum_{i=1}^{n} x(i)2^{-i}$) we can intuitively model 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 (MSB) and the lowest level being the least significant bit (LSB). 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. Clearly there is a correspondence between n and SNR in dB scale,

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

As we notice in this simplified model there is no background noise any more and hence it is a *deterministic model*. Pictorially the deterministic model corresponding to the AWGN channel is shown in Figure 1. In this figure, at the transmitter there are several small circles. Each circle represents a signal level and a binary digit can be put for transmission at each signal level. Depending on n, which represents the channel gain in dB scale, the transmitted bits at the first n signal levels will be received clearly at the destination. However the bits at other signal levels will not go through the channel.

These signal levels can potentially be created by using a multi-level lattice code in the AWGN channel [16]. 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. Therefore, if we think of the transmit signal, $\mathbf{x}$, as a binary vector of length q, then the deterministic channel delivers only its first n bits to the destination. We can algebraically write this input-output relationship by shifting $\mathbf{x}$ down by $q-n$ elements or more precisely

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

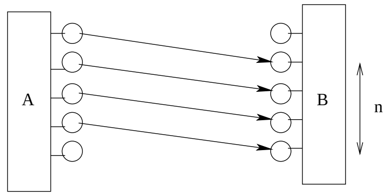


Fig. 1. Pictorial representation of the deterministic model for point-to-point channel.

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,

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

Clearly the capacity of this deterministic point-to-point channel is n, where $n = \lceil \frac{1}{2} \log \mathsf{SNR} \rceil^+$. It is interesting to note that this is a within $\frac{1}{2}$-bit approximation of the capacity of the AWGN channel[2]. In the case of complex Gaussian channel we set $n = \lceil \log \mathsf{SNR} \rceil^+$ and we get a within 1-bit approximation 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. Assume there are only two receivers. The received SNR at receiver i is denoted by SNR_i for $i = 1, 2$. Without loss of generality assume $\mathsf{SNR}_2 \leq \mathsf{SNR}_1$. Consider the binary expansion of the transmitted signal, x. Then we can deterministically model the Gaussian broadcast channel as the following:

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

[2]Note that this connection is only in the capacity without a formal connection in coding scheme or a direct translation of the capacity.

The deterministic model in some sense abstracts away the use of superposition coding and successive interference cancellation decoding in the Gaussian broadcast channel. Therefore the first 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 for a Gaussian broadcasr channel is shown in figure 2 (a). In this particular example $n_1 = 5$ and $n_2 = 2$, therefore both users receive the first two most significant bits of the transmitted signal. However user 1 (strong user) receives additional three bits from the next three signal levels of the transmitted signal. There is also the same correspondence between n and channel gains in dB: $n_i \leftrightarrow \lceil \log \mathsf{SNR}_i \rceil^+$, for $i = 1, 2$.

To analytically demonstrate how closely we are modeling the Gaussian BC channel, the capacity region of Gaussian BC channel and deterministic BC channel are shown in Figure 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 (*i.e.* if a pair (R_1, R_2) is in the capacity region of the deterministic BC then there is a pair within one bit per component of (R_1, R_2) that is in the capacity region of the Gaussian BC)[3]. However, this is only the worst case gap and in a typical case that SNR_1 and SNR_2 are very different the gap is much smaller than one bit.

C. Modeling superposition

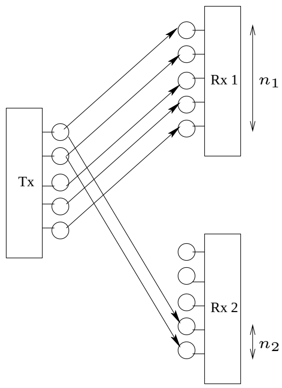
Consider a superposition scenario in which two users are simultaneously transmitting to a node. In the Gaussian model the received signal can be written as

$$y = h_1 x_1 + h_2 x_2 + z. \tag{10}$$

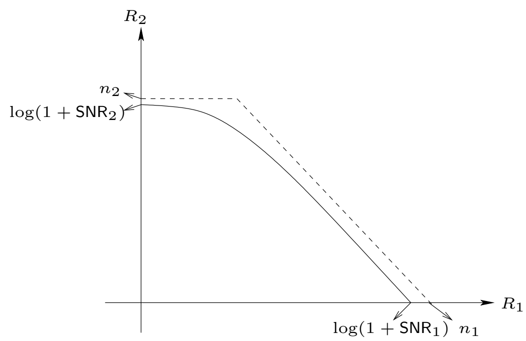
To intuitively see what happens in superposition in the Gaussian model, we again write the received signal, y, in terms of the binary expansions of x_1, x_2 and z. Assume x_1, x_2 and z are all real numbers smaller than one, and also the channel gains are

$$h_i = \sqrt{\mathsf{SNR}_i}, \quad i = 1, 2 \tag{11}$$

[3]A cautionary note is that as in the point-to-point case the connection is not formed in the coding scheme but just in capacity regions.



(a) Pictorial representation of the deterministic model for Gaussian BC



(b) Capacity region of Gaussian BC channel (solid line). Capacity region of deterministic BC channel (dashed line)

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).

Without loss of generality assume $\mathsf{SNR}_2 < \mathsf{SNR}_1$. Then we have

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i) 2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i) 2^{-i} + \sum_{i=-\infty}^{\infty} z(i) 2^{-i} \tag{12}$$

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

$$y = 2^{\frac{1}{2}\log \mathsf{SNR}_1} \sum_{i=1}^{\infty} x_1(i) 2^{-i} + 2^{\frac{1}{2}\log \mathsf{SNR}_2} \sum_{i=1}^{\infty} x_2(i) 2^{-i} + \sum_{i=1}^{\infty} z(i) 2^{-i} \tag{13}$$

or,

$$\begin{aligned} y \;\approx\; & 2^{n_1} \sum_{i=1}^{n_1-n_2} x_1(i) 2^{-i} + 2^{n_2} \sum_{i=1}^{n_2} \left(x_1(i+n_1-n_2) + x_2(i)\right) 2^{-i} \\ & + \sum_{i=1}^{\infty} \left(x_1(i+n_1) + x_2(i+n_2) + z(i)\right) 2^{-i} \end{aligned} \tag{14}$$

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 the following:

- That part of x_1 that is above SNR_2 ($x_1(i)$, $1 \le i \le n_1 - n_2$) is received clearly without any interaction from x_2.
- The remaining part of x_1 that is above noise level ($x_1(i)$, $n_1 - n_2 < i \le n_1$) and that part of x_2 that is above noise level ($x_1(i)$, $1 \le i \le n_2$) interact with 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 interaction between 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 interaction by the modulo 2 sum of the bits that are arrived at the same signal level. Pictorially the deterministic model for a Gaussian MAC channel is shown in figure 4 (a). Analogous to the deterministic model for the point-to-point channel, as seen in Figure II-C, we can write,

$$\mathbf{y} = \mathbf{S^{q-n_1} x_1} \oplus \mathbf{S^{q-n_2} x_2} \tag{15}$$

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 transmit and received signals respectively and $\mathbf{S}$ is a $q \times q$ shift matrix. There is also the same relationship between n_i's and the channel gain in dB: $n_i \leftrightarrow \lceil \log \mathsf{SNR}_i \rceil^+$, for $i = 1, 2$. Note that if one wants to make a connection between the deterministic model and real Gaussian MAC channel (rather than complex) a factor of $\frac{1}{2}$ is necessary.

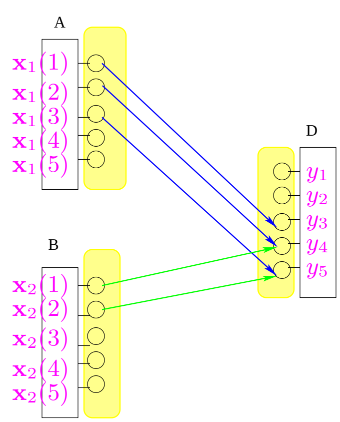
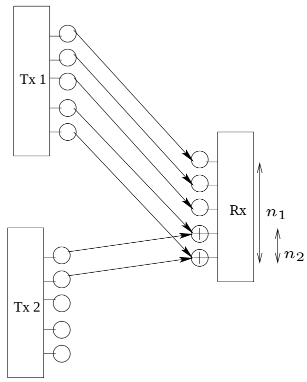
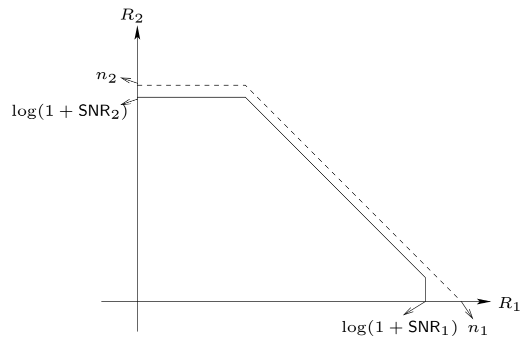


Fig. 3. Algebraic representation of shift matrix deterministic model.

Now compared to simple point-to-point case we now have interaction between the bits that are received at the same signal level at the receiver. However, we limit the receiver to observe only the modulo 2 summation of those bits that arrive at the same signal level. In some sense this way of modeling interaction is similar to the collision model. In the collision model if two packets arrive simultaneously at a receiver, both are dropped; similarly here if two bits arrive



(a) Pictorial representation of the deterministic MAC.



(b) Capacity region of Gaussian MAC. (solid line). Capacity region of deterministic MAC.(dashed line)

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

simultaneously at the same signal level the receiver gets only their modulo 2 sum, which means it cannot figure out any of them. On the other hand, unlike in the simplistic 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.

Now we can apply this model to Gaussian multiple access channel (MAC), in which

$$y = h_1 x_1 + h_2 x_2 + z \tag{16}$$

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. Without loss of generality assume $\mathsf{SNR}_2 < \mathsf{SNR}_1$. The capacity region of this channel is well-known to be the set of non-negative pairs (R_1, R_2) satisfying

$$R_i \leq \log(1+\mathsf{SNR}_i), \quad i = 1,2 \tag{17}$$

$$R_1 + R_2 \leq \log(1+\mathsf{SNR}_1+\mathsf{SNR}_2) \tag{18}$$

This region is plotted with solid line in figure 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

$$R_2 \leq n_2 \quad (19)$$

$$R_1 + R_2 \leq n_1 \quad (20)$$

where $n_i = \log \mathsf{SNR}_i$ for $i = 1, 2$. This region is plotted with dashed line in figure 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 (*i.e.* if a pair (R_1, R_2) is in the capacity region of the deterministic MAC then there is a pair within one bit per component of (R_1, R_2) that is in the capacity region of the Gaussian MAC). The intuitive explanation for this is that in real addition once two bounded signals are added together the magnitude increases however, it can only become as large as twice the maximum size of individual ones. Therefore the cardinality size of summation 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 size 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.

D. Linear finite-field deterministic model

The relay network consists of 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 a 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:

$$\mathbf{y}_j[t] = \sum_{i \in \mathcal{N}_j} \mathbf{S}^{q-n_{ij}} \mathbf{x}_i[t] \quad (21)$$

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

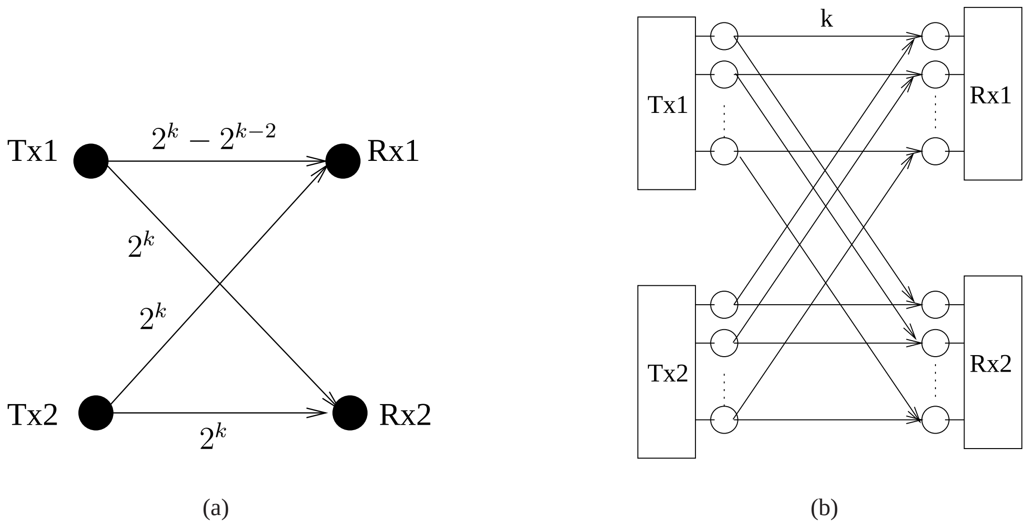


Fig. 5. An example of a 2×2 Gaussian MIMO channel is shown in (a). The corresponding linear finite field deterministic MIMO channel is shown in (b).

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 Figure 5 (a), where k is an integer larger than 2. The channel matrix is given by:

$$\mathbf{H} = 2^k \begin{pmatrix} \frac{3}{4} & 1 \\ 1 & 1 \end{pmatrix} \tag{22}$$

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

$$n_{11} = \lceil \log_2 h_{11} \rceil^+ = \lceil \log_2 (2^k - 2^{k-2}) \rceil^+ = k \tag{23}$$

$$n_{12} = n_{21} = n_{22} = \lceil \log_2 2^k \rceil^+ = k \tag{24}$$

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

$$C_{\text{LFF}} \quad = \quad \text{rank} \begin{pmatrix} I_k & I_k \\ I_k & I_k \end{pmatrix} = k \tag{25}$$

$$\tag{26}$$

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 very 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 our analysis of the Gaussian network.

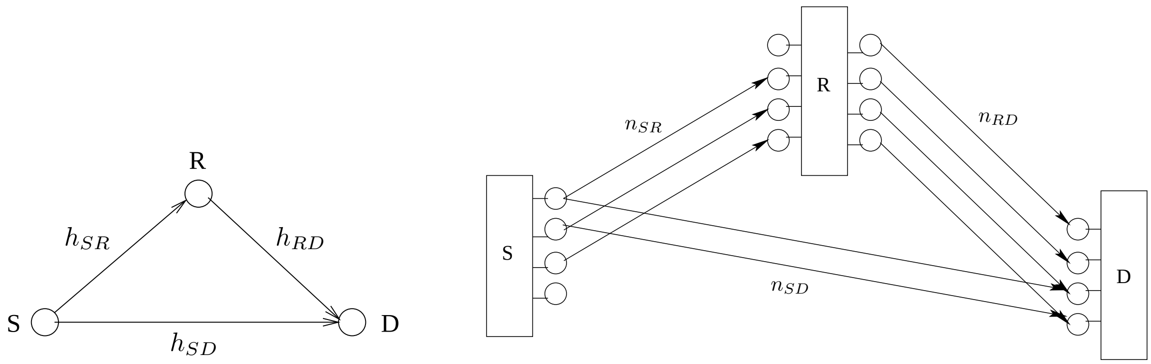
III. Motivation of our approach

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 a near optimal relaying scheme for the Gaussian case and using this insight to provably approximate the capacity of these networks under the Gaussian model within a constant number of bits. We progress from the 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. In fact we can show that the set of strategies that yield a universal approximation shrink as we progress to more complex networks. This illustrates the power of the deterministic model to provide insights into transmission techniques for the noisy networks.

For any network, there is a natural information-theoretic cut-set bound [17], which upper bounds the reliable transmission rate R. Applied to the relay network, we have the cut-set upper bound $\overline{C}$ on its capacity:

$$\overline{C} = \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega | \mathbf{X}_{\Omega^c}) \tag{27}$$

where $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts (partitions). In words, the cut 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,



(a) The Gaussian relay channel (b) The linear finite-field deterministic relay channel

Fig. 6. The relay channel: (a) Gaussian model, (b) Linear finite-field deterministic model.

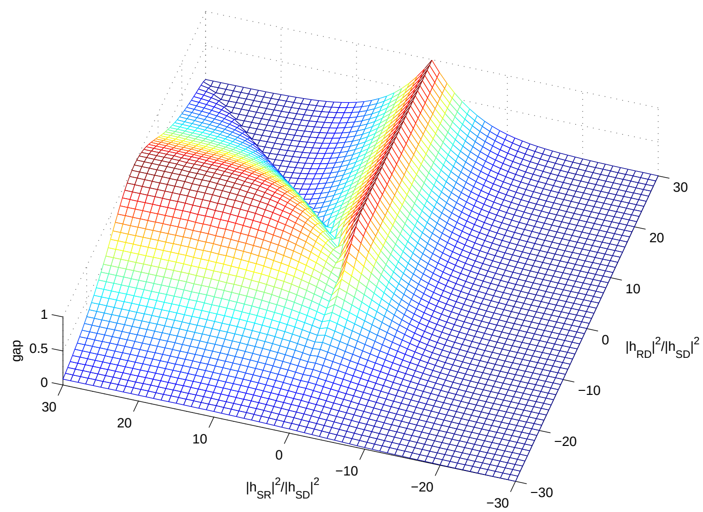


Fig. 7. The gap between cut-set upper bound and achievable rate of decode-forward scheme in the Gaussian relay channel for different channel gains (in dB scale).

this is simple 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 figure 6 (a). We examine whether it is possible to approximate its capacity uniformly (uniform over all channel gains). To answer this question positively 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 Figure 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 can be broadcasted from the relay, and the maximum number of bits that the destination can receive. Therefore.

$$C^d_{relay} \leq \min\left(\max(n_{SR}, n_{SD}), \max(n_{RD}, n_{SD})\right) \tag{28}$$

$$= \begin{cases} n_{SD}, & \text{if } n_{SD} > \min\left(n_{SR}, n_{RD}\right); \\ \min\left(n_{SR}, n_{RD}\right), & \text{otherwise.} \end{cases} \tag{29}$$

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

Note that equation (29) 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 innovative bits. In the example of Figure 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 interfere 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.

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 I. ■

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. Since in a wireless scenario the channel gains differ significantly, this happens very rarely. In figure 7 the gap between the achievable rate of decode-forward scheme and the cut-set upper bound is plotted for different channel gains. In this figure x and y axis are respectively representing the channel gains from relay to destination and source to relay normalized by the

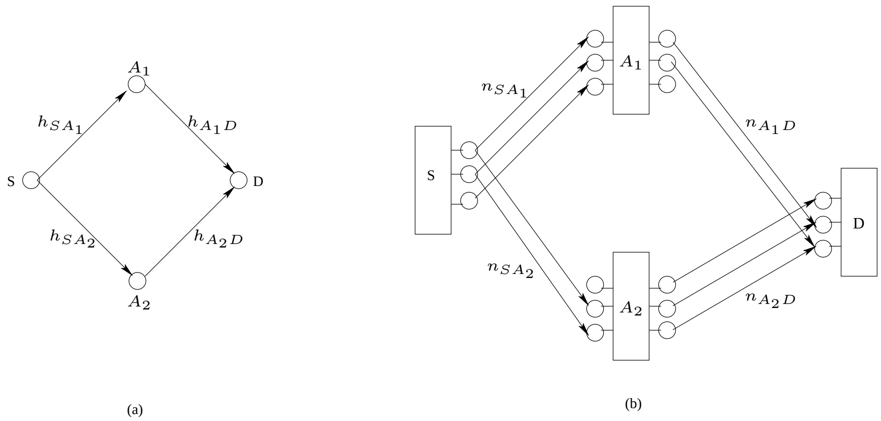


Fig. 8. Diamond network with two relays: (a) Gaussian model, (b) Linear finite-field deterministic model.

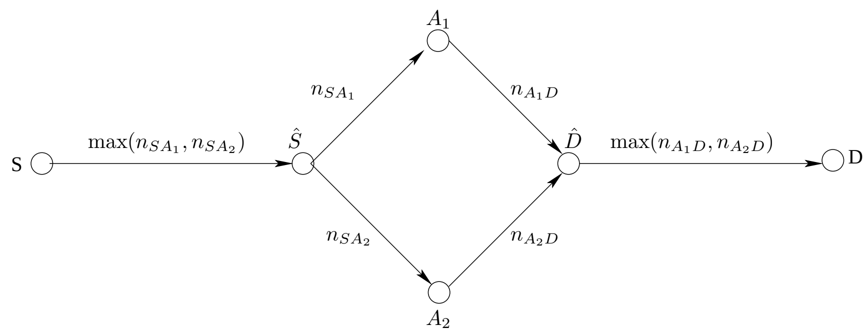


Fig. 9. Wireline diamond network

gain of the direct link (source to destination) in dB scale. The z axis shows the value of the gap (in bits/sec/Hz).

Note that the deterministic network in Figure 6 (b), suggests that several other relaying strategies are also optimal. For example doing a compress and forwarding will also achieve the cut-set bound. Moreover a ”network coding” strategy of sending the sum (or linear combination) of the received bits will also be optimal as long as the destination receives linearly independent combinations. 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. As we will see, this set shrinks as we go to larger relay networks.

B. Diamond network

Now consider the diamond Gaussian relay network, with two relays, as shown in Figure 8 (a). Brett Schein introduced this network in his Ph.D. thesis [13] and investigated its capacity. However the capacity of this network is still an open problem. We examine whether it is possible to uniformly approximate its capacity.

First we build the corresponding linear finite field deterministic model for this relay network as shown in Figure 8 (b). To investigate its capacity first we relax the interactions between incoming links at each node and create the wireline network shown in Figure 9. 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 wireline 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 wireline diamond network is equal to the value of its minimum cut. Hence

$$C^d_{\text{diamond}} \leq C^w_{\text{diamond}} = \min\left\{\max(n_{SA_1}, n_{SA_2}), \max(n_{A_1D}, n_{A_2D}), n_{SA_1} + n_{A_2D}, n_{SA_2} + n_{A_1D}\right\} \tag{30}$$

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 the wireline diamond network is achieved by a routing solution. It is not also difficult to see that we can indeed mimic this routing solution in the linear finite-field deterministic diamond network and send the same amount of information through non-interfering links from source to relays and then 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.
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 at the end 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-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 II. ■

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 relay network.

For example consider the linear-finite field deterministic diamond network shown in Figure 10 (a). Clearly 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.

Now we show that this 1-bit gap can be translated into an unbounded gap in the corresponding Gaussian network, as shown in Figure 10 (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,

$$\overline{C} \approx 3 \log a \tag{31}$$

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

$$R_{DF} \leq 2 \log a \tag{32}$$

Therefore, as a gets larger, the gap between the achievable rate of decode-forward strategy and the cut-set upper bound (31) increases.

Now 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 in Figure 10 (a). We notice that to achieve a rate of 3 from the source to the destination, the bit at the lowest signal level of the source's signal should go through A_1 while the remaining two are going 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 figure 10 (c) and 10 (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/unit time. Therefore, for this channel realization, 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

Now we look at a more complicated relay network with four relays, as shown in Figure 11. 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 Figure 12 (a). Now focus on the relaying strategy that is pictorially shown in Figure 13. 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. As one can note, in this optimal scheme the relay

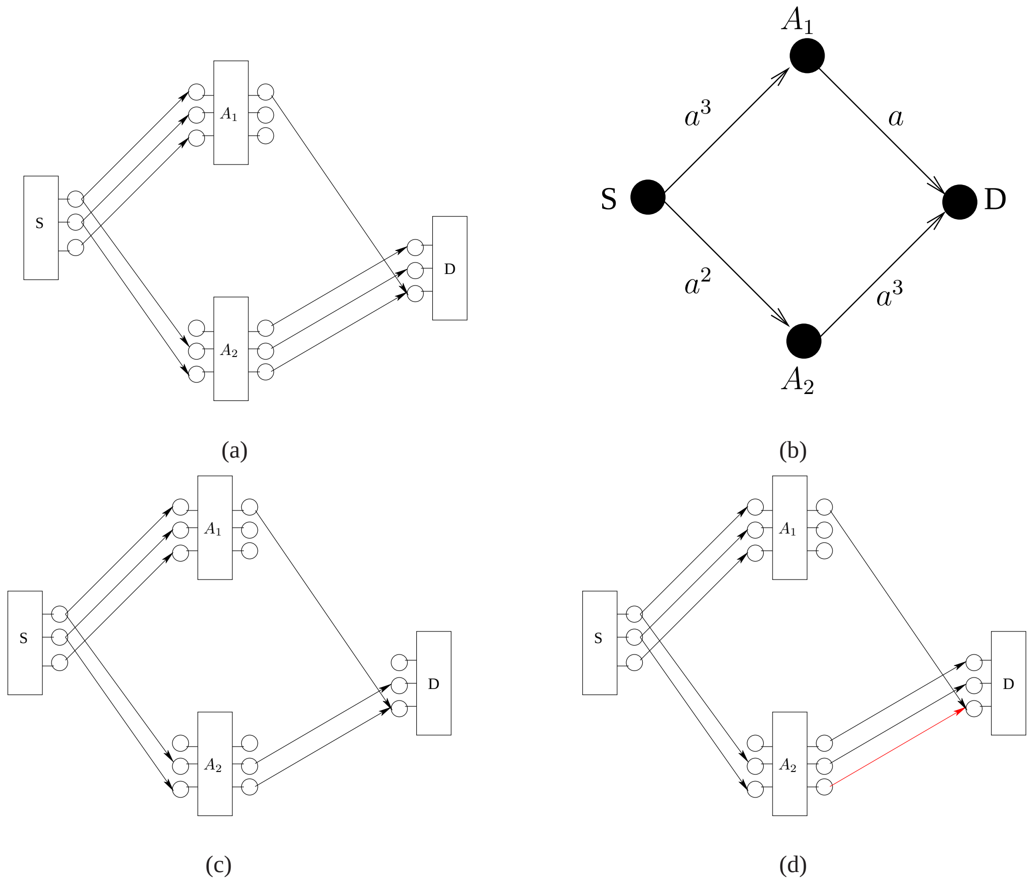
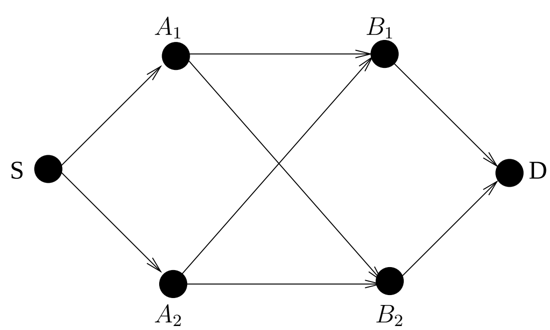


Fig. 10. An example of the linear finite field deterministic diamond network is shown in (a). The corresponding Gaussian network is shown in (b), with the gains chosen such that the ratio of the gains in dB scale match the ratios of the gains in the deterministic network. The effective network when R_2 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).



(a)

Fig. 11. A two layer relay network with four relays.

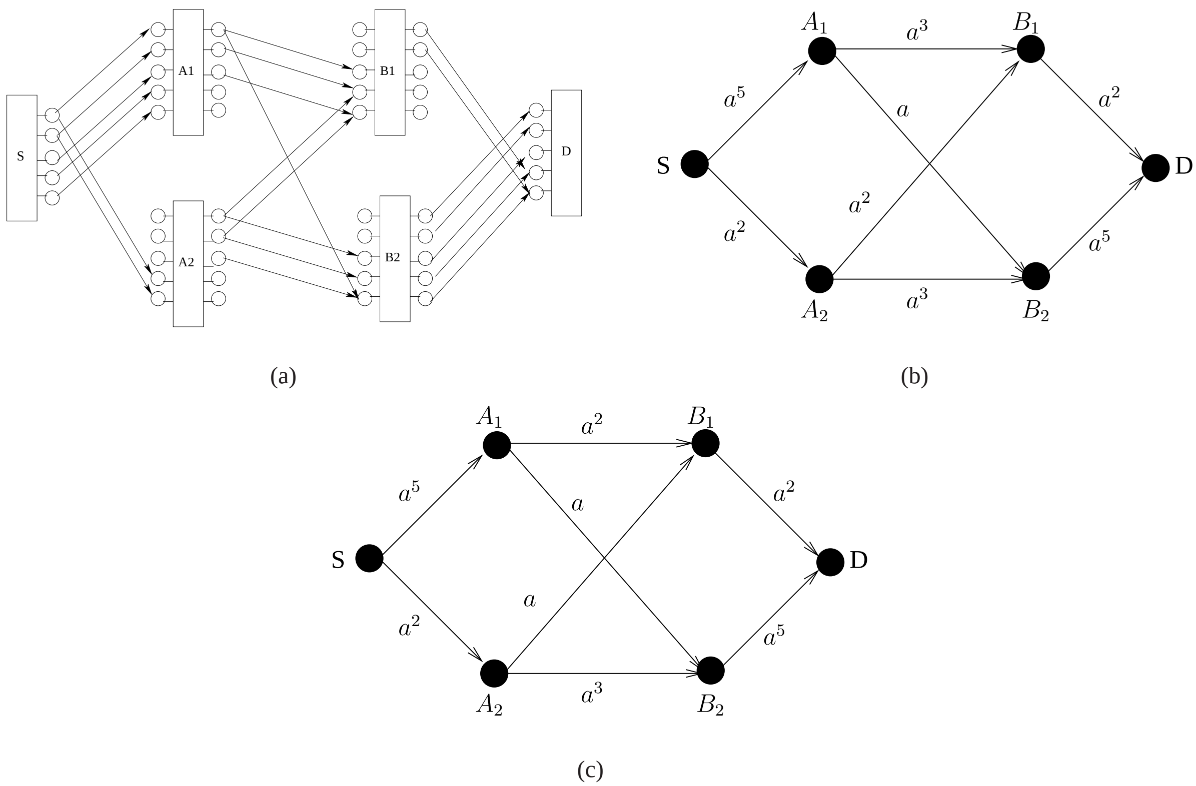


Fig. 12. An example of a four relay linear finite filed deterministic relay network is shown in (a). The corresponding Gaussian relay network is shown in (b). The effective Gaussian network for compress-forward strategy is shown in (c).

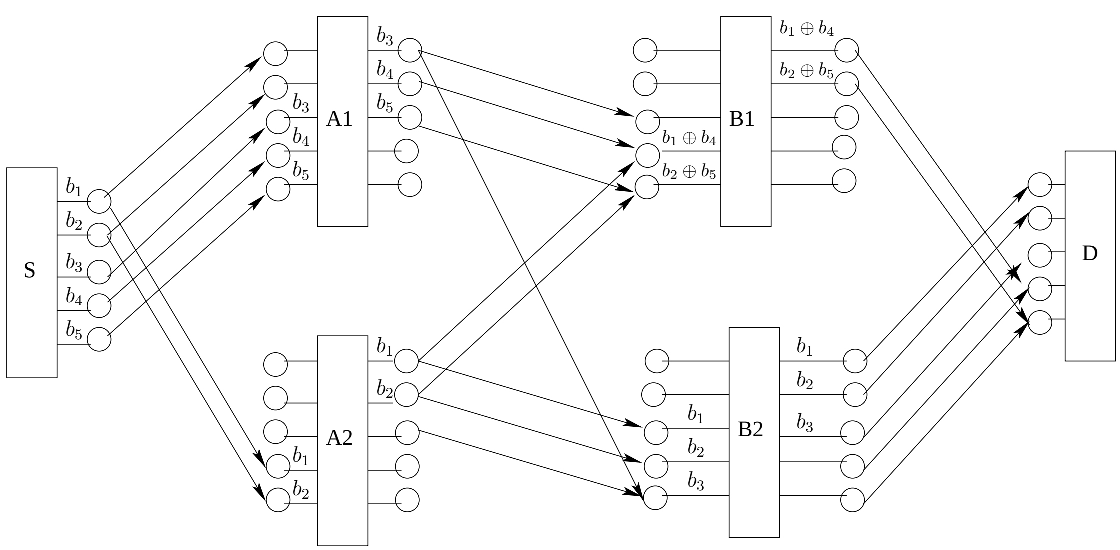


Fig. 13. Demonstration of a capacity achieving strategy.

B_1 is not decoding or partially decoding a message; it is forwarding the last two LSB's. One may wonder if this is necessary, or in another words is any choice of partial decode-forward strategy suboptimal in this example? To answer this question, note that any partial decode-forward scheme can be visualized as different flows of information going from S to D that do not get mixed in the network. Now since all transmit 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 partial decode-forward scheme in this example and hence it is always suboptimal.

The last stage in the above scheme can actually be interpreted as a compress-and-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-and-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, is the dependency between them become apparent. In fact, it can be shown that a compress-and-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-and-forward, partial decode-and-forward, amplify-and-forward and Gaussian compress-and-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 received signal at each signal level is just an equation of the message sent by the source, and the optimal strategy is to simply shuffle and combine these received equations at each relay and forward them. This insight leads to a natural *quantize-and-forward* strategy for noisy (Gaussian) relay networks. The strategy for each relay is to quantize the received signal at a 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 that 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

$$\mathbf{y}_j[t] = \mathbf{g}_j(\{\mathbf{x}_i[t]\}_{i\in\mathcal{V}}), \tag{33}$$

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 [9] which only allow broadcast.

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

$$\overline{C} = \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c};\mathbf{X}_\Omega|\mathbf{X}_{\Omega^c}) \tag{34}$$

$$\stackrel{(a)}{=} \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) \tag{35}$$

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 networks with general deterministic interaction models.

Theorem 4.1: A rate of

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{X}_i)} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}). \tag{36}$$

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

$$\max_{\prod_{i\in\mathcal{V}} p(\mathbf{X}_i)} \min_{D\in\mathcal{D}} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}). \tag{37}$$

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

This achievability result in Theorem 4.1 extends the results in [10] where only deterministic broadcast network (with no interference) were considered. Note that when we compare (36) to the cut-set upper bound in (35), 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 indeed 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, (21), and using (35) since we have a deterministic network, we get:

$$\overline{C} \;=\; \max_{p(\{\mathbf{X}_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} H(\mathbf{Y}_{\Omega^c}|\mathbf{X}_{\Omega^c}) \overset{(b)}{=} \min_{\Omega\in\Lambda_D} \operatorname{rank}(\mathbf{G}_{\Omega,\Omega^c}) \tag{38}$$

where as before, $\Lambda_D = \{\Omega : S \in \Omega, D \in \Omega^c\}$ is all source-destination cuts and $\mathbf{G}_{\Omega,\Omega^c}$ is the transfer matrix associated with that cut, *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 (21). This is also illustrated in Figure 14. Step (b) follows since in a 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 $\{x_i\}_{i\in\mathcal{V}}$ and the optimum value of each cut Ω is logarithm of the size of the range space of the transfer matrix $\mathbf{G}_{\Omega,\Omega^c}$ associated with that cut.

The following are our main results for linear finite-field deterministic relay networks,

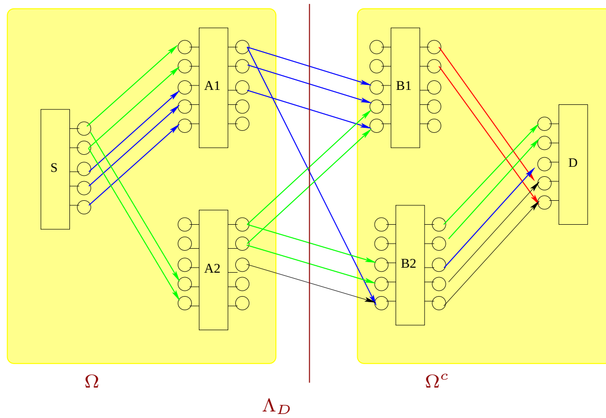


Fig. 14. Illustration of cut-set bound and cut-set transfer matrix $\mathbf{G}_{\Omega,\Omega^c}$.

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

$$C = \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{39}$$

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

$$C = \min_{D \in \mathcal{D}} \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{40}$$

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

Note that the results in Theorems 4.1, 4.2, 4.3 and 4.4, applies to networks with arbitrary topology and could have cycles. For a single source-destination pair the result in Theorem 4.3 generalizes the classical max-flow min-cut theorem for wireline networks and for multicast, the result in Theorem 4.4 generalizes the network coding result in [1] where in both these earlier results, the communication links are orthogonal, *i.e.* no broadcast or multiple access interference. Moreover, as we will see in the proof, the encoding functions at the relay nodes (for the linear finite-field model) could be restricted to linear functions to obtain the result in Theorem 4.3. We also note that, Theorems 4.3 and 4.4 are just corollaries of Theorems 4.1 and 4.2.

B. Gaussian relay networks

In the Gaussian model the signals get attenuated by complex gains and added together with Gaussian noise at each receiver (the Gaussian noises at different receivers being independent of each other.). More formally the received signal $\mathbf{y}_j$ at node $j \in \mathcal{V}$ and time t is given by

$$\mathbf{y}_j[t] = \sum_{i \in \mathcal{V}} \mathbf{H}_{ij} \mathbf{x}_i[t] + \mathbf{z}_j[t] \tag{41}$$

where $\mathbf{H}_{ij}$ is a complex matrix where element represents the channel gain from a transmitting antenna in node i to a receiving 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 following is our main result for Gaussian relay networks which is proved in Section VI.

Theorem 4.5: Given a Gaussian relay network, we can achieve all rates R up to $\overline{C} - \kappa$. Therefore the capacity of this network satisfies

$$\overline{C} - \kappa \leq C \leq \overline{C}, \tag{42}$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (27), and κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

The gap (κ) holds for all values of the channel gains and is relevant particularly in the high rate regime. This constant gap result is a far stronger result than the degree of freedom result, not only because it is non-asymptotic but also because it is uniform in the many channel SNR's. This is the first constant gap approximation of the capacity of Gaussian relay networks. As shown in Section IV, 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 an network with arbitrary topology and could have cycles.

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 (36, 37), generalizes the proof for the the linear case, Theorems 4.1 and 4.2 are proved in Appendix III. The insight from these results suggest the quantize-and-forward strategy for noisy (Gaussian) relay networks. We use this insight as well as proof ideas generated 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 going through an example which then is generalized.

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. Since the proof idea of Theorems 4.1 and 4.2, which are achievable rates for arbitrary deterministic networks, generalize the proof for linear finite field networks, we present their proofs in Appendix III.

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. [1]. We do this in sections V-A. Next, in Section V-B, we extend the result to an arbitrary network by expanding the network over time[4]. Since the time-expanded network is layered and we can apply our result in the first step to it and complete the proof.

A. Layered networks

The network given in Figure 16 is an example of a *layered* network where the number of hops for each path from S to D is equal to 3[5].

In this section we give the encoding scheme for the layered linear finite-field deterministic relay networks in Section V-A.1. In Section V-A.2 we illustrate the proof techniques on a simple

[4]The concept of time-expanded network is also used in [1], 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.

[5]Note that in the equal path network we do not have "self-interference" since all path-lengths from S to D in terms of "hops" are equal, though as we will see in the analysis that can easily be taken care of. However we do allow for self-interference in the model and we choose to handle such loops, and more generally cyclic networks, through time-expansion as will be seen in Section V-B.

linear unicast relay network example. In Section V-A.3 we prove main Theorems 4.3, 4.4 for layered networks.

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 transmit signals in the next block. For the model (21), we will use linear mappings $f_j(\cdot)$, *i.e.*,

$$\mathbf{x}_j = \mathbf{F}_j \mathbf{y}_j, \tag{43}$$

where $\mathbf{F}_j$ is chosen uniformly randomly over all matrices in $\mathbb{F}_2^{qT \times qT}$. Each relay does the encoding prescribed by (43). Given the knowledge of all the encoding functions $\mathbf{F}_j$ at the relays, the decoder $D \in \mathcal{D}$, attempts to decode each message w_k sent by the source. This encoding strategy is illustrated in Figure 15.

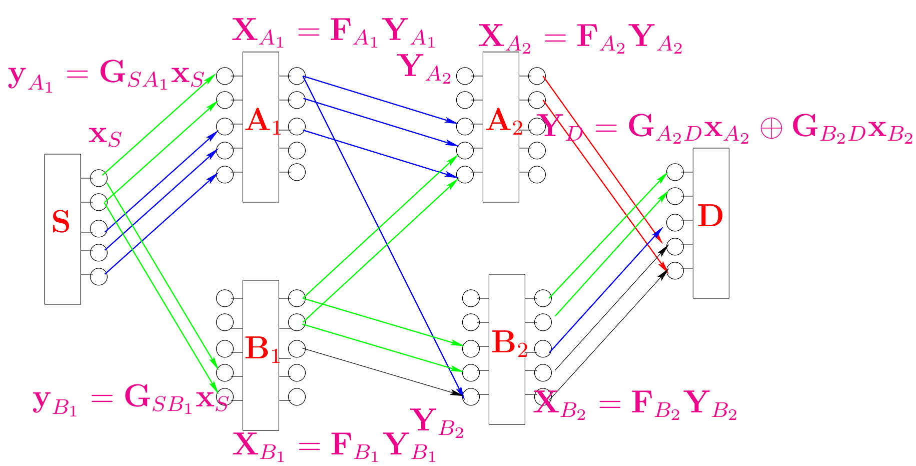


Fig. 15. Illustration of linear encoding strategy. Note that the vectors $\mathbf{Y}_j$ represent received signals $\mathbf{y}_j[t], t = 1, \ldots, T$ which have been collected over T time units. Similarly $\mathbf{X}_j = [\mathbf{x}_j[1], \ldots, \mathbf{x}_j[T]]$ are transmitted over T time-units.

Now suppose message w_k is sent by the source in block k, then since each relay j operates only on block of lengths T and we have a layered structure, 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. As a result the key simplification that occurs for layered networks is that the messages do not get mixed with each other.

Now, given the knowledge of all the encoding functions $\mathbf{F}_j$ at the relays and signals received over block $k + l_D$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_k sent by the source.

2) Proof illustration: In order to illustrate the proof ideas of Theorem (4.1) we examine the network shown in Figure 16.

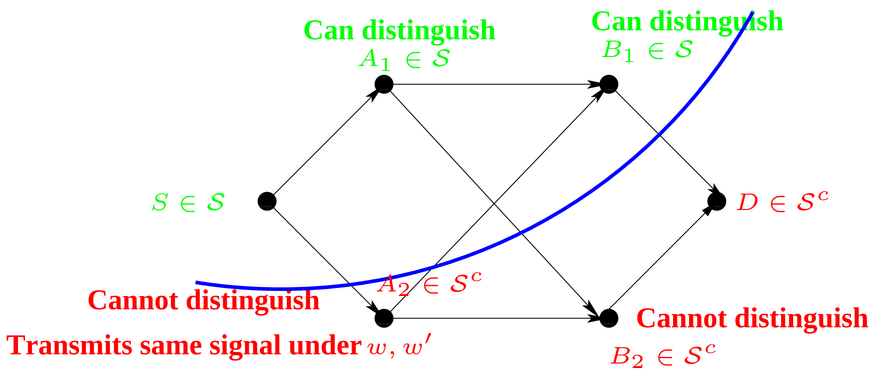


Fig. 16. An example of layered relay network. Nodes on the left hand side of the cut can distinguish between messages w and w', while nodes on the right hand side can not.

Since we have a layered network, 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*, which is that messages w, w' are distinguishable at any node j if $\mathbf{y}_j(w) \neq \mathbf{y}_j(w')$.

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

$$P_e \leq 2^{RT} \mathbb{P}\left\{w \to w'\right\} = 2^{RT} \mathbb{P}\left\{\mathbf{y}_D(w) = \mathbf{y}_D(w')\right\}. \tag{44}$$

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,

$$\mathbb{P}\left\{w \to w'\right\} = \sum_{\Omega \in \Lambda_D} \underbrace{\mathbb{P}\left\{\text{Nodes in } \Omega \text{ can distinguish } w, w' \text{ and nodes in } \Omega^c \text{ cannot}\right\}}_{\mathcal{P}} \tag{45}$$

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 (45). For example, consider the cut $\Omega = \{S, A_1, B_1\}$ shown in Figure 16. A necessary condition for the distinguishability set to be this cut 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:

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

Now we can write

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{A}_2, \mathcal{B}_2, \mathcal{D}, \mathcal{A}_1^c, \mathcal{B}_1^c\} && (46)\\
&= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2, \mathcal{A}_1^c | \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D}, \mathcal{B}_1^c | \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} && (47)\\
&\leq \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} && (48)\\
&= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{B}_2\} && (49)
\end{aligned}$$

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

$$X_S \rightarrow (Y_{A_1}, Y_{A_2}) \rightarrow (Y_{B_1}, Y_{B_2}) \rightarrow Y_D \tag{50}$$

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

$$\mathbb{P}\{\mathcal{A}_2\} = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{S,A_2})(\mathbf{x}_S(w) - \mathbf{x}_S(w')) = \mathbf{0}\right\} = 2^{-T\text{rank}(\mathbf{G}_{S,A_2})}, \tag{51}$$

since the random mapping given in (43) induces independent uniformly distributed $\mathbf{x}_S(w), \mathbf{x}_S(w')$. Here, $\otimes$ is the Kronecker matrix product[6]. 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

[6]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 $A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$

if $\mathbf{y}_{A_1}(w) \neq \mathbf{y}_{A_1}(w')$, then the random mapping given in (43) induces independent uniformly distributed $\mathbf{x}_{A_1}(w), \mathbf{x}_{A_1}(w')$ Therefore, we get

$$\mathbb{P}\{\mathcal{B}_2|\mathcal{A}_1^c, \mathcal{A}_2\} = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{A_1,B_2})(\mathbf{x}_{A_1}(w) - \mathbf{x}_{A_1}(w')) = \mathbf{0}\right\} = 2^{-T\text{rank}(\mathbf{G}_{A_1,B_2})}. \quad (52)$$

Similarly we get,

$$\mathbb{P}\{\mathcal{D}|\mathcal{B}_1^c, \mathcal{B}_2\} = \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{B_1,D})(\mathbf{x}_{B_1}(w) - \mathbf{x}_{B_1}(w')) = \mathbf{0}\right\} = 2^{-T\text{rank}(\mathbf{G}_{B_1,D})}. \quad (53)$$

Putting these together, since all three would need to occur, we see that in (45), for the network in Figure 16, we have,

$$\begin{aligned} \mathcal{P} &\leq 2^{-T\text{rank}(\mathbf{G}_{S,A_2})} 2^{-T\text{rank}(\mathbf{G}_{A_1,B_2})} 2^{-T\text{rank}(\mathbf{G}_{B_1,D})} \\ &= 2^{-T\{\text{rank}(\mathbf{G}_{S,A_2}) + \text{rank}(\mathbf{G}_{A_1,B_2}) + \text{rank}(\mathbf{G}_{B_1,D})\}}. \end{aligned} \quad (54)$$

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 (54) is exactly $2^{-T\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}$. Therefore, by substituting this back into (45) and (44), we see that

$$P_e \leq 2^{RT} |\Lambda_D| 2^{-T \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega,\Omega^c})}, \quad (55)$$

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.

3) Proof of Theorems 4.3 and 4.4 for 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 decoder D can be upper bounded by,

$$P_e \leq 2^{RT} \mathbb{P}\left\{w \to w'\right\} = 2^{RT} \mathbb{P}\left\{\mathbf{y}_D(w) = \mathbf{y}_D(w')\right\}. \quad (56)$$

Similar to Section V-A.2, we can write

$$\mathbb{P}\left\{w \to w'\right\} = \sum_{\Omega \in \Lambda_D} \underbrace{\mathbb{P}\left\{\text{Nodes in } \Omega \text{ can distinguish } w, w' \text{ and nodes in } \Omega^c \text{ cannot}\right\}}_{\mathcal{P}} \quad (57)$$

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 can not distinguish between w and w', i.e. $\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w')$.

Similar to Section V-A.2 we can write,

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1}, l = 2, \dots, l_D\} && (58)\\
&= \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1} | \mathcal{R}_j, \mathcal{L}_{j-1}, j = 2, \dots, l-1\} && (59)\\
&\leq \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_j, \mathcal{L}_j, j = 2, \dots, l-1\} && (60)\\
&\overset{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} && (61)
\end{aligned}$$

where (a) is true due to Markovian nature of the layered network. Note that as in the example network of Section V-A.2, 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'. Therefore, all the nodes in R_{l-1} cannot distinguish between w, w' and therefore

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

Therefore, just as in Section V-A.2, we see that the probability that

$$\begin{aligned}
\mathbb{P}\{\mathcal{R}_l | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} &= \mathbb{P}\left\{\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w') | \mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w'), \mathbf{y}_{R_{l-1}}(w) = \mathbf{y}_{L_{R-1}}(w')\right\} && (62)\\
&= \mathbb{P}\left\{\mathbf{y}_{R_l}(w) = \mathbf{y}_{R_l}(w') | \mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w'), \mathbf{x}_{R_{l-1}}(w) = \mathbf{x}_{L_{R-1}}(w')\right\} && (63)\\
&= \mathbb{P}\left\{(\mathbf{I}_T \otimes \mathbf{G}_{L_{l-1}, R_l})(\mathbf{x}_{L_{l-1}}(w) - \mathbf{x}_{L_{l-1}}(w')) = \mathbf{0} | \mathbf{y}_{L_{l-1}}(w) \neq \mathbf{y}_{L_{l-1}}(w')\right\} && (64)\\
&\overset{(a)}{=} 2^{-T \text{rank}(\mathbf{G}_{L_{l-1}, R_l})}. && (65)
\end{aligned}$$

where $\mathbf{G}_{L_{l-1}, R_l}$ is the transfer matrix from transmit 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 (43) induces independent uniformly distributed $\mathbf{x}_{L_{l-1}}(w), \mathbf{x}_{L_{l-1}}(w')$.

Therefore we get

$$\mathcal{P} \leq \prod_{l=2}^{d} 2^{-T\text{rank}(\mathbf{G}_{L_{l-1},R_l})} = 2^{-T\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}. \tag{66}$$

Therefore, by substituting this back into (57) and (56), we see that

$$P_e \leq 2^{RT}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D}\text{rank}(\mathbf{G}_{\Omega,\Omega^c})}, \tag{67}$$

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 for layered networks:

Theorem 5.1: The multicast capacity C of a layered linear finite-field deterministic relay network is given by,

$$C = \min_{D\in\mathcal{D}}\min_{\Omega\in\Lambda_D}\text{rank}(\mathbf{G}_{\Omega,\Omega^c}). \tag{68}$$

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

B. Arbitrary networks (not necessarily layered)

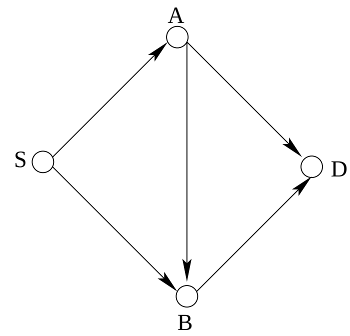
First we formally describe the encoding strategy:

1) Encoding for arbitrary linear deterministic relay network : We have a single source S with message $W \in \{1, 2, \ldots, 2^{TKR}\}$ which is encoded by the source S into a signal over KT transmission times (symbols), giving an overall transmission rate of R. Relay j operates over blocks of time T symbols, and at the k-th block uses a linear mapping $f_j^{[k]}(.)$ to map its received symbols from all the previous $k-1$ blocks of T symbols to transmit signals in the next block, *i.e.*,

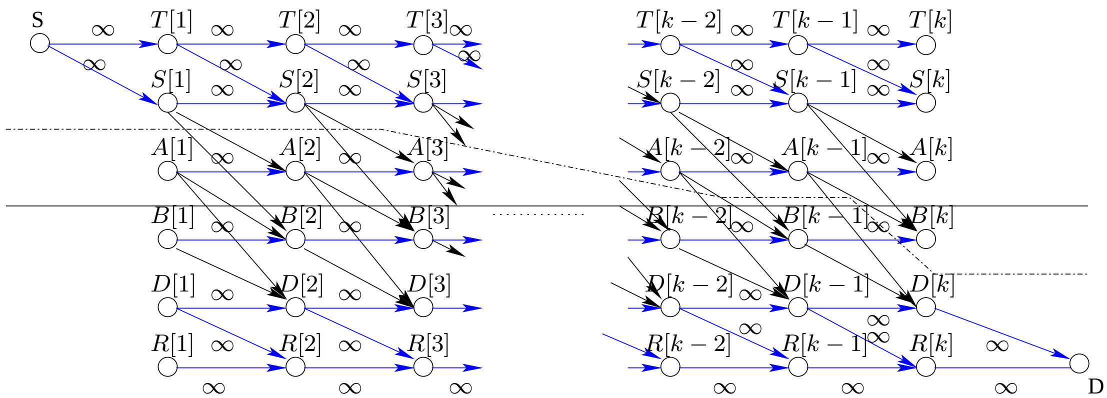
$$\mathbf{x}_j^{(k)} = \mathbf{F}_j^{(k)}\mathbf{y}_j^{(1):(k-1)} \tag{69}$$

where $\mathbf{F}_j^{(k)}$ is chosen uniformly randomly over all matrices in $\mathbb{F}_2^{qT\times(k-1)qT}$. Each relay does the encoding prescribed by (69). Given the knowledge of all the encoding functions $\mathbf{F}_j^{(k)}$ at the relays, the decoder $D \in \mathcal{D}$, attempts to decode each message w_k sent by the source. This encoding strategy is illustrated in Figure 15.

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. First is that we can explicitly represent our relaying scheme by unfolded the network over time



(a) An example of general deterministic network



(b) Unfolded deterministic network. An example of steady cuts and dipping cuts are respectively shown by solid and dotted lines.

Fig. 17. An example of a general deterministic network with un equal paths from S to D is shown in (a). The corresponding unfolded network is shown in (b).

to create a layered deterministic network. The idea is to unfold the network to K stages such that i-th stage is representing what happens in the network during $(i-1)T$ to $iT-1$ symbol times. For example in figure 17 (a) a network with unequal paths from S to D is shown. Figure 17(b) shows the unfolded form of this network. As we notice each node $v \in \mathcal{V}$ is appearing at stage $1 \leq i \leq K$ as $v[i]$. There are additional nodes: $T[i]$'s and $R[i]$'s. These nodes are just virtual transmitters and receivers that are put to buffer and synchronize the network. Since all communication links connected to these nodes ($T[i]$'s and $R[i]$'s) are modelled as wireline links without any capacity limit they would not impose any constraint on the network. Note that there are also infinite-capacity links between nodes such as $A[1]$, $A[2]$, $\ldots$, that are copies of the same node at different blocks.

Lemma 5.2: Assume $\mathcal{G}$ is a linear deterministic network and $\mathcal{G}_{\text{unf}}^{(K)}$ is a network obtained by unfolding $\mathcal{G}$ over K time steps (as shown in figure 17). Then a communication rate of:

$$R < \frac{1}{K} \min_{\Omega_{\text{unf}} \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega_{\text{unf}}, \Omega_{\text{unf}}^c}) \tag{70}$$

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

Proof: By unfolding $\mathcal{G}$ we get an acyclic layered deterministic network. Therefore by theorem 5.1 we can achieve the rate

$$R_{\text{unf}} < \min_{\Omega_{\text{unf}} \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega_{\text{unf}}, \Omega_{\text{unf}}^c}) \tag{71}$$

in the time-expanded graph. Since it takes K steps to translate and achievable scheme in the time-expanded graph to an achievable scheme in the original graph, then the Lemma is proved.

■

Note that the achievability scheme that we used to prove Lemma 5.2 was obtained by applying the encoding scheme described in section V-A.1 to the network that is unfolded over K blocks. This translates to the encoding scheme defined in Section V-B.1 for a general deterministic relay network.

If we look at the different cuts in the time-expanded graph we notice that there are two types of cuts with non-infite value. One type separates the nodes at different stages identically. An example of such a steady cut is drawn with solid line in figure 17 (b) which separates $\{S, A\}$ from $\{B, D\}$ at all stages. Clearly each steady cut in the time-expanded graph corresponds to a cut in the original graph and moreover its value is K times the value of the corresponding cut in the original network. However there is another type of cut which does not behave identically at different stages. An example of such a dipping cut is drawn with dotted line in figure 17 (b). A dipping cut can be thought of as a list of straight cuts with a number of downward transitions between them. Note that if a cut has an upward transition, then one of the infinite capacity links will cross him (from left to right), hence the cut-value becomes infinity. So we will only need to consider the cuts with downward transitions. However, since the number of downward transitions is at most $K - |\mathcal{V}|$, the value of a dipping cut is at least equal to a weighted sum of $K - |\mathcal{V}|$ cut-values of the original network, therefore its value can not be smaller than $K - |\mathcal{V}|$ times the min-cut value of in the original network. Therefore

$$\min_{\Omega_{\text{unf}} \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega_{\text{unf}}, \Omega_{\text{unf}}^c}) \geq (K - |\mathcal{V}|) \min_{\Omega \in \Lambda_D} \text{rank}(\mathbf{G}_{\Omega, \Omega^c}) \tag{72}$$

This combined with Lemma 5.2 completes the proof of Theorem 4.3.

VI. Gaussian relay networks

So far, we have focused on noiseless relay networks. As we discussed in Sections II and III, our 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 noisy Gaussian relay networks. This is what we will accomplish in this section.

Theorem 4.5 is our main result for Gaussian relay networks and the rest of this section is devoted to prove it. Similar to the deterministic case, first we focus on networks that have a layered structure that the messages do not get mixed in the network. The proof of the result for layered network is done in section VI-A. Next, we extend the result to an arbitrary network by expanding the network over time, as done in Section V. Since the time-expanded network is layered and we can apply our result in the first step to it and complete the proof. We first prove the theorem for the single antenna case, then at the end we extend it to a multiple antenna scenario.

A. Layered Gaussian relay networks

In this section we prove Theorem 4.5 for a 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 noiseless 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 together with the whole network creates an inner code with provides certain mutual information between the source and 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 inner code, and hence the proof is complete. The system diagram of our coding strategy is illustrated in Figure 18.

We now explicitly describe our encoding strategy:

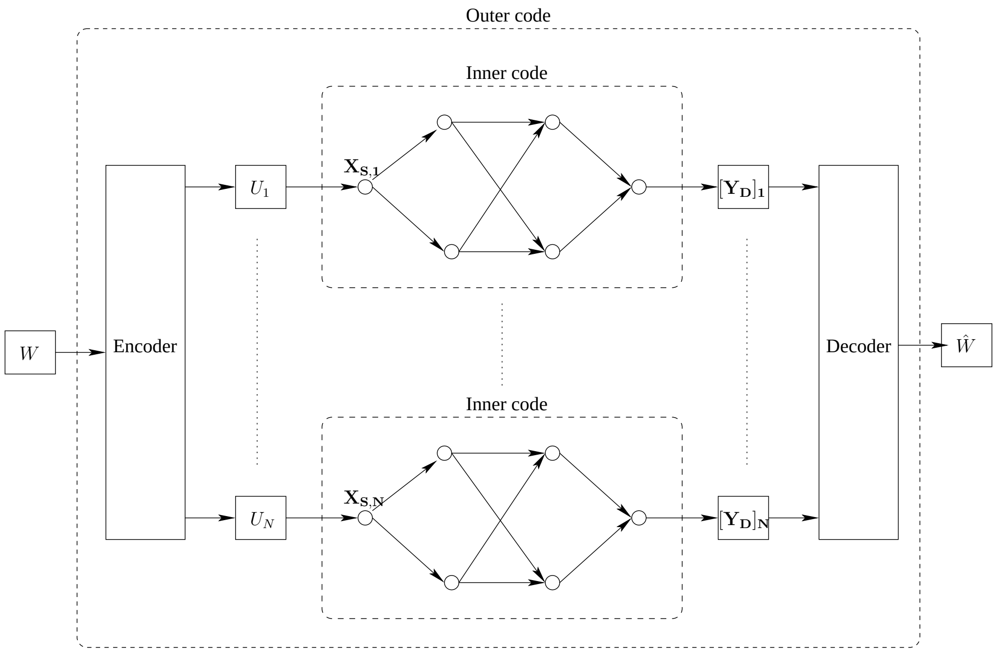


Fig. 18. System diagram.

2) Encoding for layered Gaussian relay networks: We first define the following quantization operation:

Definition 6.1: The quantization operation $[.] : \mathbb{C} \mapsto \mathbb{C}$, maps a complex number $c = x + iy$ to $[c] = [x] + i[y]$, where $[x]$ and $[y]$ are the closest integers to x and y, respectively.

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

a) Inner code: Each symbol of the inner code is represented by $u \in \{1, \dots, 2^{R_{\text{in}}T}\}$, 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 $F_i : (\mathbb{Z}^T, \mathbb{Z}^T) \mapsto \mathcal{T}_{x_i}$ which maps each quantized received signal vector of length T independently into an i.i.d. $\mathcal{CN}(0,1)$ random vector of length T. A particular realization of F_i is denoted by f_i. Now

- Source: maps each inner code symbol $u \in \{1, \dots, 2^{R_{\text{in}}T}\}$ 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: Source S has a sequence of messages $w_j \in \{1, 2, \dots, 2^{TNR}\}$, $j = 1, 2, \dots$. Each message is encoded by the source into N inner code symbols, $u_1, \dots, u_N$. Each inner code symbol, is then sent via the inner code over T transmission times, giving an overall transmission rate of R.

Now, given the knowledge of all the encoding functions F_i's at the relays and received signals $[\mathbf{y}_D]_1, \ldots, [\mathbf{y}_D]_N$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_j 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-A.2.

Note that

$$\frac{1}{T} I(U; [\mathbf{Y}_D] | F_{\mathcal{V}}) \geq \frac{1}{T} I(U; [\mathbf{Y}_D] | \mathbf{Z}_{\mathcal{V}}, F_{\mathcal{V}}) - \frac{1}{T} H([\mathbf{Y}_D] | U, F_{\mathcal{V}}) \tag{73}$$

where $\mathbf{Z}_{\mathcal{V}}$ is the sequence of the channel noises at all nodes in the network. The first term on the right hand side of (73) is the average end to end mutual information conditioned on the noise sequence. Once we condition on a noise sequence, the network turns into a deterministic network. We then 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 (73) 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.

Definition 6.2: We define

$$\overline{C}_{i.i.d.} \triangleq \min_{\Omega} I(X_\Omega; Y_{\Omega^c} | X_{\Omega^c}) \tag{74}$$

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

Lemma 6.3: Given a layered Gaussian relay network $\mathcal{G}$, assume all nodes perform the operation described in subsection VI-A.2 (a) and the inner code symbol U is distributed uniformly over $\{1, \ldots, 2^{R_{\text{in}} T}\}$. Then

$$I(U; [\mathbf{Y}_D] | \mathbf{Z}_{\mathcal{V}}, F_{\mathcal{V}}) \geq R_{\text{in}} T - (1 + \min\{1, 2^{|\mathcal{V}|} 2^{-T(\overline{C}_{iid} - |\mathcal{V}| \log \max_l |\mathcal{V}_l|) - R_{\text{in}})}\} R_{\text{in}} T) \tag{75}$$

where $\overline{C}_{iid}$ is defined in Definition 6.2 and $\mathcal{V}_l$ is the set of nodes at layer l.

Proof:

Suppose the destination attempts to detect the transmitted symbol U at the source given the received signal, all the mappings, channel gains and $\mathbf{z}_{\mathcal{V}}$. 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 that would have been received under u'. This leads to a notion of *distinguishability* for a fixed $\mathbf{z}_{\mathcal{V}}$, which is that symbol values u, u' are distinguishable at any node j if $[\mathbf{y}_j(u)] \neq [\mathbf{y}_j(u')]$. Hence,

$$\mathbb{P}\left\{u \to u' | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\right\} = \sum_{\Omega \in \Lambda_D} \underbrace{\mathbb{P}\left\{\text{Nodes in } \Omega \text{ can distinguish } u, u' \text{ and nodes in } \Omega^c \text{ cannot} | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\right\}}_{\mathcal{P}} \quad (76)$$

Consider 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 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 by assumption $\mathbb{P}\{\mathcal{L}_1\} = 1$ (source distinguishes).

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1}, l = 2, \ldots, l_D | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\} && (77)\\
&= \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1} | \mathcal{R}_j, \mathcal{L}_{j-1}, j = 2, \ldots, l-1, \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\} && (78)\\
&\leq \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_j, \mathcal{L}_j, j = 2, \ldots, l-1, \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\} && (79)\\
&\stackrel{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\} && (80)\\
&= \prod_{l=2}^{l_D} \mathbb{P}\{[\mathbf{y}_{R_l}(w)] = [\mathbf{y}_{R_l}(w')] | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\} && (81)
\end{aligned}$$

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

Now note that if A and B are complex matrices, then

$$[A] = [B] \Rightarrow ||A - B||_\infty \leq \sqrt{2} \quad (82)$$

Therefore by (81) and (82) we have

$$\mathcal{P} \leq \prod_{l=2}^{l_D} \mathbb{P}\{||\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')||_\infty \leq \sqrt{2}|\mathcal{R}_{l-1}, \mathcal{L}_{l-1}, \mathbf{Z}_\mathcal{V} = \mathbf{z}_\mathcal{V}\} \tag{83}$$

$$\stackrel{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{||\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')||_\infty \leq \sqrt{2}|\mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} \tag{84}$$

$$= \prod_{l=2}^{l_D} \mathbb{P}\{\forall 1 \leq j \leq T : ||H_l \left(x_{L_{l-1},j}(u) - x_{L_{l-1},j}(u')\right)||_\infty \leq \sqrt{2}|\mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} \tag{85}$$

$$\stackrel{(b)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\forall 1 \leq j \leq T : ||H_l \left(x_{L_{l-1},j}(u) - x_{L_{l-1},j}(u')\right)||_\infty \leq \sqrt{2}|\mathcal{L}_{l-1}\} \tag{86}$$

where (a) is true since $\mathbf{y}_{R_l}(u) - \mathbf{y}_{R_l}(u')$ does not depend on the noise realization, and (b) is true since the nodes in $R_{l-1}(\Omega)$ transmit the same codeword under both u and u'. H_l denotes the MIMO channel from L_{l-1} to R_l (the transfer matrix from left side of the cut to the right side of the cut at the l-th stage). Now since $\mathbf{x}_{L_{l-1}}(u) \neq \mathbf{x}_{L_{l-1}}(u')$, due to the random mapping they are independent and the difference is just an i.i.d complex Gaussian random vector with distribution $\mathcal{CN}(0,1)$. Now, we state the following Lemma which is proved in Appendix IV.

Lemma 6.4: Assume $\mathbf{x}_i = [x_{i,1}, \cdots, x_{i,T}]$, $i = 1, \ldots, m$, are vectors of i.i.d. $\mathcal{CN}(0, \sigma^2)$ Gaussian vectors of length T, and $H \in \mathbb{C}^{n\times m}$ is an $n \times m$ matrix. Then, we have

$$\mathbb{P}\left\{\forall 1 \leq j \leq T : ||H[x_{1,j}, \cdots, x_{m,j}]^t||_\infty \leq \sqrt{2}\right\} \leq 2^{-T\left(I(X;HX+Z) - \min(m,n)(1+\log\frac{n}{\sigma^2})\right)} \tag{87}$$

where X and Z are i.i.d complex normal Gaussian vectors of length m and n respectively.

By applying Lemma 6.4 to (86), we get

$$\mathbb{P}\{\forall 1 \leq j \leq T : ||H_l \left(x_{L_{l-1},j}(u) - x_{L_{l-1},j}(u')\right)||_\infty \leq \sqrt{2}|\mathcal{L}_{l-1}\} \leq 2^{-T\left(I\left(X_{L_{l-1}};Y_{R_l}|X_{R_{l-1}}\right) - \min(|L_{l-1}|,|R_l|)(1+\log|R_l|)\right)} \tag{88}$$

where X_i, $i \in \mathcal{V}$, are iid with normal (Gaussian) distribution. Hence

$$\mathcal{P} \leq \prod_{l=2}^{l_D} 2^{-T\left(I\left(X_{L_{l-1}};Y_{R_l}|X_{R_{l-1}}\right) - \min(|L_{l-1}|,|R_l|)(1+\log\frac{|R_l|}{\pi})\right)} \tag{89}$$

$$\leq 2^{-T(\overline{C}_{iid} - |\mathcal{V}|\log\max_l|\mathcal{V}_l|)} \tag{90}$$

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

Now, the average probability of symbol detection error at the destination can be upper bounded by,

$$P_e = \mathbb{P}\left\{\hat{u} \neq u | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\right\} \leq 2^{R_{\text{in}}T}\mathbb{P}\left\{u \rightarrow u' | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}\right\} \tag{91}$$

By the union bound we have

$$P_e \leq \sum_{\Omega} 2^{-T(\overline{C}_{iid} - |\mathcal{V}| \log \max_l |\mathcal{V}_l|) - R)} \leq 2^{|\mathcal{V}|} 2^{-T(\overline{C}_{iid} - |\mathcal{V}| \log \max_l |\mathcal{V}_l|) - R)} \tag{92}$$

Now, by using the Fano's inequality we get,

$$\begin{aligned}
I(U; [\mathbf{Y}_D] | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) &= H(U) - H(U | [\mathbf{Y}_D], \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) && (93)\\
&= R_{\text{in}}T - H(U | [\mathbf{Y}_D], \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}}) && (94)\\
&= R_{\text{in}}T - \mathbb{E}_{F_{\mathcal{V}}}[H(U | [\mathbf{Y}_D], \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}} = f_{\mathcal{V}})] && (95)\\
&\overset{\text{Fano}}{\geq} R_{\text{in}}T - (1 + \mathbb{E}_{F_{\mathcal{V}}}[\mathbb{P}\left\{\hat{u} \neq u | \mathbf{Z}_{\mathcal{V}} = \mathbf{z}_{\mathcal{V}}, F_{\mathcal{V}} = f_{\mathcal{V}})\right\}] R_{\text{in}}T) && (96)\\
&= R_{\text{in}}T - (1 + P_e R_{\text{in}}T) && (97)\\
&\geq R_{\text{in}}T - (1 + \min\{1, 2^{|\mathcal{V}|} 2^{-T(\overline{C}_{iid} - |\mathcal{V}| \log \max_l |\mathcal{V}_l|) - R_{\text{in}})}\} R_{\text{in}}T) && (98)
\end{aligned}$$

Hence, the proof is complete.

■

We now bound the second term on the RHS of (73),

Lemma 6.5: Given a layered Gaussian relay network $\mathcal{G}$, assume all nodes perform the operation described in subsection VI-A.2 (a). Then

$$H([\mathbf{Y}_D] | U, F_{\mathcal{V}})] \leq 14T|\mathcal{V}| \tag{99}$$

Proof: See Appendix V. ■

Lemma 6.6: Given a Gaussian relay network $\mathcal{G}$, then

$$\overline{C} - \overline{C}_{iid} < \kappa_2 \tag{100}$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$, $\overline{C}_{iid}$ is defined in Definition 6.2 and $\kappa_2 = |\mathcal{V}|$.

Proof: See Appendix VI. ■

Lemma 6.7: Given a layered Gaussian relay network $\mathcal{G}$, assume all nodes perform the operation described in subsection VI-A.2 (a) and the inner code symbol U is distributed uniformly over $\{1,\ldots,2^{R_{\text{in}}T}\}$. Then

$$\frac{1}{T}I(U;[\mathbf{Y}_D]|F_{\mathcal{V}}) \geq R_{\text{in}} - 14|\mathcal{V}| - (\frac{1}{T} + \min\{1, 2^{|\mathcal{V}|}2^{-T(\overline{C}-|\mathcal{V}|(1+\log\max_l|\mathcal{V}_l|))-R_{\text{in}})}\}R_{\text{in}}) \quad (101)$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ and $\mathcal{V}_l$ is the set of nodes at layer l.

Proof: By using equation (73) and Lemmas 6.3, 6.5 and 6.6 we have,

$$\frac{1}{T}I(U;[\mathbf{Y}_D]|F_{\mathcal{V}}) \geq \frac{1}{T}I(U;[\mathbf{Y}_D]|\mathbf{Z}_{\mathcal{V}},F_{\mathcal{V}}) - \frac{1}{T}H([\mathbf{Y}_D]|U,F_{\mathcal{V}}) \quad (102)$$

$$\overset{\text{Lemma 6.3 and 6.5}}{\geq} R_{\text{in}}T - (1 + \min\{1, 2^{|\mathcal{V}|}2^{-T(\overline{C}_{iid}-|\mathcal{V}|\log\max_l|\mathcal{V}_l|)-R_{\text{in}})}\}R_{\text{in}}T) - 14T|\mathcal{V}| \quad (103)$$

$$\overset{\text{Lemma 6.6}}{\geq} R_{\text{in}} - 14|\mathcal{V}| - (\frac{1}{T} + \min\{1, 2^{|\mathcal{V}|}2^{-T(\overline{C}-|\mathcal{V}|(1+\log\max_l|\mathcal{V}_l|))-R_{\text{in}})}\}R_{\text{in}}) \quad (104)$$

■

An immediate corollary of this Lemma is that by choosing R_{in} arbitrarily close to $\overline{C}-|\mathcal{V}|(1+\log\max_l|\mathcal{V}_l|)$, and letting T be arbitrary large, for any $\delta > 0$ we get

$$\frac{1}{T}I(U;[\mathbf{Y}_D]|F_{\mathcal{V}}) \geq \overline{C} - |\mathcal{V}|(15 + \log\max_l|\mathcal{V}_l|) - \delta \quad (105)$$

Therefore there exists a choice of mappings that provides an end to end mutual information arbitrarily close to $\overline{C} - |\mathcal{V}|(15 + \log\max_l|\mathcal{V}_l|)$. Hence, we have effectively created a point to point channel from U to $[\mathbf{Y}_D]$ with certain mutual information. We can now use a good outer code to reliably send a message over N uses of this channel (as illustrated in Figure 18) at any rate up to

$$R \leq \overline{C} - |\mathcal{V}|(15 + \log\max_l|\mathcal{V}_l|) \quad (106)$$

Hence we get our main result for layered Gaussian relay networks.

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,

$$R < \overline{C} - \kappa_{\text{Lay}} \quad (107)$$

where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (27), $\kappa_{\text{Lay}} = |\mathcal{V}|(15 + \log\max_l|\mathcal{V}_l|)$ is a constant not depending on the channel gains and $\mathcal{V}_l$ is the set of nodes at layer l.

B. General Gaussian relay networks (not necessarily layered)

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

First we formally describe the encoding strategy:

1) Encoding for general Gaussian relay network: We have a single source S with a sequence of messages $w_j \in \{1, 2, \ldots, 2^{NKTR}\}$, $j = 1, 2, \ldots$. Each message is encoded by the source S into a signal over NKT transmission times (symbols), giving an overall transmission rate of R. Similar to the layered case, we have an inner code and an outer code.

a) Inner code: Similar to Section VI-A.2, the source node generates a set of $2^{KR_{\text{in}}T}$ independent complex Gaussian codewords of length KT with components distributed as i.i.d. $\mathcal{CN}(0,1)$, denoted by $\mathcal{T}_{x_S}$. Associated with each relay node i, and for each k, $k = 1, \ldots, K$, there is also a random mapping $F_i^{(k)} : (\mathbb{Z}^{kT}, \mathbb{Z}^{kT}) \mapsto \mathcal{T}_{x_i}$ which maps the quantized received signals up to block k independently into an i.i.d. $\mathcal{CN}(0,1)$ random vector of length T.

- Source: maps each inner code symbol $u \in \{1, \ldots, 2^{KR_{\text{in}}T}\}$ to $F_S(u) \in \mathcal{T}_{x_S}$ ($|\mathcal{T}_{x_S}| = 2^{R_{\text{in}}T}$) and send it in KT transmission times.
- Relay i: operates over blocks of time T symbols, and at the k-th block quantizes all received sequence ($\mathbf{y}_i^{(1)}$, ..., $\mathbf{y}_i^{(k)}$) into ($[\mathbf{y}_i^{(1)}]$, ..., $[\mathbf{y}_i^{(k)}]$), which is then randomly mapped to $F_i^{(k)}(([\mathbf{y}_i^{(1)}], \ldots, [\mathbf{y}_i^{(k)}])) \in \mathcal{T}_{x_i}$.

b) Outer code: Source S has a sequence of messages $w_j \in \{1, 2, \ldots, 2^{NKTR}\}$, $j = 1, 2, \ldots$. Each message is encoded by the source into N inner code symbols of size 2^{KRT}, $u_1, \ldots, u_N$. Each inner code symbol, is then sent via the inner code over KT transmission times (symbols).

Given the knowledge of all the encoding functions at the relays and signals received over $K - |V|$ blocks, the decoder D, attempts to decode the message w sent by the source.

Similar to the deterministic case (Section V-B), we use time expansion idea to analyze this relaying scheme. We first state the following lemma which is a corollary of Theorem 6.8.

Lemma 6.9: Given a Gaussian relay network, $\mathcal{G}$, all rates R satisfying the following condition are achievable,

$$R < \frac{1}{K}\overline{C}_{\text{unf}}^{(K)} - \kappa \tag{108}$$

where $\overline{C}_{\text{unf}}^{(K)}$ is the cut-set upper bound on the capacity of the time expanded graph associated with $\mathcal{G}$, $\kappa = |\mathcal{V}|(15 + \log|\mathcal{V}|)$.

Proof: By unfolding $\mathcal{G}$ we get an acyclic network such that all the paths from the source to the destination have equal length. Therefore, by theorem 6.8, all rates R_{unf}, satisfying the following condition are achievable in the time-expanded graph

$$R_{\text{unf}} < \overline{C}^{(K)}_{\text{unf}} - \kappa_{\text{unf}} \tag{109}$$

where $\kappa_{\text{unf}} = K|\mathcal{V}|(15 + \log \max_l |\mathcal{V}^{(K)}_{\text{unf,l}}|)$. But note that the number of nodes (not buffers) at each layer of the unfolded graph is exactly $|\mathcal{V}|$, hence $\kappa_{\text{unf}} = K|\mathcal{V}|(15 + \log |\mathcal{V}|)$. Now since it takes K steps to translate and achievable scheme in the time-expanded graph to an achievable scheme in the original graph we can achieve $\frac{1}{K}R_{\text{unf}}$ and the proof is complete. ■

Now smilar to the deterministic case, it is easy to see that,

$$\overline{C}^{(K)}_{\text{unf}} = (|K - |\mathcal{V}|)\overline{C} \tag{110}$$

Hence, by lemma 6.9 and (110), we can achieve all rates upto

$$R < \frac{K - |\mathcal{V}|}{K}\overline{C} - \kappa \tag{111}$$

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

To prove Theorem 4.5 for 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

$$R < \overline{C}_D - \kappa \tag{112}$$

where $\kappa < |\mathcal{V}|(15+\log |\mathcal{V}|)$ is a constant and $\overline{C}_D = \max_{p(\{x_j\}_{j\in\mathcal{V}})} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_\Omega|X_{\Omega^c})$ is the cut-set upper bound on the capacity from the source to D. Therefore as long as $R < \min_D \overline{C}_D - \kappa$, 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 will first quantize the received signal of each antenna at noise level and then map it to another transmit codeword. 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 by assuming that we have a network with at most *i.e.* $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)$ virtual nodes (each correspond to an antenna). Therefore from our previous analysis we know that the gap is at most $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$ 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 deterministic channel model captures certain high SNR aspects 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

$$\mathbf{y}_j[t] = \left[\sum_{i\in\mathcal{V}} \mathbf{H}_{ij} x_i[t]\right], \qquad j = 1, \ldots, |\mathcal{V}| \tag{113}$$

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.

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

$$|C_{\text{Gaussian}} - C_{\text{Truncated}}| \leq |\mathcal{V}|(36 + \log|\mathcal{V}|) \tag{114}$$

To prove this Theorem first we need the following lemma,

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 $P_{\mathbf{X}}$,

$$|I(\mathbf{X}; G\mathbf{X} + Z) - I(\mathbf{X}; [G\mathbf{X}])| \leq 21n \tag{115}$$

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

Proof: Look at appendix VII. ∎

Now we prove Theorem 7.1.

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

$$|\overline{C}_{\text{Gaussian}} - \overline{C}_{\text{Truncated}}| \leq 21|\mathcal{V}| \tag{116}$$

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

$$C_{\text{Truncated}} \geq \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}^{\text{truncated}}; X_\Omega | X_{\Omega^c}) \tag{117}$$

Now by Lemma 6.6 and Lemma 7.2 we know have the following

$$\begin{aligned} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}^{\text{truncated}}; X_\Omega | X_{\Omega^c}) &\geq I(Y_{\Omega^c}^{\text{Gaussian}}; X_\Omega | X_{\Omega^c}) - 21|\mathcal{V}| & (118)\\ &\geq \overline{C}_{\text{Gaussian}} - 22|\mathcal{V}| & (119) \end{aligned}$$

Then from equations (116) and (119) we have

$$\overline{C}_{\text{Gaussian}} - 22|\mathcal{V}| \leq C_{\text{Truncated}} \leq \overline{C}_{\text{Gaussian}} + 21|\mathcal{V}| \tag{120}$$

Also from main Theorem 4.5 we know that

$$\overline{C}_{\text{Gaussian}} - |\mathcal{V}|(15 + \log|\mathcal{V}|) \leq C_{\text{Gaussian}} \leq \overline{C}_{\text{Gaussian}} \tag{121}$$

Therefore

$$|C_{\text{Gaussian}} - C_{\text{Truncated}}| \leq |\mathcal{V}|(36 + \log|\mathcal{V}|) \tag{122}$$

■

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 that 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 exist a scheme such that for any channel gain realizations, still the source can communicate to the destination at rate R, without the knowledge of the channel realizations at the source, the relays and the destination.

Theorem 8.1: Given a compound Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C_{cn} satisfies

$$\overline{C}_{cn} - \kappa \leq C_{cn} \leq \overline{C}_{cn} \tag{123}$$

Where $\overline{C}_{cn}$ is the cut-set upper bound on the compound capacity of $\mathcal{G}$ as described below

$$\overline{C}_{cn} = \max_{p(\{\mathbf{X}_i\}_{j\in\mathcal{V}})} \inf_{h\in\mathcal{H}} \min_{\Omega\in\Lambda_D} I(\mathbf{Y}_{\Omega^c}; \mathbf{X}_\Omega | \mathbf{X}_{\Omega^c}) \tag{124}$$

And κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(16 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$, 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 main 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. Now 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 $\frac{1}{\mathsf{SNR}}$ level so that overall with signal it is at noise level. 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 1 at each relay node (or reducing all channel SNR's of the links 3dB), 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 $|\mathcal{V}|(16 + \log |\mathcal{V}|)$.

Furthermore, as shown in [18] there exists a universal decoder for this finite group 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

$$R < \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \tag{125}$$

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

$$\overline{C}_{cn} - |\mathcal{V}| \leq \max_{p(\{x_i\}_{j\in\mathcal{V}})} \inf_{\hat{h}\in\hat{\mathcal{H}}} \min_{\Omega\in\Lambda_D} I(Y_{\Omega^c}; X_\Omega | X_{\Omega^c}) \tag{126}$$

Therefore from equations (125) and (126) all rates up to $\overline{C}_{cn} - |\mathcal{V}|(16 + \log|\mathcal{V}|)$ are achievable and the proof can be completed.

Now by using the ideas in [19] and [20], 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 is operating at a different frequency band[7], this extension is a just straight forward corollary of the case that nodes have multiple antennas.

Theorem 8.2: Given a frequency selective Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with F different frequency bands. The capacity of this network, C, satisfies

$$\overline{C} - \kappa \leq C \leq \overline{C} \tag{127}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (27), and κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$, where M_i and N_i are respectively the number of transmit and receive antennas at node i.

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

C. Half duplex relay network (fixed transmission scheduling)

One of the practical constraints on wireless networks is that the transceivers can not 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. In this section we study the capacity of wireless relay networks under the half-duplex constraint. The model that we use to study this problem is the same as [21]. 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[8]. 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 portion 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 [21]:

$$C_{hd} \leq \overline{C}_{hd} = \max_{\substack{p(\{x_j^m\}_{j \in \mathcal{V}, m \in \{1,\ldots,M\}}) \\ t_m:\ 0 \leq t_m \leq 1,\ \sum_{m=1}^{M} t_m = 1}} \min_{\Omega \in \Lambda_D} \sum_{m=1}^{M} t_m I(Y_{\Omega^c}^m; X_\Omega^m | X_{\Omega^c}^m) \tag{128}$$

Theorem 8.3: Given a Gaussian relay network with half-duplex constraint, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity, C_{hd}, satisfies

$$\overline{C}_{hd} - \kappa \leq C_{hd} \leq \overline{C}_{hd} \tag{129}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity of $\mathcal{G}$ as described in equation (128), and κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$, 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 Figure 19.

Now consider the t_i's that maximize $\overline{C}_{hd}$ in (128). 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

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

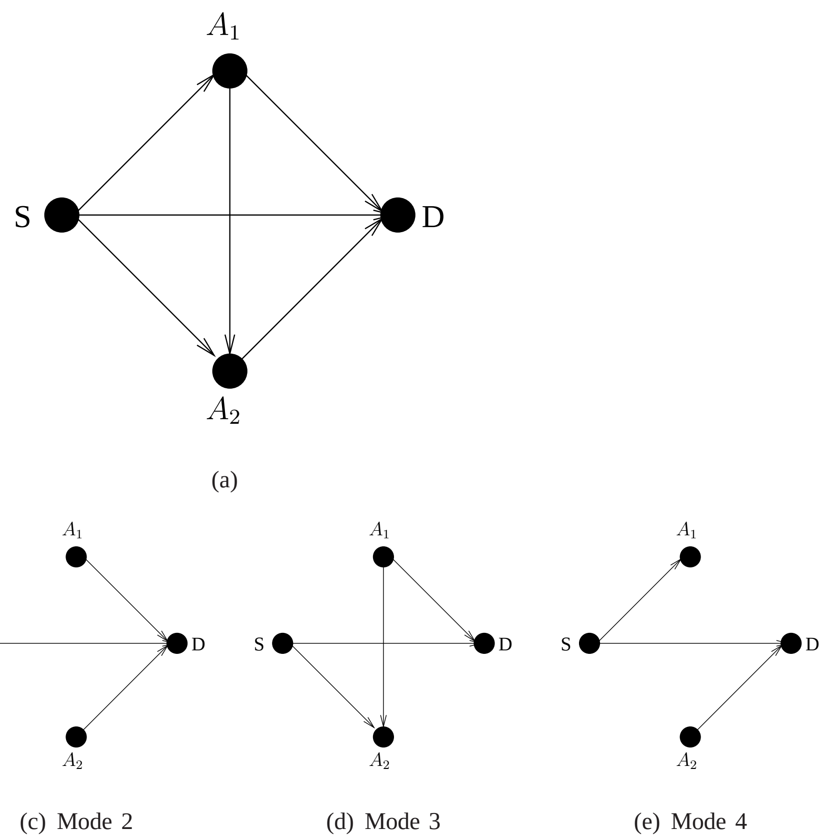


Fig. 19. An example of a relay network with two relays is shown in (a). All four modes of half-duplex operation of the relays are shown in $(b) - (e)$.

allocate Wt_i of bandwidth to mode i, $i = 1, \ldots, M$. Now each mode is running at a different frequency band, therefore as shown in Figure 20 we can combine all these modes and create a frequency selective relay network. Since the links are orthogonal to each other, still the cut-set upper bound on the capacity of this frequency selective relay network (in bits/sec/Hz) is the same as (128). Now by theorem 8.2 we know that our quantize-map-forward scheme achieves, within a constant gap, κ, of $\overline{C}_{hd}$ 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. Now 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. Then in the Wt_i blocks that are allocated to mode i, all relays do exactly what they do in frequency band i. This

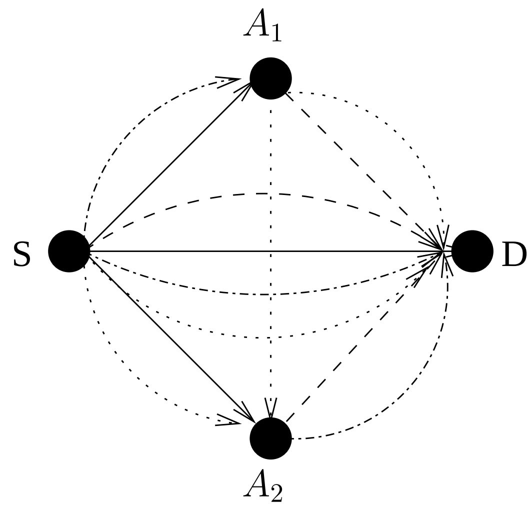


Fig. 20. Combination of all half-duplex modes of the network shown in figure 19. Each mode operates at a different frequency band.

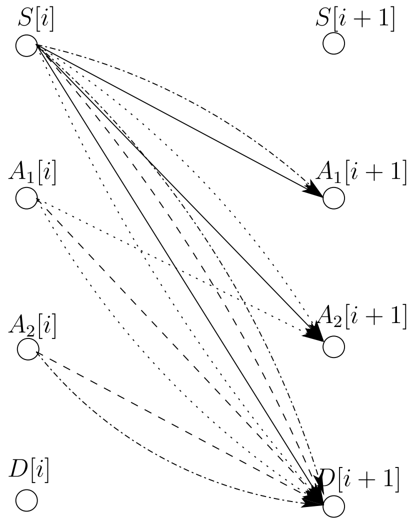
is pictorially described in Figure 21 for the network of Figure 20. This figure shows how one communication block of the frequency selective network (a) is expanded over W blocks of the the original half-duplex network (b). Now since the transmitted signal at each frequency band is only a function of 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.

The main difference between this strategy and our original strategy for full duplex networks is that now the relays are required to have a much larger memory. As a matter of fact, in the full duplex scenario the relays had only memory over one block (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 clearly W can be arbitrary large.

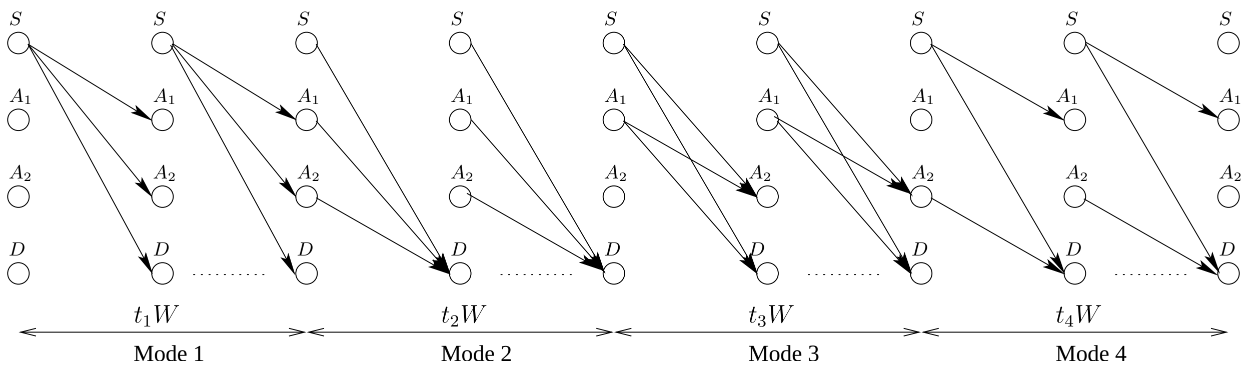
■

D. Quasi-static fading relay network (underspread regime)

In a wireless environment channel gains are not fixed and change over time. In this section we consider a typical scenario in which although the channel gains are changing, they can be considered time invariant over a long time scale (for example during the transmission of a block).



(a) One communication block of the frequency selective half-duplex network



(b) Expanding over W blocks of the original network

Fig. 21. One communication block of the frequency selective network (a) is expanded over W blocks of the original half-duplex network (b).

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 inter-symbol 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: Given a fast fading quasi-static fading Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the ergodic capacity C_{ergodic} satisfies

$$\mathcal{E}_{h_{ij}} \left[\overline{C}(\{h_{ij}\}) \right] - \kappa \leq C_{\text{ergodic}} \leq \mathcal{E}_{h_{ij}} \left[\overline{C}(\{h_{ij}\}) \right] \tag{130}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (27), and the expectation is taken over the channel gain distribution, and κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \right)$, 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. Upper bound is just the cut-set upper bound. For the achievability note that the relaying strategy that we proposed for general wireless 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

$$\frac{1}{L} \sum_{l=1}^{L} \left(\overline{C}(\{h_{ij}\}^l) - \kappa \right) \tag{131}$$

Now as $L \to \infty$,

$$\frac{1}{L} \sum_{l=1}^{L} \overline{C}(\{h_{ij}\}^l) \to \mathcal{E}_{h_{ij}} \left[\overline{C} \right] \tag{132}$$

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,

$$\mathcal{P}_{out}(R) = \mathbb{P} \left\{ C(\{h_{ij}\}) < R \right\} \tag{133}$$

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

$$C_\epsilon = \mathcal{P}_{out}^{-1}(\epsilon) \tag{134}$$

Here is our main result to approximate the outage probability

Theorem 8.5: Given a slow fading quasi-static fading Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the outage probability, $\mathcal{P}_{out}(R)$ satisfies

$$\mathbb{P}\left\{\overline{C}(\{h_{ij}\}) < R\right\} \leq \mathcal{P}_{out}(R) \leq \mathbb{P}\left\{\overline{C}(\{h_{ij}\}) < R + \kappa\right\} \tag{135}$$

Where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (27), and the probability is calculated over the distribution of the channel gains, and κ is a constant and is upper bounded by $\sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i) \left(15 + \log \sum_{i=1}^{|\mathcal{V}|} \max(M_i, N_i)\right)$, 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 $\overline{C}(\{h_{ij}\}) - \kappa < R$ there will not be an outage. ■

E. Low rate capacity approximation of Gaussian relay network

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

Theorem 8.6: Given a Gaussian relay network, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the capacity C satisfies

$$\lambda\overline{C} \leq C \leq \overline{C} \tag{136}$$

where $\overline{C}$ is the cut-set upper bound on the capacity, as described in equation (27), and λ is a constant and is lower bounded by $\frac{1}{2d(d+1)}$ and 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[9] 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

[9] For example see [22] p.153

each of these links is used a total $\frac{1}{2d(d+1)}$ of the time but with d times more power. Now by coding we can convert each links $h_{i,j}$ into a noise free link with capacity

$$c_{i,j} = \frac{1}{2d(d+1)} \log(1 + d|h_{i,j}|^2) \tag{137}$$

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

$$C_{\text{orthogonal}} = \min_{\Omega} \sum_{i,j: i \in \Omega, j \in \Omega^c} c_{i,j} \tag{138}$$

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

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)} \overline{C} \tag{139}$$

To show this assume in the orthogonal network each node transmit the same signal on its outgoing links, and also each node takes the summation of all incoming links (normalized by $\frac{1}{\sqrt{d}}$) and denote it as the received signal. Then the received signal at each node is j is

$$y_j[t] = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \left(h_{ij} \sqrt{d} x_i[t] + z_{ij}[t] \right) \tag{140}$$

$$= \sum_{i=1}^{d} h_{ij} x_i[t] + \tilde{z}_j[t] \tag{141}$$

where

$$\tilde{z}_j[t] = \frac{\sum_{i=1}^{d} z_{ij}[t]}{\sqrt{d}} \sim \mathcal{CN}(0,1) \tag{142}$$

Therefore we get a network which is statically similar to the original non-orthogonal 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

$$C_{\text{orthogonal}} \geq \frac{1}{2d(d+1)} \overline{C} \tag{143}$$

■

IX. CONCLUSIONS

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 foundation to analyze Gaussian models. The main results are a new scheme for general Gaussian relay networks called quantize-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 presented 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-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. In this case we believe that a better upper bound than the cutset bound is needed.

ACKNOWLEDGEMENTS

The research of D. Tse and A. Avestimehr were supported in part by the National Science Foundation through grant CCR-01-18784. The research of S. Diggavi was supported in part by the Swiss National Science Foundation NCCR-MICS center. We would like to thank Anant 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, and R. Yeung.

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[2] S.-Y.R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 43(2):371–81, February 2003.

[3] R. Koetter and M. Médard. An algebraic approach to network coding. *IEEE/ACM Trans. Netw.*, 11(5):782–795, 2003.

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

[5] C. Fragouli and E. Soljanin. Network coding fundamentals. *Foundations and Trends in Networking*, 2(1):1–133, 2007.

[6] T. Ho S. Ray D. Karger R. Koetter M. Effros, M. Medard. Linear network codes: A unified framework for source channel, and network coding. *DIMACS workshop on network information theory*, 2003.

[7] E. C. van der Meulen. Three-terminal communication channels. *Ad. Appl. Pmb.*, 3:120–154, September 1971.

[8] T.M. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Info. Theory*, 25(5):572–584, September 1979.

[9] M. Aref. Information flow in relay networks. *PhD. Thesis, Stanford University*, October 1980.

[10] N. Ratnakar and G. Kramer. The multicast capacity of deterministic relay networks with no interference. *IEEE Transactions on Information Theory*, 52(6):2425–2432, June 2006.

[11] P. Gupta, S. Bhadra, and S. Shakkottai. On network coding for interference networks. *International Symposium on Information Theory (ISIT), Seattle*, July 2006.

[12] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros. Capacity of wireless erasure networks. *IEEE Transactions on Information Theory*, 52(3):789–804, March 2006.

[13] B. Schein. Distributed coordination in network information theory. *Massachusetts Institute of Technology*, 2001.

[14] L. L. Xie and P. R. Kumar. A network information theory for wireless communication: Scaling laws and optimal operation. *IEEE Transactions on Information Theory*, IT-50(5):748–767, May 2004.

[15] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.

[16] G. D. Forney and G. Ungerboeck. Modulation and coding for linear gaussian channels. *IEEE Transactions on Information Theory*, 44(6):2384–2415, October 1998.

[17] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing, 2nd edition, 2006.

[18] M. Feder and A. Lapidoth. Universal decoding for channels with memory. *IEEE Transactions on Information Theory*, 44(9):1726–1745, September 1998.

[19] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *The Annals of Mathematical Statistics*, 30(4):1229–1241, December 1959.

[20] W. L. Root and P. P. Varaiya. Capacity of classes of gaussian channels. *SIAM Journal on Applied Mathematics*, 16(6):1350–1393, November 1968.

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

[22] B. Bollobas. *Modern Graph Theory*. Graduate Texts in Mathematics, vol. 184, Springer, New York, 1998.

[23] A. Orlitsky and J. Roche. Coding for computing. *IEEE Transactions on Information Theory*, 47(3):903–917, March 2001.

APPENDIX I
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 we can think of a decode-forward scheme as described in [8]. Then by using a block-Markov encoding scheme the following communication rate is achievable:

$$R = \min\left(\log\left(1 + |h_{SR}|^2\right), \log\left(1 + |h_{SD}|^2 + |h_{RD}|^2\right)\right) \tag{144}$$

Therefore overall the following rate is always achievable:

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

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

$$C \leq \overline{C} = \max_{|\rho|\leq 1} \min\{\log\left(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)\right), \log\left(1+|h_{SD}|^2+|h_{RD}|^2+2\rho|h_{SD}||h_{RD}|\right)\}$$

Note that if $|h_{SR}| > |h_{SD}|$ then

$$R_{DF} = \min\left(\log\left(1+|h_{SR}|^2\right), \log\left(1+|h_{SD}|^2+|h_{RD}|^2\right)\right) \tag{145}$$

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

$$\log\left(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)\right) \leq \log\left(1+|h_{SR}|^2\right)+1 \tag{146}$$

$$\log\left(1+|h_{SD}|^2+|h_{RD}|^2+2\rho|h_{SD}||h_{RD}|\right) \leq \log\left(1+|h_{SD}|^2+|h_{RD}|^2\right)+1 \tag{147}$$

Hence

$$R_{\text{DF}} \geq \overline{C}_{\text{relay}} - 1 \tag{148}$$

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

$$R_{DF} = \log(1+|h_{SD}|^2) \tag{149}$$

and

$$\log\left(1+(1-\rho^2)(|h_{SD}|^2+|h_{SR}|^2)\right) \leq \log\left(1+|h_{SD}|^2\right)+1 \tag{150}$$

therefore again,

$$R_{\text{DF}} \geq \overline{C}_{\text{relay}} - 1 \tag{151}$$

Appendix II
Proof of Theorem 3.2

The cut-set upper bound on the capacity of diamond network is:

$$\begin{aligned} C_{\text{diamond}} \quad \leq \quad & \overline{C} \leq \min\{\log\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \\ & , \log\left(1+(|h_{A_1D}|+|h_{A_2D}|)^2\right) \\ & , \log(1+|h_{SA_1}|^2)+\log(1+|h_{A_2D}|^2) \\ & , \log(1+|h_{SA_2}|^2)+\log(1+|h_{A_1D}|^2)\} \end{aligned} \tag{152}$$

Without loss of generality assume

$$|h_{SA_1}| \geq |h_{SA_2}| \tag{153}$$

Then we have the following cases:

1) $|h_{SA_1}| \leq |h_{A_1D}|$:

   In this case

$$R_{PDF} \geq \log(1+|h_{SA_1}|^2) \geq \overline{C}-1 \tag{154}$$

2) $|h_{SA_1}| > |h_{A_1D}|$:

   Let $\alpha = \frac{|h_{A_1D}|^2}{|h_{SA_1}|^2}$ then

$$R_{PDF} = \log(1+|h_{A_1D}|^2) + \min\left\{\log\left(1+\frac{(1-\alpha)\,|h_{SA_2}|^2}{\alpha|h_{SA_2}|^2+1}\right), \log\left(1+\frac{|h_{A_2D}|^2}{1+|h_{A_1D}|^2}\right)\right\} \tag{155}$$

   or

$$R_{PDF} = \min\left\{\log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right), \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right)\right\} \tag{156}$$

   Now if

$$\log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right) \geq \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right) \tag{157}$$

   we have

$$\begin{aligned} R_{PDF} &= \log\left(1+|h_{A_1D}|^2+|h_{A_2D}|^2\right) && (158)\\ &\geq \log\left(1+(|h_{A_1D}|+|h_{A_2D}|)^2\right)-1 && (159)\\ &\geq \overline{C}-1 && (160)\end{aligned}$$

   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

$$R_{PDF} = \log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right) \tag{161}$$

   In this case consider two possibilities:

   - $\alpha|h_{SA_2}|^2 \leq 1$:

     In this case we have

$$\begin{aligned} R_{PDF} &= \log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right) && (162)\\ &\geq \log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{2}\right) && (163)\\ &= \log(1+|h_{SA_2}|^2)+\log(1+|h_{A_1D}|^2)-1 && (164)\\ &\geq \overline{C}-1 && (165)\end{aligned}$$

- $\alpha|h_{SA_2}|^2 \geq 1$:

  In this case we are going to show that

$$\begin{aligned} R_{PDF} &= \log\left(\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1}\right) \qquad (166) \\ &\geq \log\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)-1 \qquad (167) \\ &\geq \overline{C}-1 \qquad (168) \end{aligned}$$

  To show this we just need to prove

$$\frac{(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)}{\alpha|h_{SA_2}|^2+1} \geq \frac{1}{2}\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right) \qquad (169)$$

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

$$2|h_{SA_1}|^2(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2) \geq \left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right) \qquad (170)$$

  But note that

$$\begin{aligned} &2|h_{SA_1}|^2(1+|h_{SA_2}|^2)(1+|h_{A_1D}|^2)-\left(1+|h_{SA_1}|^2+|h_{SA_2}|^2\right)\left(|h_{SA_1}|^2+|h_{SA_2}|^2|h_{A_1D}|^2\right) \\ &=|h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2+(|h_{SA_1}|^2|h_{SA_2}|^2-|h_{SA_2}|^4|h_{A_1D}|^2)+ \\ &\quad+(|h_{SA_1}|^2|h_{A_1D}|^2-|h_{SA_2}|^2|h_{A_1D}|^2)+(|h_{SA_1}|^2|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^4) \qquad (171) \\ &=|h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2+|h_{SA_2}|^2(|h_{SA_1}|^2-|h_{SA_2}|^2|h_{A_1D}|^2)+ \\ &\quad+|h_{A_1D}|^2(|h_{SA_1}|^2-|h_{SA_2}|^2)+|h_{SA_1}|^2(|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^2) \qquad (172) \\ &=|h_{SA_1}|^2+|h_{SA_1}|^2|h_{A_1D}|^2+(|h_{SA_1}|^2-|h_{SA_2}|^2)(|h_{SA_2}|^2|h_{A_1D}|^2-|h_{SA_1}|^2+|h_{A_1D}|^2) \qquad (173) \\ &\geq 0 \qquad (174) \end{aligned}$$

  Where the last step is true since

$$\begin{aligned} |h_{SA_1}|^2 &\geq |h_{SA_2}|^2 \qquad (175) \\ |h_{SA_2}|^2|h_{A_1D}|^2 &\geq |h_{SA_2}|^2 \quad (\text{since } \alpha|h_{SA_2}|^2 \geq 1) \qquad (176) \end{aligned}$$

APPENDIX III
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 (33) in Section III-A. We then illustrate the ingredients of the proof using the same example as in Section V-A.2. The complete proof of our result for layered networks is brought in Section III-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 [23]. The notion of joint typicality is naturally extended from Definition 3.1.

Definition 3.1: We define $\underline{x}$ being 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 $f_j : \mathcal{Y}_j^T \to \mathcal{X}_j^T$ its received symbols from the previous block of 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 $\mathcal{Y}_j^T$ *i.e.*, $T_\delta(Y_j)$ onto typical set of transmit sequences *i.e.*, $T_\delta(X_j)$, as

$$\mathbf{x}_j^{(k)} = f_j(\mathbf{y}_j^{(k-1)}), \tag{177}$$

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 (177). Now, given the knowledge of all the encoding functions f_j at the relays and signals received over block $k + l_D$, the decoder $D \in \mathcal{D}$, attempts to decode the message w_k sent by the source.

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-A.2, which was done for the linear deterministic model. Since we are dealing with deterministic networks, the logic up to (45) in Section V-A.2 remains the same. We will again illustrate the ideas using the cut $\Omega = \{S, A_1, B_1\}$. As in Section V-A.2, we can write

$$\begin{aligned} \mathcal{P} &= \mathbb{P}\{\mathcal{A}_2, \mathcal{B}_2, \mathcal{D}, \mathcal{A}_1^c, \mathcal{B}_1^c\} && (178)\\ &= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2, \mathcal{A}_1^c | \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D}, \mathcal{B}_1^c | \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} && (179)\\ &\leq \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{A}_2, \mathcal{B}_2, \mathcal{A}_1^c\} && (180)\\ &= \mathbb{P}\{\mathcal{A}_2\} \times \mathbb{P}\{\mathcal{B}_2 | \mathcal{A}_1^c, \mathcal{A}_2\} \times \mathbb{P}\{\mathcal{D} | \mathcal{B}_1^c, \mathcal{B}_2\} && (181) \end{aligned}$$

where the events $\{\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2, \mathcal{D}\}$ are defined in (46), 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 transmit signals $\mathbf{x}_S(w), \mathbf{x}_S(w') \in T_\delta(X_S)$, hence we can see that

$$\mathbb{P}\{\mathcal{A}_2\} = \mathbb{P}\left\{(\mathbf{x}_S(w'), \mathbf{y}_{A_2}(w)) \in T_\delta(X_S, Y_{A_2})\right\} \doteq 2^{-TI(X_S;Y_{A_2})}, \tag{182}$$

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,

$$\mathbb{P}\{\mathcal{B}_2|\mathcal{A}_1^c,\mathcal{A}_2\} = \mathbb{P}\left\{(\mathbf{x}_{A_1}(w'),\mathbf{x}_{A_2}(w),\mathbf{y}_{B_2}(w)) \in T_\delta(X_{A_1},X_{A_2}Y_{B_2})\right\} \doteq 2^{-TI(X_{A_1};Y_{B_2},X_{A_2})} \overset{(a)}{=} 2^{-TI(X_{A_1};Y_{B_2}|X_{A_2})}, \quad (183)$$

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 of (183) is given by $\frac{|T_\delta(\mathbf{X}_{A_1}|\mathbf{x}_{A_2},\mathbf{y}_{B_2})|}{|T_\delta(\mathbf{X}_{A_1})|}$, which by using properties of (robustly) typical sequences [23] yields the same expression as in (183). Note that the calculation in (183) is similar to one of the error event calculations in a multiple access channel,

Using a similar logic we can write,

$$\mathbb{P}\{\mathcal{D}|\mathcal{B}_1^c,\mathcal{B}_2\} = \mathbb{P}\left\{(\mathbf{x}_{B_1}(w'),\mathbf{x}_{B_2}(w),\mathbf{y}_{D}(w)) \in T_\delta(X_{B_1},X_{B_2}Y_{D})\right\} \doteq 2^{-TI(X_{B_1};Y_{D},X_{B_2})} \overset{(a)}{=} 2^{-TI(X_{B_1};Y_{D}|X_{B_2})}. \quad (184)$$

Therefore, putting (182)–(184) together as done in (54) we get

$$\mathcal{P} \leq 2^{-T\{I(X_S;Y_{A_2})+I(X_{A_1};Y_{B_2}|X_{A_2})+I(X_{B_1};Y_D|X_{B_2})\}}$$

Note that for this example, due to the Markovian structure of the network we can see that[10] $I(Y_{\Omega^c};X_\Omega|X_{\Omega^c}) = I(X_S;Y_{A_2}) + I(X_{A_1};Y_{B_2}|X_{A_2}) + I(X_{B_1};Y_D|X_{B_2})$, hence as in (55) we get that,

$$P_e \leq 2^{RT}|\Lambda_D|2^{-T\min_{\Omega\in\Lambda_D} I(Y_{\Omega^c};X_\Omega|X_{\Omega^c})}, \quad (185)$$

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 III-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 $S \in L_1(\Omega)$),

[10] Note that 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.

- $R_l(\Omega)$: 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 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 $l = 2, \dots, l_D$. Each of these clusters have a transfer function $\mathbf{G}_l(\cdot), l = 1, \dots, l_D$ 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 (177) as

$$\mathbf{x}_j^{(k)}(w) = f_j^{(k)}(\mathbf{y}_j^{(k-1)}(w)). \tag{186}$$

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 Section III-B we can write,

$$\begin{aligned}
\mathcal{P} &= \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1}, l = 2, \ldots, l_D\} && (187)\\
&= \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l, \mathcal{L}_{l-1} | \mathcal{R}_j, \mathcal{L}_{j-1}, j = 2, \ldots, l-1\} && (188)\\
&\leq \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_j, \mathcal{L}_j, j = 2, \ldots, l-1\} && (189)\\
&\overset{(a)}{=} \prod_{l=2}^{l_D} \mathbb{P}\{\mathcal{R}_l | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} && (190)
\end{aligned}$$

Note that as in the example network of Section III-B, 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'. Therefore, all the nodes in R_{l-1} cannot distinguish between w, w' and therefore

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

Hence it is clear that 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 Section III-B, we see that the probability that

$$\begin{aligned}
\mathbb{P}\{\mathcal{R}_l | \mathcal{R}_{l-1}, \mathcal{L}_{l-1}\} &= \mathbb{P}\left\{(\mathbf{x}_{L_{l-1}}(w'), \mathbf{x}_{R_{l-1}}(w), \mathbf{y}_{R_l}(w)) \in T_\delta(X_{L_{l-1}}, X_{R_{l-1}}, Y_{R_l})\right\} && (191)\\
&\doteq 2^{-TI(X_{L_{l-1}}; Y_{R_l} | X_{R_{l-1}})}. && (192)
\end{aligned}$$

Therefore we get

$$\mathcal{P} \leq \prod_{l=2}^{d} 2^{-TI(X_{L_{l-1}}; Y_{R_l} | X_{R_{l-1}})} = 2^{-T \sum_{l=2}^{d} H(Y_{R_l} | X_{R_{l-1}})}. \tag{193}$$

Note that due to the Markovian nature of the layered network, we see that $\sum_{l=2}^{d} H(Y_{R_l} | X_{R_{l-1}}) = H(Y_{\Omega^c} | X_{\Omega^c})$. From this point the proof closely follows the steps from (185) onwards, as in Section III-B. Similarly in multicast scenario we declare an error if *any* receiver $D \in \mathcal{D}$ makes an error, we see that since we have 2^{RT} messages, from the union bound we can drive the error probability to zero if we have,

$$R < \max_{\prod_{i \in \mathcal{V}} p(x_i)} \min_{D \in \mathcal{D}} \min_{\Omega \in \Lambda_D} H(Y_{\Omega^c} | X_{\Omega^c}). \tag{194}$$

APPENDIX IV

PROOF OF LEMMA 6.4

Consider the SVD decomposition of H: $H = U\Sigma V^\dagger$. We have

$$||\Sigma V x||_\infty = ||U^\dagger H x||_\infty \tag{195}$$

$$\leq ||U^\dagger H x||_2 \tag{196}$$

$$\overset{(a)}{=} ||Hx||_2 \tag{197}$$

$$\leq \sqrt{n}||Hx||_\infty \tag{198}$$

where (a) is true since U is unitary.

Therefore, If $||Hx||_\infty \leq \sqrt{2}$, then $||\Sigma V x||_\infty \leq \sqrt{2n}$, which means

$$\mathbb{P}\left\{||Hx||_\infty \leq \sqrt{2}\right\} \leq \mathbb{P}\left\{||\Sigma V x||_\infty \leq \sqrt{2n}\right\} \tag{199}$$

$$= \mathbb{P}\left\{||\Sigma x||_\infty \leq \sqrt{2n}\right\} \tag{200}$$

where the last step is true since x is i.i.d white Gaussian vector, and V is unitary, hence the distribution of x and Vx are the same.

Now by using (200), we get

$$\mathbb{P}\left\{\forall 1 \leq j \leq T : ||H[x_{1,j}, \cdots, c_{m,j}]^t||_\infty \leq \sqrt{2}\right\} \leq \mathbb{P}\left\{\forall 1 \leq j \leq T : ||\Sigma[x_{1,j}, \cdots, c_{m,j}]^t||_\infty \leq \sqrt{2n}\right\} \tag{201}$$

$$\leq \mathbb{P}\left\{\forall 1 \leq i \leq \min(n,m) : ||\sigma_i[x_{i,1}, \cdots, x_{i,T}]||_\infty \leq \sqrt{2n}\right\} \tag{202}$$

$$= \prod_{i=1}^{\min(m,n)} \prod_{j=1}^{T} \mathbb{P}\left\{\sigma_i|x_{i,j}| \leq \sqrt{2n}\right\} \tag{203}$$

$$\leq \prod_{i:\sigma_i \geq 1} \prod_{j=1}^{T} \mathbb{P}\left\{\sigma_i|x_{i,j}| \leq \sqrt{2n}\right\} \tag{204}$$

$$= \prod_{i:\sigma_i \geq 1} \left(1 - e^{-\frac{n}{\sigma^2\sigma_i^2}}\right)^T \tag{205}$$

$$\overset{(e^x \geq 1+x)}{\leq} \prod_{i:\sigma_i \geq 1} \left(\frac{n}{\sigma^2\sigma_i^2}\right)^T \tag{206}$$

$$= 2^{-T\sum_{i:\sigma_i \geq 1}\left(\log(\sigma^2\sigma_i^2) - \log n\right)} \tag{207}$$

$$\leq 2^{-T\left(\log(\prod_{i:\sigma_i \geq 1}\sigma_i^2) - \min(m,n)\log\frac{n}{\sigma^2}\right)} \tag{208}$$

But note that

$$\log(\prod_{i:\sigma_i\geq 1}\sigma_i^2) \geq \log\prod_{i=1}^{\min(m,n)}(1+\sigma_i^2)-\min(m,n) \quad (209)$$

$$= I\left(X;HX+Z\right)-\min(m,n) \quad (210)$$

where X and Z are i.i.d complex normal Gaussian vectors of length m and n respectively.

Therefore

$$\mathbb{P}\left\{\forall 1\leq j\leq T: ||H[x_{1,j},\cdots,c_{m,j}]^t||_\infty\leq\sqrt{2}\right\}\leq 2^{-T\left(I(X;HX+Z)-\min(m,n)(1+\log\frac{n}{\sigma^2})\right)} \quad (211)$$

where X and Z are i.i.d complex normal Gaussian vectors of length m and n respectively.

APPENDIX V
PROOF OF LEMMA 6.5

We first prove the following Lemmas.

Lemma 5.1: Consider integer-valued random variables X, R and S such that

$$X \perp R \quad (212)$$

$$S \in \{-L,\ldots,0,\ldots,L\} \quad (213)$$

$$\mathbb{P}\left\{|R|\geq k\right\} \leq e^{-f(k)}, \quad \text{for all } k\in\mathcal{Z}^+ \quad (214)$$

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

$$Y=X+R+S \quad (215)$$

Then

$$H(Y|X) \leq 2\log_2 e\left(\sum_{k=1}^{\infty}f(k)e^{-f(k)}\right)+\frac{2L+1}{2}+\mathcal{N}_f \quad (216)$$

$$H(X|Y) \leq \log\left(2L+1\right)+2\log_2 e\left(\sum_{k=1}^{\infty}f(k)e^{-f(k)}\right)+\frac{2L+1}{2}+\mathcal{N}_f \quad (217)$$

where

$$\mathcal{N}_f=\left|\{n\in\mathcal{Z}^+|e^{-f(n)}>\frac{1}{2}\}\right| \quad (218)$$

Proof: By definition we have

$$\begin{aligned} H(Y|X) &= H(X+R+S|X) && (219)\\ &= H(R+S|X) && (220)\\ &\leq H(R+S) && (221)\\ &= -\sum_k \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} && (222) \end{aligned}$$

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

$$-\sum_{k=-L}^{L} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} \leq \frac{2L+1}{2} \tag{223}$$

Now note that for $|k| > L$ we have

$$\mathbb{P}\{R+S=k\} \leq \mathbb{P}\{|R| \geq |k|-L\} \leq e^{-f(|k|-L)} \tag{224}$$

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

$$\begin{aligned} -\sum_{k=L+1}^{\infty} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} &= -\sum_{\substack{k>L\\ k-L\in\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} -\\ &\quad -\sum_{\substack{k>L\\ k-L\notin\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} && (225)\\ &\leq \frac{\mathcal{N}_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L)\log e && (226) \end{aligned}$$

and similarly

$$\begin{aligned} -\sum_{k=-\infty}^{-L} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} &= -\sum_{\substack{k<-L\\ |k|-L\in\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} -\\ &\quad -\sum_{\substack{k<-L\\ |k|-L\notin\mathcal{N}_f}} \mathbb{P}\{R+S=k\}\log\mathbb{P}\{R+S=k\} && (227)\\ &\leq \frac{\mathcal{N}_f}{2} + \sum_{k=L+1}^{\infty} e^{-f(k-L)} f(k-L)\log e && (228) \end{aligned}$$

Now by combining (223), (226) and (228) we get

$$H(Y|X) \leq 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \tag{229}$$

Now we prove the second inequality

$$\begin{aligned}
H(X|Y) &= H(X|X+R+S) && (230)\\
&= H(X) - I(X;X+R+S) && (231)\\
&= H(X) - H(X+R+S) + H(X+R+S|X) && (232)\\
&\leq H(X) - H(X+R+S|S) + H(Y|X) && (233)\\
&= H(X) - H(X+R|S) + H(Y|X) && (234)\\
&= H(X) - H(X+R) + I(X+R;S) + H(Y|X) && (235)\\
&\leq H(X) - H(X+R) + H(S) + H(Y|X) && (236)\\
&\leq H(X) - H(X+R) + \log(2L+1) + H(Y|X) && (237)\\
&\leq H(X) - H(X+R|R) + \log(2L+1) + H(Y|X) && (238)\\
&= H(X) - H(X) + \log(2L+1) + H(Y|X) && (239)\\
&= \log(2L+1) + H(Y|X) && (240)
\end{aligned}$$

Therefore

$$H(X|Y) \leq \log(2L+1) + 2\log_2 e\left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \qquad (241)$$

■

Corollary 5.2: Assume v is a continuous complex random variable, then

$$\begin{aligned}
H([v+z]|||[v]) &\leq 14 && (242)\\
H([v]|||[v+z]) &\leq 14 && (243)
\end{aligned}$$

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

Proof: We use lemma 5.1 with variables

$$\begin{aligned}
X &= [\text{Re}(v)] && (244)\\
R &= [\text{Re}(z)] && (245)\\
S &= [\{\text{Re}(v)\} + \{\text{Re}(z)\}] && (246)
\end{aligned}$$

Then $L = 1$ and since

$$\mathbb{P}\{|[\text{Re}(z)]| \geq k\} \leq \mathbb{P}\left\{|[\text{Re}(z)]| - \frac{1}{2} \geq k\right\} \tag{247}$$

$$= 2Q(k - \frac{1}{2}) \tag{248}$$

$$\leq e^{-\frac{(k-\frac{1}{2})^2}{2}} \tag{249}$$

Therefore

$$f(k) = \frac{(k-\frac{1}{2})^2}{2} \tag{250}$$

Also since

$$e^{-\frac{(k-\frac{1}{2})^2}{2}} < \frac{1}{2}, \quad \text{for } k \geq 3 \tag{251}$$

we have

$$\mathcal{N}_f = \{1, 2\} \tag{252}$$

Now we have

$$\log(2L+1) + 2\log_2 e \left(\sum_{k=1}^{\infty} f(k)e^{-f(k)}\right) + \frac{2L+1}{2} + \mathcal{N}_f \tag{253}$$

$$= 2\log_2 e \left(\sum_{k=1}^{\infty} \frac{(k-\frac{1}{2})^2}{2} e^{-\frac{(k-\frac{1}{2})^2}{2}}\right) + 3.5 + \log_2 3 \tag{254}$$

$$\approx 6.89 < 7 \tag{255}$$

As a result

$$H([\text{Re}(v+z)]|||[\text{Re}(v)]) \leq 7 \tag{256}$$

$$H([\text{Re}(v)]|||[\text{Re}(v+z)]) \leq 7 \tag{257}$$

Similarly

$$H([\text{Im}(v+z)]|||[\text{Im}(v)]) \leq 7 \tag{258}$$

$$H([\text{Im}(v)]|||[\text{Im}(v+z)]) \leq 7 \tag{259}$$

Therefore

$$H([v+z]|||[v]) \leq H([\text{Re}(v+z)]|||[\text{Re}(v)]) + H([\text{Im}(v+z)]|||[\text{Im}(v)]) \leq 14 \tag{260}$$

$$H([v]|||[v+z]) \leq H([\text{Re}(v)]|||[\text{Re}(v+z)]) + H([\text{Im}(v)]|||[\text{Im}(v+z)]) \leq 14 \tag{261}$$

■

$$H([\mathbf{Y}_D]|U,F_{\mathcal{V}})] \leq H([\mathbf{Y}_{\mathcal{V}}]|w',F_{\mathcal{V}}) \tag{262}$$

$$= \sum_{l=2}^{l_D} H([\mathbf{Y}_{\mathcal{V}_l}]|[\mathbf{Y}_{\mathcal{V}_{l-1}},F_{\mathcal{V}}]) \tag{263}$$

$$= \sum_{l=2}^{l_D} H([\mathbf{Y}_{\mathcal{V}_l}]|\mathbf{X}_{\mathcal{V}_{l-1}},F_{\mathcal{V}}) \tag{264}$$

$$= \sum_{l=2}^{l_D} H([\text{Re}(\mathbf{Y}_{\mathcal{V}_l})]|\mathbf{X}_{\mathcal{V}_{l-1}},F_{\mathcal{V}}) + H([\text{Im}(\mathbf{Y}_{\mathcal{V}_l})]|\mathbf{X}_{\mathcal{V}_{l-1}},F_{\mathcal{V}}) \tag{265}$$

$$\overset{\text{Corollary 5.2}}{\leq} \sum_{l=2}^{l_D} 14T|\mathcal{V}_l| \tag{266}$$

$$= 14T|\mathcal{V}| \tag{267}$$

APPENDIX VI
PROOF OF LEMMA 6.6

First note that $\overline{C}_\Omega$ is the capacity of the MIMO 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 $m \times n$ MIMO channel ($m \leq n$),

$$Y^n = HX^m + z^n \tag{268}$$

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

$$C = C_{wf} = \sum_{i=1}^{m} \log(1 + \tilde{Q}_{ii}\lambda_i) \tag{269}$$

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

$$\sum_{i=1}^{m} \tilde{Q}_{ii} = mP \tag{270}$$

Now with equal power allocation we have

$$C_{ep} = \sum_{i=1}^{m} \log(1 + P\lambda_i) \tag{271}$$

Now note that

$$C_{wf} - C_{ep} = \log\left(\frac{\prod_{i=1}^{m}(1+\tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{m}(1+P\lambda_i)}\right) \tag{272}$$

$$\leq \log\left(\frac{\prod_{i=1}^{m}(1+\tilde{Q}_{ii}\lambda_i)}{\prod_{i=1}^{m}\max(1,P\lambda_i)}\right) \tag{273}$$

$$= \log\left(\prod_{i=1}^{m}\frac{1+\tilde{Q}_{ii}\lambda_i}{\max(1,P\lambda_i)}\right) \tag{274}$$

$$= \log\left(\prod_{i=1}^{m}\left(\frac{1}{\max(1,P\lambda_i)}+\frac{\tilde{Q}_{ii}\lambda_i}{\max(1,P\lambda_i)}\right)\right) \tag{275}$$

$$\leq \log\left(\prod_{i=1}^{m}\left(1+\frac{\tilde{Q}_{ii}\lambda_i}{P\lambda_i}\right)\right) \tag{276}$$

$$= \log\left(\prod_{i=1}^{m}\left(1+\frac{\tilde{Q}_{ii}}{P}\right)\right) \tag{277}$$

Now note that

$$\sum_{i=1}^{m}(1+\frac{\tilde{Q}_{ii}}{P}) = 2m \tag{278}$$

and therefore by arithmetic mean-geometric mean inequality we have

$$\prod_{i=1}^{m}\left(1+\frac{\tilde{Q}_{ii}}{P}\right) \leq \left(\frac{\sum_{i=1}^{m}(1+\frac{\tilde{Q}_{ii}}{P})}{m}\right)^{m} = 2^m \tag{279}$$

and hence

$$C_{wf} - C_{ep} \leq m \tag{280}$$

Therefore the loss from restricting ourselves to use equal transmit powers at each antenna of an $m \times n$ MIMO channel is at most m bits.

Appendix VII
Proof of Lemma 7.2

To prove this lemma we need to first prove the following two main lemmas:

Lemma 7.1: 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 $P_{\mathbf{X}}$,

$$|I(\mathbf{X};[G\mathbf{X}+Z]) - I(\mathbf{X};[G\mathbf{X}])| \leq 14n \tag{281}$$

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

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 $P_{\mathbf{X}}$,

$$|I(\mathbf{X}; G\mathbf{X}+Z) - I(\mathbf{X}; [G\mathbf{X}+Z])| \leq 7n \tag{282}$$

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

Note that the main Lemma that we want to prove in this section (Lemma 7.2) is just a corollary of these two lemmas. The reason is the following:

$$\left.\begin{array}{r} |I(\mathbf{X}; [G\mathbf{X}+Z]) - I(\mathbf{X}; [G\mathbf{X}])| \leq 14n \\ |I(\mathbf{X}; G\mathbf{X}+Z) - I(\mathbf{X}; [G\mathbf{X}+Z])| \leq 7n \end{array}\right\} \Rightarrow |I(\mathbf{X}; G\mathbf{X}+Z) - I(\mathbf{X}; [G\mathbf{X}])| \leq 21n \tag{283}$$

Therefore we just need to prove Lemma 7.1 and Lemma 7.2. In order to prove Lemma 7.1 we need the following lemma and its corollary.

Now we prove Lemma 7.1.

Proof: **(proof of Lemma 7.1)**
First note that

$$\begin{aligned} I(\mathbf{X}; [G\mathbf{X}]) &\leq I(\mathbf{X}; [G\mathbf{X}+Z]) + I(\mathbf{X}; [G\mathbf{X}] | [G\mathbf{X}+Z]) & (284) \\ &= I(\mathbf{X}; [G\mathbf{X}+Z]) + H([G\mathbf{X}] | [G\mathbf{X}+Z]) & (285) \\ &\leq I(\mathbf{X}; [G\mathbf{X}+Z]) + 14n & (286) \end{aligned}$$

where the last step is true because of Corollary 5.2. Also

$$\begin{aligned} I(\mathbf{X}; [G\mathbf{X}+Z]) &\leq I(\mathbf{X}; [G\mathbf{X}]) + I(\mathbf{X}; [G\mathbf{X}+Z] | [G\mathbf{X}]) & (287) \\ &\leq I(\mathbf{X}; [G\mathbf{X}]) + H([G\mathbf{X}+Z] | [G\mathbf{X}]) & (288) \\ &\leq I(\mathbf{X}; [G\mathbf{X}]) + 14n & (289) \end{aligned}$$

where the last step is true because of Corollary 5.2. Now from equations (286) and (289) we have

$$|I(\mathbf{X}; [G\mathbf{X}+Z]) - I(\mathbf{X}; [G\mathbf{X}])| \leq 14n \tag{290}$$

■

Now we prove Lemma 7.2

Proof: **(proof of Lemma 7.2)**

Define the following random variables:

$$\mathbf{Y} = G\mathbf{X} + Z \tag{291}$$

$$\hat{\mathbf{Y}} = [G\mathbf{X} + Z] \tag{292}$$

$$\tilde{\mathbf{Y}} = \hat{\mathbf{Y}} + \mathbf{U} \tag{293}$$

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 Z.

Now by data processing inequality we have

$$I(\mathbf{X};\mathbf{Y}) \geq I(\mathbf{X};\hat{\mathbf{Y}}) \geq I(\mathbf{X};\tilde{\mathbf{Y}}) \tag{294}$$

Now note that,

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\tilde{\mathbf{Y}}) &= h(\mathbf{Y}) - h(\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - h(\mathbf{Y}|\mathbf{X}) && (295)\\
&= h(\mathbf{Y}) - h(\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - n\log(\pi e) && (296)\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) - h(\tilde{\mathbf{Y}}|\mathbf{Y}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - n\log(\pi e) && (297)\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) - h(\mathbf{U}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - n\log(\pi e) && (298)\\
&= h(\mathbf{Y}|\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - n\log(\pi e) && (299)
\end{aligned}$$

where the last step is true since $h(\mathbf{U}) = nh(U_1) = 2n\log 1 = 0$. Now note that

$$|\text{Re}(y) - \text{Re}(\tilde{y})| \leq \max\left(|[\text{Re}(x+z)] - \text{Re}(x)|\right) + \max|\text{Re}(u)| = \frac{3}{2} \tag{300}$$

and similarly

$$|\text{Im}(y) - \text{Im}(\tilde{y})| \leq \max\left(|[\text{Im}(x+z)] - \text{Im}(x)|\right) + \max|\text{Im}(u)| = \frac{3}{2} \tag{301}$$

Therefore

$$\begin{aligned}
h(\mathbf{Y}|\tilde{\mathbf{Y}}) &= h(\mathbf{Y} - \tilde{\mathbf{Y}}|\tilde{\mathbf{Y}}) && (302)\\
&\leq n\log\left(2\pi e\sqrt{\max\left(|\text{Re}(y) - \text{Re}(\tilde{y})|\right)\max\left(|\text{Im}(y) - \text{Im}(\tilde{y})|\right)}\right) && (303)\\
&= n\log 3\pi e && (304)
\end{aligned}$$

For the second term, we have lets look at the i-th element of $\tilde{y}$

$$\begin{aligned}\tilde{Y}_i &= [\mathbf{g}_i\mathbf{X} + Z_i] + U_i && (305)\\ &= \mathbf{g}_i\mathbf{X} + Z_i + \delta(\mathbf{g}_i\mathbf{X} + Z_i) + U_i && (306)\end{aligned}$$

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

$$\text{Var}\left[\text{Re}(\tilde{Y}_i)|\mathbf{X}\right] = \text{Var}\left[\text{Re}(Z_i) + \text{Re}(\delta(\mathbf{g}_i\mathbf{X} + Z_i)) + \text{Re}(U_i)\right] \quad (307)$$

$$\leq \text{Var}\left[\text{Re}(Z_i)\right] + \text{Var}\left[\text{Re}(\delta(\mathbf{g}_I\mathbf{X} + Z_i))|\mathbf{X}\right] + 2\text{Cov}\left[\text{Re}(Z_i), \text{Re}(\delta(\mathbf{g}_i\mathbf{X} + Z_i))|\mathbf{X}\right] + \text{Var}\left[\text{Re}(U)\right] \quad (308)$$

$$\leq \text{Var}\left[\text{Re}(Z_i)\right] + |\max \text{Re}(\delta(.))|^2 + 2\sqrt{\text{Var}\left[\text{Re}(Z_i)\right] \times |\max \text{Re}(\delta(.))|} + \text{Var}\left[\text{Re}(U_i)\right]) \quad (309)$$

$$= \frac{1}{2} + \frac{1}{4} + 1 + \frac{1}{12} = \frac{11}{6} \quad (310)$$

Similarly

$$\text{Var}\left[\text{Im}(\tilde{Y}_i)|\mathbf{X}\right] \leq \frac{11}{6} \quad (311)$$

Therefore

$$\begin{aligned}h(\tilde{\mathbf{Y}}|\mathbf{X}) &\leq \sum_{i=1}^{n} h(\tilde{Y}_i|\mathbf{X}) && (312)\\ &\leq \sum_{i=1}^{n} \log 2\pi e\sqrt{|K_{\tilde{Y}_i|X}|} && (313)\\ &\overset{(310)}{\leq} n \log \frac{11}{3}\pi e && (314)\end{aligned}$$

Now from equation (299), (304) and (314) we have

$$\begin{aligned}I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\tilde{\mathbf{Y}}) &\leq h(\mathbf{Y}|\tilde{\mathbf{Y}}) + h(\tilde{\mathbf{Y}}|\mathbf{X}) - \frac{n}{2}\log(2\pi e) && (315)\\ &\leq n \log 11\pi e \approx 6.55n < 7n && (316)\end{aligned}$$

■