Towards an Information Theory of Large Networks: An Achievable Rate Region

Piyush Gupta, *Member, IEEE,* and P. R. Kumar, *Fellow, IEEE*

***Abstract*— We study communication networks of arbitrary size and topology and communicating over a general vector discrete memoryless channel. We propose an information-theoretic constructive scheme for obtaining an achievable rate region in such networks. Many well-known capacity-defining achievable rate regions can be derived as special cases of the proposed scheme. A few such examples are the physically degraded and reversely-degraded relay channels, the Gaussian multiple-access channel, and the Gaussian broadcast channel. The proposed scheme also leads to inner bounds for the multicast and allcast capacities.**

Applying the proposed scheme to a specific wireless network of n nodes located in a region of unit area, we show that a *transport capacity* of $\Theta(n)$ bit-meters/sec is feasible in a certain family of networks, as compared to the best possible transport capacity of $\Theta(\sqrt{n})$ bit-meters/sec in [16] where the receiver capabilities were limited. Even though the improvement is shown for a specific class of networks, a clear implication is that designing and employing more sophisticated multi-user coding schemes can provide sizable gains in at least some large wireless networks.

***Index Terms*— Discrete memoryless channels, Gaussian channels, multiuser communications, network information theory, relay networks, transport capacity, wireless networks.**

I. INTRODUCTION

The last few decades have seen a tremendous growth in wireless communication. The most popular examples are cellular voice and data networks and satellite communication systems. These and other similar applications have motivated researchers to extend Shannon's information theory for a single-user channel to some that involve communication among multiple users. A few such examples are the multiple-access channel, the broadcast channel, and the interference channel. The exact capacity region is, however, known in the most general case only for the multiple-access channel, while the broadcast capacity region is known only for few specific channels, like the additive white Gaussian noise channel and the deterministic channel [7], and even fewer results are available for the interference channel [23]. It should be further noted that the above applications as well as the channel models used for analyzing them involve mainly single-hop wireless communication.

Lately, there has been considerable interest in another class of wireless networks, namely, *ad hoc* wireless networks or multi-hop wireless networks. These networks consist of a group of nodes that communicate with each other over a wireless channel without any centralized control. Examples of such networks are in coordinating an emergency rescue operation, networking mobile users of portable-yet-powerful computing devices (laptops, PDAs, smart-phones) on a campus using, for instance, IEEE 802.11 wireless LAN technology [20], sensor networks [3], automated transportation systems [10], *Bluetooth* [17], and *HomeRF* [24]. Ad hoc wireless networks differ from conventional cellular networks in that all links are wireless and there is no centralized control. As every node may not be in direct communication range of every other node, nodes in ad hoc networks cooperate in routing each other's data packets. Lack of centralized control and possible node mobility give rise to a number of challenging design and performance evaluation issues in such wireless networks, many of which have no counterparts in the cellular networks or in the wired networks like the Internet.

An important performance analysis issue is to determine what the traffic-carrying capacity of such multi-hop wireless networks is. An attempt to address this issue was made in [16] under certain models of communications motivated by current technology. For this, an important notion of the *transport capacity* of a network was introduced. Consider a network of n nodes. Suppose the network has U source-destination pairs, communicating independent information over the network. Let $(R_1, R_2, \ldots, R_U)$ be a vector of feasible rates for the source-destination pairs and r_u be the distance between the u-th source and its destination. Then, $\sup \sum_{u=1}^{U} R_u \cdot r_u$, where the supremum is taken over all feasible rate vectors, is defined as the *transport capacity* of the network. In other words, the transport capacity of a network is the maximum bit-distance product that can be transported by the entire network per second and is measured in bit-meters/sec, where a bit transported over a distance of 1 meter toward its destination is counted as 1 bit-meter. It was shown in [16] that under some models of noninterference motivated by current technology, the transport capacity of a network of n nodes located in a region of unit area is $O(\sqrt{n})$ bit-meters/sec, even assuming the node locations, traffic patterns, and the range/power of each transmission, are all optimally chosen. (A specific network construction that achieves a transport capacity of $\Theta(\sqrt{n})$ bit-

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Washington, D.C., June 2001, and at the IMA Workshop on Wireless Networks, Minneapolis, MN, August 2001. The paper is based upon work partially supported by the U.S. Army Research Office under Contracts DAAD19-00-1-0466 and DAAD19-01010-465, DARPA under contracts N00014-01-1-0576 and F33615-01-C-1905, and the Office of Naval Research under Contract N00014-99-1-0696. Any opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the above agencies.

Piyush Gupta is with Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, USA. Email: pgupta@research.bell-labs.com

P. R. Kumar is with the University of Illinois at Urbana-Champaign, Department of Electrical and Computer Engineering, and Coordinated Science Laboratory, 1308 West Main Street, Urbana, IL 61801, USA. Email: prkumar@uiuc.edu

meters/sec was also given.) If the network's transport capacity were to be equitably divided among all n nodes, then each node could only obtain $O(\frac{1}{\sqrt{n}})$ bit-meters/sec. An implication of this scaling law is that an ad hoc wireless network furnishes an average throughput to each user for non-vanishingly far away destinations that diminishes to zero as the number of nodes increases in the network. This suggests that only small ad hoc networks or networks supporting mainly nearest neighbor communications are feasible with current technology.

A natural question is whether the above constriction on the average throughput provided to each node with the size of the network is a limitation of current technology, or whether it can be overcome by employing more sophisticated multi-user coding schemes. In other words, what are the fundamental information-theoretic limits on the traffic-carrying capacity of multi-hop wireless networks? This paper aims at taking a step in addressing this challenging question. As mentioned earlier, the exact information-theoretic capacity region is not known for even simple networks like a general broadcast channel or an interference channel. Thus, here we will only attempt to obtain an achievable rate region in general networks. Even though the obtained rate region will not be capacity-defining in general, it will nevertheless include many well-known capacity-achieving rate regions as special cases. A few such examples are the physically degraded and reversely-degraded relay channels, the Gaussian multiple access channel, and the Gaussian broadcast channel. Furthermore, the proposed scheme will achieve a transport capacity of $\Theta(n)$ bit-meters/sec in a specific wireless network of n nodes located in a unit-area region, which will be motivated by the multiple-input multiple-output architecture.

Consider a set of nodes $\mathcal{N} = \{1, 2, \ldots, n\} =: [1, n]$. Let $X_{j;i}$ denote the signal transmitted by node j in the i-th transmission, and $Y_{j;i}$ be the signal received by node j in the i-th transmission. Let the range space of $X_{j;i}$ be $\mathcal{X}_j$, and the range space of $Y_{j;i}$ be $\mathcal{Y}_j$. Nodes in $\mathcal{N}$ communicate with each other over a general n-user vector discrete memoryless channel (V-DMC) that is specified by $p((y_1, y_2, \ldots, y_n) \,|\, (x_1, x_2, \ldots, x_n))$, the *joint* probability that $\{Y_{j;i} = y_j \in \mathcal{Y}_j : j \in [1, n]\}$ is received in the i-th transmission given that $\{X_{j;i} = x_j \in \mathcal{X}_j : j \in [1, n]\}$ was transmitted.

Suppose $\mathcal{N}$ has U source-destination pairs $\{(s_u, d_u) : u \in [1, U]\}$. We propose a constructive scheme to determine an information-theoretic achievable rate region $\{(R(1), R(2), \cdots, R(U))\}$ in $\mathcal{N}$. As a communication strategy, consider a *feedforward flow graph* $\mathcal{G}(s_u, d_u)$ for each s-d pair $(s_u, d_u), u \in [1, U]$, in $\mathcal{N}$. (One can subsequently optimize over the choice of all such feedforward flow graphs). In each such flow graph $\mathcal{G}(s, d)$, nodes in $\mathcal{N}$ are grouped into disjoint level sets $L_0, L_1, \ldots, L_M, L_{M+1}$, as follows: $L_0 = \{s\}$, $L_{M+1} = \{d\}$, and, for $m \in [1, M]$, nodes in L_m will receive information only from nodes in levels L_{m-k} for $k \in [1, m]$, while they will send information only to nodes in levels L_{m+k} for $k \in [1, M+1-m]$. Thus $\mathcal{G}(s, d)$ will only have arcs of the form (L_m, L_{m+k}), for $k > 0$, $m \in [0, M]$. Finally, let $L_F := \mathcal{N} \setminus \{\cup_{m=0}^{M+1} L_m\}$.

To illustrate the main idea, let us first consider the case of $U = 1$, i.e., when $\mathcal{N}$ has a single source-destination pair (s, d). Let $\mathcal{G}(s, d)$ be a flow graph for (s, d) as above. Let $X_{m,j}$ denote the signal transmitted by $L_{m,j}$, j-th node in level L_m, in one usage of the channel, and $X_m := (X_{m,1}, X_{m,2}, \ldots, X_{m,n_m})$, where n_m denotes the number of nodes in level L_m. Also, let $Y_{m,j}$ denote the signal received by $L_{m,j}$.

Our main result is the following.

Main Result: Rate R_0 is achievable for source-destination pair (s, d) in a network $\mathcal{N}$ of nodes communicating over a V-DMC whenever there exists a flow graph $\mathcal{G}(s, d)$, $x_F \in \mathcal{X}_F$, and a joint probability distribution $p(X_0, X_1, \ldots, X_M \,|\, X_F = x_F)$ corresponding to $\mathcal{G}(s, d)$, such that, for some $\{(R_1, R_2, \ldots, R_M) : R_0 \geq R_1 \geq \cdots \geq R_M \geq 0\}$, R_0 satisfies

$$\begin{aligned} R_M &< I\big(X_M; Y_{M+1,1} \big| x_F\big), \\ R_m &< \min \Big\{ \min_{j \in [1, n_{m+1}]} I\big(X_m; Y_{m+1,j} \big| X_{[m+1,M]}, x_F\big), \\ &\qquad R_{m+1} + \min_{\substack{k \in [2, M+1-m] \\ j \in [1, n_{m+k}]}} I\big(X_m; Y_{m+k,j} \big| X_{[m+1,M]}, x_F\big) \Big\}, \end{aligned} \tag{1}$$

for $m = M-1, M-2, \ldots, 0$. Above $X_{[m+1,M]} := (X_{m+1}, X_{m+2}, \ldots, X_M)$ and $I(X; Y|Z)$ is the *conditional mutual information* between X and Y given Z.

Later in Theorem 3.1 we will apply the above result to obtain achievable rates for the additive white Gaussian noise channels with and without fading. We will also extend the procedure in Section IV to obtain an achievable rate region for a network with multiple source-destination pairs communicating over the Gaussian channels.

The basic idea behind our main result, which will be made precise in the sequel, is as follows. As nodes in L_F cannot relay any information from s to d, they constantly transmit some symbol x_F that facilitates transmissions by nodes in L_m, $m \in [0, M]$. Now if in each level L_{m+1+k}, $k \in [1, M-m]$, every node $L_{m+1+k,j}$, $j \in [1, n_{m+1+k}]$, can receive information at rate R_{m+1} from the component of its received signal $Y_{m+1+k,j}$ that is correlated with $\{X_{m+1}, X_{m+2}, \ldots, X_{m+k}\}$, the signal vectors transmitted by the nodes in levels $\{L_{m+1}, L_{m+2}, \ldots, L_{m+k}\}$, then level L_m will be able to simultaneously send common information at rate R_m to every node in each L_{m+k}, $k \in [1, M+1-m]$, whenever the following are true:

- In level L_{m+1}, every node $L_{m+1,j}$, $j \in [1, n_{m+1}]$, can receive information at rate R_m from the component of its received signal $Y_{m+1,j}$ that is correlated with X_m, the signal vector sent by the nodes in L_m (first term within minimum in (1)).
- In each level L_{m+k}, $k \in [2, M+1-m]$, every node $L_{m+k,j}$, $j \in [1, n_{m+k}]$, can obtain additional information at rate $(R_m - R_{m+1})$ from the component of its received signal $Y_{m+k,j}$ that is correlated with X_m (minimum in the second term in (1)).

It should be noted that in the above scheme the information sent by source s at rate R_0 reaches not only destination d but also all the nodes in the intermediate levels

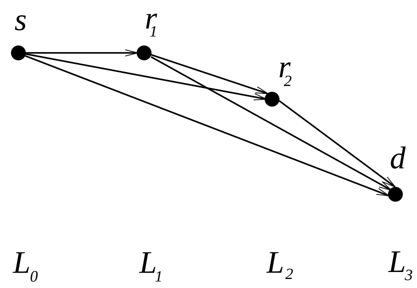

Fig. 1. A 2-level relay channel.

$\{L_1, L_2, \ldots, L_M\}$. Hence the scheme also leads to an inner bound on the *multicast capacity* from source s to a set of destination nodes $\mathcal{A}$, for any $\mathcal{A} \subseteq \cup_{k=1}^{M+1} L_k$.

We next illustrate the main result on a *2-level relay* channel.

Example: 2-level Relay Channel

Consider a network $\mathcal{N}$ of four nodes: source s, destination d, and two relays, r_1 and r_2. Consider a flow graph $\mathcal{G}(s,d)$ for $\mathcal{N}$ such that $L_0 = \{s\}$, $L_1 = \{r_1\}$, $L_2 = \{r_2\}$, and $L_3 = \{d\}$; see Fig. 1. Let X_m and Y_m denote respectively the signal transmitted and the signal received by the node in L_m, $m = 0,1,2,3$. Then a rate R_0 is achievable from s to d with vanishingly small probability of error whenever there exist rates R_1, R_2 and probability distribution $p(x_0, x_1, x_2)$ such that the following hold (which again will be made precise in the sequel).

1) Relay r_1 can receive information at rate R_0 from the transmission of source s. This is achievable if
$$R_0 \;<\; I(X_0; Y_1 | X_1, X_2). \qquad (2)$$
2) Relay r_2 can receive information at rate R_1 from relay r_1, and an additional information at rate $(R_0 - R_1)$ from the transmission of source s. To achieve this, r_2 employs *successive cancellation* decoding. Thus, it first decodes information at rate R_1 from the signal component of its received signal Y_2 that is correlated to the signal transmitted by relay r_1, X_1, treating the uncorrelated component of the transmission of s as noise. It then cancels from its received signal Y_2 the component correlated to X_1 and decodes additional information at rate $(R_0 - R_1)$ from the remaining component of Y_2 that is correlated with the signal transmitted by s, X_0. This is feasible if
$$\begin{aligned} R_1 &< I(X_1; Y_2 | X_2) \\ R_0 &< I(X_0; Y_2 | X_1, X_2) + R_1. \end{aligned} \qquad (3)$$
3) Destination d can receive information at rate R_2 from relay r_2, and additional information at rate $(R_1 - R_2)$ from relay r_1 and at rate $(R_0 - R_1)$ from the transmission of source s. Again employing successive cancellation decoding, this is achievable if
$$\begin{aligned} R_2 &< I(X_2; Y_3) \\ R_1 &< I(X_1; Y_3 | X_2) + R_2 \\ R_0 &< I(X_0; Y_3 | X_1, X_2) + R_1. \end{aligned} \qquad (4)$$

For each $m = 0,1,2$, collecting the bounds on rate R_m in (2-4), we obtain the expression given by (1) for the 2-level relay channel. ■

We next identify some special cases of our general model that have been considered in the literature.

When $\mathcal{N}$ has only three nodes – s, d, and one additional node, r – the resultant network is commonly referred to as a *relay channel*, which was introduced in [22]. The capacities of two special types of relay channels – the *physically degraded* and *reversely degraded* relay channels – were obtained in [8]. The achievable rate given by our result for a relay channel is exactly the maximum of those in Theorems 1 and 2 of [8]. Specifically, two flow graphs are possible for a relay channel:

1) $L_0 = \{s\}$, $L_1 = \{r\}$, and $L_2 = \{d\}$. For a given joint distribution $p(X_s, X_r)$, we have
$$\begin{aligned} R_1 &< I(X_r; Y_d) \\ R_0 &< \min\{I(X_s; Y_r | X_r),\ R_1 + I(X_s; Y_d | X_r)\}. \end{aligned}$$
Hence
$$\begin{aligned} R_0 &< \min\{I(X_s; Y_r | X_r),\ I(X_r; Y_d) + I(X_s; Y_d | X_r)\} \\ &= \min\{I(X_s; Y_r | X_r),\ I(X_s, X_r; Y_d)\}. \end{aligned}$$
2) $L_0 = \{s\}$, $L_1 = \{d\}$, and $L_F = \{r\}$. For $x_r \in \mathcal{X}_r$ and probability distribution $p(X_s | X_r = x_r)$, we have
$$R_0 \;<\; I(X_s; Y_d | X_r = x_r).$$

Choosing the maximum of the suprema of the above two over $p(X_s, X_r)$, and x_r and $p(X_s | X_r = x_r)$, respectively, our main result states that a rate R is achievable in a relay channel, whenever
$$\begin{aligned} R \;<\; \max\Big\{ & \sup_{p(X_s, X_r)} \min\{I(X_s; Y_r | X_r),\ I(X_s, X_r; Y_d)\}, \\ & \max_{x_r \in \mathcal{X}_r} \sup_{p(X_s \mid X_r = x_r)} I(X_s; Y_d \mid X_r = x_r)\Big\}, \end{aligned}$$
which is exactly the maximum of the achievable rates given in Theorems 1 and 2 of [8]. Hence our result is a generalization of the achievable rate for the one-relay channel in Theorems 1 and 2 of [8] to networks having more than one relay node.

A network having four nodes – s, d, and two relays, r_1 and r_2 – with $\mathcal{G}(s,d)$ being $L_1 = \{r_1, r_2\}$, is analyzed in [25] when the network is communicating over the Gaussian channel. In this case, one of the achievable rates obtained there is a special case of our result.

Later in Section IV we will extend our main result to networks having multiple source-destination pairs. We will also then discuss how the capacity-defining achievable rate regions of the Gaussian multiple-access and broadcast channels can be obtained as special cases of our constructive scheme.

The rest of the paper is organized as follows. In Section II we obtain an achievable rate for a source-destination pair in a network of nodes communicating over a general vector discrete memoryless channel. We apply this result in Section III to determine an achievable rate over the additive white Gaussian noise channels with and without fading. In Section IV we extend the procedure given in the previous section to obtain an achievable rate-vector region for networks with multiple source-destination pairs. We apply this result in Section V to arrive at a better achievable bound on the transport capacity of wireless networks than in [16]. In Section VI we discuss a specific wireless network of n nodes located in a region of unit area in which the proposed scheme achieves a transport capacity of $\Theta(n)$ bit-meters/sec.

II. Achievable Rate for a Source-Destination Pair over V-DMC

Consider a set of nodes $\mathcal{N} = \{j : j \in [1, n]\}$ communicating over a general vector discrete memoryless channel (V-DMC) specified by $p((y_1, y_2, \ldots, y_n) \,|\, (x_1, x_2, \ldots, x_n))$, the *joint* probability that node j receives $Y_{j;i} = y_j \in \mathcal{Y}_j$, in i-th transmission, $j \in [1, n]$, given that $\{X_{j;i} = x_j \in \mathcal{X}_j, j \in [1, n]\}$ was transmitted.

Suppose that $\mathcal{N}$ has a single source-destination pair (s, d). In order to specify when a rate R is achievable from source s to destination d, we next provide a few definitions.

Definition 2.1: A (T, b)*-code* for the network with source-destination pair (s, d) consists of the following:

1) An index set of integers $\mathcal{T} = \{1, 2, \ldots, T\} =: [1, T]$.
2) An encoding function $x_s : \mathcal{T} \to \mathcal{X}_s^b$ for the source, yielding codewords $x_s(1), x_s(2), \ldots, x_s(T)$, each of length b. The set of codewords is referred to as codebook $\mathcal{C}_s$.
3) Encoding functions $f_{j;i} : \mathcal{Y}_j^{i-1} \to \mathcal{X}_j$, $j \in [1, n] \setminus \{s, d\}$, $i \in [1, b]$, s.t.
$$X_{j;i} = f_{j;i}(Y_{j;1}, Y_{j;2}, \ldots, Y_{j;i-1}).$$
4) A decoding function $g : \mathcal{Y}_d^b \to \mathcal{T}$.

The above definition, and those to follow, are straightforward generalizations of the ones for the single-relay channel analyzed in [8].

If the message $w \in \mathcal{T}$ is sent, let
$$\lambda(w) = P\big(g(Y_{d;1}, Y_{d;2}, \ldots, Y_{d;b}) \neq w \mid X_s = x_s(w)\big)$$
denote the conditional probability of error in the message decoded by destination d. Then the *maximal probability of error*, $\lambda^{(b)}$, for the (T, b)-code is defined as
$$\lambda^{(b)} = \max_{w \in \mathcal{T}} \lambda(w).$$
The *rate* of a (T, b)-code is $\frac{\log_2 T}{b}$ bits per transmission.

Definition 2.2: A rate R is said to be *achievable* for source-destination pair (s, d) in $\mathcal{N}$ if there exists a sequence of $(T^{(b)}, b)$-codes with $T^{(b)} \geq 2^{bR}$ such that the maximal probability of error $\lambda^{(b)} \to 0$ as $b \to \infty$.

Given the set of nodes $\mathcal{N}$ with source-destination pair (s, d), we construct a *feedforward flow graph* $\mathcal{G}(s, d)$ as follows. As a communication strategy, let us group nodes in $\mathcal{N}$ into disjoint level sets $L_0, L_1, \ldots, L_M, L_{M+1}, L_F$ such that the following hold:

1) $L_0 = \{s\}$, $L_{M+1} = \{d\}$,
2) For $m \in [1, M]$, nodes in L_m will receive information only from nodes in levels L_{m-k} for $k \in [1, m]$, while they will send information only to nodes in levels L_{m+k} for $k \in [1, M+1-m]$. Thus $\mathcal{G}(s, d)$ will only have arcs of the form (L_m, L_{m+k}), for $k > 0$,
3) $L_F := \mathcal{N} \setminus \{\cup_{m=0}^{M+1} L_m\}$.[1]

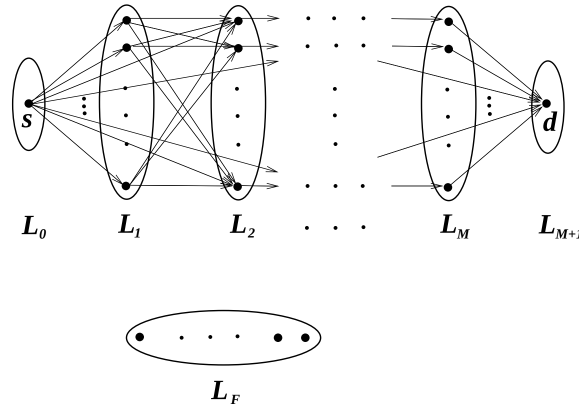

Fig. 2. A feedforward flow graph $\mathcal{G}(s, d)$.

Fig. 2 illustrates such a feedforward flow graph. Let n_m be the number of nodes in L_m, for $m \in [0, M+1]$, and n_F be the number of nodes in L_F. We will use $L_{m,j}$ to indicate the j-th node at level m, according to some arbitrary but fixed numbering of nodes in L_m.

We will denote by $X_{m,j}$ the signal transmitted by node $L_{m,j}$ in one usage of the channel, and by $X_m = (X_{m,1}, X_{m,2}, \ldots, X_{m,n_m}) \in \mathcal{X}_{m,1} \times \mathcal{X}_{m,2} \times \cdots \times \mathcal{X}_{m,n_m}$ the signal vector transmitted by all the nodes in L_m. Similarly, let $X_F \in \mathcal{X}_F := \mathcal{X}_{F,1} \times \mathcal{X}_{F,2} \times \cdots \times \mathcal{X}_{F,n_F}$ denote the signal vector transmitted by the nodes in L_F. Let $X_{[m,m+k]} := (X_m, X_{m+1}, \ldots, X_{m+k})$. Finally, let $Y_{m,j}$ denote the signal received by node $L_{m,j}$.

Then an achievable rate in network $\mathcal{N}$ of arbitrary size and topology and having a single source-destination pair (s, d) is as follows.

Theorem 2.1: A rate R from source s to destination d in network $\mathcal{N}$ of nodes communicating over a V-DMC is achievable whenever
$$R < \max_{\mathcal{G}(s,d)} \max_{x_F \in \mathcal{X}_F} \sup_{p(X_0, X_1, \ldots, X_M \,|\, x_F)} R_0,$$
where, for some $\{(R_1, R_2, \ldots, R_M) : R_0 \geq R_1 \geq \cdots \geq R_M \geq 0\}$, R_0 satisfies
$$\begin{aligned} R_M &< I\big(X_M; Y_{M+1,1} \big| x_F\big), \\ R_m &< \min\Big\{ \min_{j \in [1, n_{m+1}]} I\big(X_m; Y_{m+1,j} \big| X_{[m+1,M]}, x_F\big), \\ &\qquad R_{m+1} + \min_{\substack{k \in [2, M+1-m] \\ j \in [1, n_{m+k}]}} I\big(X_m; Y_{m+k,j} \big| X_{[m+1,M]}, x_F\big) \Big\}, \end{aligned} \tag{5}$$
for $m = M-1, M-2, \ldots, 0$.

Proof: See Appendix II. ∎

[1] Nodes in L_F are those nodes in $\mathcal{N}$ which will not participate in "relaying" information from s to d (but they may "facilitate" the relaying process by constantly transmitting symbols independent of x_s that "open" the channel for other transmissions).

III. Achievable Rate over Gaussian Channels

We next obtain an achievable rate for a single source-destination pair (s,d) in a network $\mathcal{N}$ of n nodes communicating over the additive white Gaussian noise (AWGN) channels with and without fading.

For $l \in [1,n]$, let $X_{l;i}$ be the signal transmitted by node l in the i-th transmission. Then the signal received by node l at the end of the i-th transmission is given by

$$Y_{l;i} \;=\; \sum_{l' \in [1,n]\setminus l} h_{l',l}\, X_{l';i} \,+\, Z_{l;i}\,, \tag{6}$$

where $h_{l',l}$ is the channel gain from node l' to node l and $\mathbf{Z}_l = (Z_{l;1}, Z_{l;2}, \ldots, Z_{l;b})$ is a sequence of i.i.d. Gaussian random variables with mean 0 and variance N_l, for $l \in [1,n]$. An example of the channel gain term $h_{l',l}$ is to model the signal power path loss from node l' to l, i.e.,

$$h_{l',l} \;=\; \frac{1}{r_{l',l}^{\frac{\alpha}{2}}}, \tag{7}$$

where $r_{l',l}$ is the distance between nodes l' and l and α is the signal power path loss exponent. This is the physical model studied, for instance, in [16]. For this model of the channel gain, $X_{l;i}$, $Y_{l;i}$, and $Z_{l;i}$ will all be real-valued random variables.

Here, we will also consider a more general model that, in addition to the signal-power path loss, also incorporates channel variation due to Rayleigh flat fading [6], [28]. Specifically, the channel gain from node l' to node l is given by

$$h_{l',l} \;=\; \frac{\xi_{l',l}}{r_{l',l}^{\frac{\alpha}{2}}}, \tag{8}$$

where $\{\xi_{l',l} \in \mathbb{C} : l', l \in [1,n]\}$ is an i.i.d. complex Gaussian collection with $\xi_{l',l} \sim \mathcal{N}_{\mathbb{C}}(0,1)$.[2] Furthermore, each use of the channel involves independent $\{\xi_{l',l}\}$. Under this channel model, $X_{l;i}$, $Y_{l;i}$, and $Z_{l;i}$ will all be complex-valued random variables, with noise $Z_{l;i} \sim \mathcal{N}_{\mathbb{C}}(0, N_l)$.

As is by now a standard practice [6], we will analyze the Gaussian fading channel under two scenarios concerning the availability of the channel state information at various nodes in the network:

1) *CSIR:* For each $l \in [1,n]$, the channel gains $\mathcal{H}_l := \{h_{l',l} : l' \in [1,n]\}$ will be known to the receiving node l.
2) *CSIRT:* The channel gains $\mathcal{H} := \{h_{l',l} : l', l \in [1,n]\}$ will be known to each node in $\mathcal{N}$.[3]

Let P_l denote the constraint on the average transmission power available to node l, i.e., $X_{l;i}$ needs to satisfy

$$E\left[\,|X_{l;i}|^2\,\right] \;\le\; P_l, \tag{9}$$

where the expectation is also over $\mathcal{H}$ under CSIRT if node l varies its transmit power with the channel state information.

For the above network channel settings, we next obtain an achievable rate from a given source node s to a given destination node d in $\mathcal{N}$. For this, recall the definitions of flow graph $\mathcal{G}(s,d)$ and the associated variables from Section II. In addition, let $h_{(m',j'),(m,j)}$ and $r_{(m',j'),(m,j)}$ denote respectively the channel gain and the distance from node $L_{m',j'}$ to node $L_{m,j}$. Also, for the sake of simplicity in notation, let us relabel the transmission power constraint and the noise power at node $L_{m,j}$ as $P_{m,j}$ and $N_{m,j}$, respectively.

An achievable rate from s to d in $\mathcal{N}$ is obtained as follows. Each node $L_{m,j}$, $m \in [0,M]$, $j \in [1,n_m]$, assigns a fraction $\beta_{m,j,m+k}$ of its power $P_{m,j}$ to aid level L_{m+k} in transmitting its information, for each $k \in [0, M-m]$. This is done in a "coherent" way so that the square roots of the powers received from different nodes in aid of the transmission of the same level add up at each node. Let $P^{(R)}_{m+k,j,m-k'}$ denote the total "coherent" power received by node $L_{m+k,j}$ from the transmissions of level $L_{m-k'}$ as well as those aiding them (made precise in the sequel). Then, the following holds.

Theorem 3.1: i) In network $\mathcal{N}$ of nodes communicating over the AWGN channel without fading given by (6) and (7), rate R is achievable from source s to destination d whenever

$$R \;<\; \max_{\mathcal{G}(s,d)} \sup_{\substack{\{\beta_{m,j,m+k},\, m\in[0,M],\, j\in[1,n_m],\, k\in[0,M-m]:\\ \beta_{m,j,m+k}\in[0,1) \text{ and } \sum_{k=0}^{M-m}\beta_{m,j,m+k}<1\}}} R_0, \tag{10}$$

where, for some $\{(R_1, R_2, \ldots, R_M) : R_0 \ge R_1 \ge \cdots \ge R_M \ge 0\}$, R_0 satisfies

$$\begin{aligned}
R_M &< C\left(\frac{P^{(R)}_{M+1,1,M}}{N_{M+1,1} + \sum_{k'=1}^{M} P^{(R)}_{M+1,1,M-k'}}\right),\\
R_m &< \min\left\{ \min_{j\in[1,n_{m+1}]} C\left(\frac{P^{(R)}_{m+1,j,m}}{N_{m+1,j} + \sum_{k'=1}^{m} P^{(R)}_{m+1,j,m-k'}}\right),\right.\\
&\qquad \left. R_{m+1} + \min_{\substack{k\in[2,M+1-m]\\ j\in[1,n_k]}} C\left(\frac{P^{(R)}_{m+k,j,m}}{N_{m+k,j} + \sum_{k'=1}^{m} P^{(R)}_{m+k,j,m-k'}}\right)\right\},
\end{aligned} \tag{11}$$

for $m = M-1,\, M-2,\, \ldots,\, 0$. Above,

$$P^{(R)}_{m+k,j,m-k'} = \left(\sum_{\substack{k''\in[k',m]\\ j''\in[1,n_{m-k''}]}} \sqrt{\frac{\beta_{m-k'',j'',m-k'} P_{m-k'',j''}}{r^{\alpha}_{(m-k'',j''),(m+k,j)}}}\right)^2, \tag{12}$$

and

$$C(x) \;:=\; \frac{1}{2}\log_2(1+x). \tag{13}$$

ii) For the AWGN channel with fading given by (6) and (8) and the channel state information known only to the receiving nodes (CSIR), an achievable rate R from s to d in $\mathcal{N}$ is also

[2] A random vector $\zeta \in \mathbb{C}^k$ is said to be complex Gaussian if $(\mathcal{R}e(\zeta), \mathcal{I}m(\zeta)) \in \mathbb{R}^{2k}$ is Gaussian. $\vartheta \sim \mathcal{N}_{\mathbb{C}}(\mu, \Sigma)$ denotes that ϑ is a complex circularly-symmetric Gaussian random vector with mean μ and covariance matrix Σ.

[3] Alternatively, one can consider a block-fading channel model where the fading coefficients remain fixed over a block of transmissions and take i.i.d. values from block to block [6], [18], [26]. If the length of the blocks over which the fading coefficients remain constant is sufficiently long, the above scenario holds by learning the channel gains implicitly or explicitly.

given by (10) and (11), except that now

$$P^{(R)}_{m+k,j,m-k'} = \Bigg| \sum_{\substack{k''\in[k',m] \\ j''\in[1,n_{m-k''}]}} \sqrt{\beta_{m-k'',j'',m-k'} P_{m-k'',j''}} \cdot h_{(m-k'',j''),(m+k,j)} \Bigg|^2, \quad (14)$$

and

$$C(x) := E[\log_2(1+x)], \quad (15)$$

where the expectation is taken over the variation in the respective channel gain coefficients.

iii) For the AWGN channel with fading and CSIRT (i.e., the channel state information $\mathcal{H}$ is known to each node in $\mathcal{N}$), a rate R is achievable from s to d whenever there exist a flow graph $\mathcal{G}(s,d)$ and complex transmission gains $\{\gamma_{m,j,m+k}(\mathcal{H}) \in \mathbb{C}, m \in [0,M], j \in [1,n_m], k \in [0,M-m]\}$ such that, for each $m \in [0,M]$, $j \in [1,n_m]$,

$$E_{\mathcal{H}}\left[\sum_{k\in[0,M-m]} |\gamma_{m,j,m+k}(\mathcal{H})|^2\right] < P_{m,j}, \quad (16)$$

and, for some $\{(R_1,R_2,\dots,R_M) : R \geq R_1 \geq \cdots \geq R_M \geq 0\}$, $R_0 = R$ satisfies (11) with

$$P^{(R)}_{m+k,j,m-k'} = \Bigg| \sum_{\substack{k''\in[k',m] \\ j''\in[1,n_{m-k''}]}} \gamma_{m-k'',j'',m-k'}(\mathcal{H}) \cdot h_{(m-k'',j''),(m+k,j)} \Bigg|^2, \quad (17)$$

and

$$C(x) := E_{\mathcal{H}}\left[\log_2(1+x)\right]. \quad (18)$$

Proof: See Appendix III. ■

As stated earlier, in the above scheme, information from source s at rate R_0 not only reaches destination d but also reaches all the nodes in intermediate levels $\{L_1, L_2, \dots, L_M\}$. Hence, the above scheme also leads to an achievable bound on the multicast capacity. Nevertheless, it may sometimes be more profitable to send different information over different paths, or in general over different subnets. In the sequel we will give an approach that will allow splitting the rate of information transfer from s to d over different flow graphs between them. The approach will be a special case of a network with multiple source-destination pairs, discussed next.

IV. Networks with Multiple Source-Destination Pairs

Consider a set of nodes $\mathcal{N} = \{j : j \in [1,n]\}$ communicating over the AWGN channel with or without fading, described in the previous section. Suppose $\mathcal{N}$ has U s-d pairs $\{(s_1,d_1),(s_2,d_2),\dots,(s_U,\, d_U)\}$. As for the single s-d pair case, we next provide a few definitions to specify when a rate vector for the s-d pairs in $\mathcal{N}$ is said to be achievable.

Definition 4.1: A $((T_1,T_2,\dots,T_U),\, b)$*-code* for $\mathcal{N}$ consists of:

1) Index sets of integers $\mathcal{T}_u = [1,T_u]$, $u \in [1,U]$.
2) Encoding functions $f_{j;i} : \left(\prod_{u=1}^{U} (\mathcal{T}_u)^{\delta_j(u)}\right) \times \mathcal{Y}_j^{i-1} \to \mathcal{X}_j$, $j \in [1,n]$, $i \in [1,b]$, where $\delta_j(u) = 1$ if $s_u = j$, 0 otherwise.
3) Decoding functions $g_j : \mathcal{Y}_j^b \to \prod_{u=1}^{U} (\mathcal{T}_u)^{\hat{\delta}_j(u)}$, $j \in [1,n]$, where $\hat{\delta}_j(u) = 1$ if $d_u = j$, 0 otherwise.

Corresponding to each s-d pair (s_u,d_u), $u \in [1,U]$, is a message $W_u \in \mathcal{T}_u$, which is assumed to be independently and uniformly distributed over its range. Then the probability of error that the message W_u is not decoded correctly, is given by

$$P_e^{(b)}(u) = P\Big(\big(g_{d_u}(Y_{d_u;1},Y_{d_u;2},\dots,Y_{d_u;b})\big)_u \neq W_u\Big),$$

and the maximal probability of error for s-d pair (s_u,d_u) is given by

$$\lambda^{(b)}(u) = \max_{w_u\in\mathcal{T}_u} P\Big(\big(g_{d_u}(Y_{d_u;1},\dots,Y_{d_u;b})\big)_u \neq w_u | W_u = w_u\Big).$$

Definition 4.2: A rate vector $(R_1,R_2,\dots,R_U)$ is said to be *achievable* if there exists a sequence of $((T_1^{(b)}, T_2^{(b)}, \dots, T_U^{(b)}),\, b)$-codes with $T_u^{(b)} \geq 2^{bR_u}, u \in [1,U]$, such that $\max_{u\in[1,U]} \lambda^{(b)}(u) \to 0$ as $b \to \infty$.

Next, recall the definition of flow graph $\mathcal{G}(s,d)$ from Section II. For each $u \in [1,U]$, let $\mathcal{G}(s_u,d_u)$ be a flow graph for source-destination pair (s_u,d_u). Suppose that $\mathcal{G}(s_u,d_u)$ has levels $L_0^u, L_1^u, \dots, L_{M_u}^u, L_{M_u+1}^u$. Let $m_j(u) \in [0, M_u+2]$ denote the index of the level to which node j belongs in $\mathcal{G}(s_u,d_u)$, where $m_j(u) = M_u + 2$ denotes that j does not participate in relaying information from s_u to d_u. Let $\mathcal{U}_j$ denote the set of all s-d pairs for which node j does relay information, i.e., $\mathcal{U}_j = \{u \in [1,U] : m_j(u) < M_u+2\}$. Then, an achievable rate-vector region in $\mathcal{N}$ for source-destination pairs $\{(s_1,d_1),\,(s_2,d_2),\dots,(s_U,d_U)\}$ is as follows.

Theorem 4.1: An achievable rate-vector region $\mathcal{R}$ in network $\mathcal{N}$ of nodes communicating over a Gaussian channel and having U source-destination pairs $\{(s_1,d_1),\, (s_2,d_2),\dots,(s_U,d_U)\}$ is given by the closure of the convex hull of the set of all rate vectors $(R_1,R_2,\dots,R_U)$ such that, for some flow graphs $\{\mathcal{G}(s_u,d_u) : u \in [1,U]\}$, intermediate level rates $\{(R_{u,1},R_{u,2},\dots,R_{u,M_u}) : R_{u,0} := R_u \geq R_{u,1} \geq \cdots \geq R_{u,M_u} \geq 0,\, u \in [1,U]\}$, and random variable Q with probability distribution $\{p(q) : q \in \mathcal{Q}\}$, for some finite $|\mathcal{Q}|$, the following constraints are satisfied for each node $j \in [1,n]$:

$$R_{u,k} < E_Q\left[C\left(\frac{P^{(R)(Q)}_{j,u,k}}{N_j + \sum_{\substack{\{(u',k') :\, \sigma_j^{(Q)}(u',k') \\ > \sigma_j^{(Q)}(u,k)\}}} P^{(R)(Q)}_{j,u',k'}}\right)\right],$$

for each $u \in \mathcal{U}_j$ and $k = m_j(u) - 1$, and

$$R_{u,k} < R_{u,k+1} + E_Q\left[C\left(\frac{P^{(R)(Q)}_{j,u,k}}{N_j + \sum_{\substack{\{(u',k') :\, \sigma_j^{(Q)}(u',k') \\ > \sigma_j^{(Q)}(u,k)\}}} P^{(R)(Q)}_{j,u',k'}}\right)\right],$$

for each $u \in \mathcal{U}_j$ and $k \in [0, m_j(u)-2]$. Above, for each $q \in \mathcal{Q}$, $\{\sigma_j^{(q)}(u,k) \in \{0,1,2,\dots\} : k \in [0,M_u],\, u \in \mathcal{U}_j\}$

are some successive decoding orderings satisfying

$$\sigma_j^{(q)}(u, m_j(u)-1) < \sigma_j^{(q)}(u, m_j(u)-2) < \cdots < \sigma_j^{(q)}(u,0),$$

and

$$P_{j,u,k}^{(R)(q)} = \left| \sum_{\substack{j' \neq j \\ j' \in [1,n]}} \gamma_{j',u,k}^{(q)} \, h_{j',j} \right|^2,$$

for some transmission gains $\{\gamma_{j,u,k}^{(q)} : k \in [0, M_u+2], u \in [1,U]\}$ with $E[|\gamma_{j,u,k}^{(q)}|^2] = \beta_{j,u,k}^{(q)} P_j$, for some power allocations $\{\beta_{j,u,k}^{(q)} \in [0,1) : k \in [0, M_u+2], u \in [1,U]\}$ satisfying $\{\beta_{j,u,k}^{(q)} = 0, u \notin \mathcal{U}_j \text{ or } k < m_j(u)\}$ and

$$\sum_{q \in \mathcal{Q}} p(q) \sum_{u \in \mathcal{U}_j} \sum_{k=m_j(u)}^{M_u} \beta_{j,u,k}^{(q)} < 1.$$

Proof: See Appendix IV. ■

It is worth noting that the achievable-rate regions given by Theorem 4.1 for the U-user Gaussian multiple-access channel and the U-user Gaussian broadcast channel are exactly the capacity-defining achievable rate regions given for these channels in [1], [21] and [4], [14], respectively.

We next extend Theorems 3.1 and 4.1 so as to allow different information to be sent over different flow graphs between a source-destination pair (s,d). The basic idea is to colocate virtual source-destination pairs at the same pair of nodes s and d, in the above scheme. We illustrate this with an example. Consider a network with a single source-destination pair (s,d). Suppose that (s,d) sends information at rate R_v over path, or more generally, feedforward flow graph, $\mathcal{G}_v$, $v = 1, 2, \ldots, V$. Then we replace (s,d) by $\{(s_1,d_1),(s_2,d_2),\ldots,(s_V,d_V)\}$, where s_v and d_v are located at s and d, respectively, for each $v \in [1,V]$. Now the rate partition $\{R_v : v \in [1,V]\}$ is achievable for (s,d) if $(R_1, R_2, \ldots, R_V)$ belongs to the rate region $\mathcal{R}$ given by Theorem 4.1 for $\{(s_1,d_1),(s_2,d_2),\ldots,(s_V,d_V)\}$ with flow graphs $\mathcal{G}(s_v,d_v) = \mathcal{G}_v$, $v \in [1,V]$. Generalization of the approach for the case where there are multiple source-destination pairs is straightforward.

A limitation of the above approach is that it does not use *network coding* [2]. That is, since intermediate nodes handle information sent by different virtual source-destination pairs independently, they do not perform coding of bits received over different paths. It is shown in [2] that such coding strictly increases the achievable rate in multicasting, i.e., when the same information is to be sent from a source to multiple destinations. Generalizing the network coding scheme of [2] for a network of independent DMCs to the more general channel model that we consider is a challenging problem.

V. Achievable Bound on Transport Capacity of Wireless Networks

We next apply Theorem 4.1 to obtain an achievable bound on the transport capacity of arbitrary wireless networks (cf. Section I). Even though the achievable bound is not in an explicit form, it nevertheless improves upon that given in [16] for networks with limited receiver capabilities. In fact, in the next section we will construct a specific wireless network of n nodes located in a unit-area region where the proposed scheme will achieve a transport capacity of $\Theta(n)$ bit-meters/sec, as compared to the best possible transport capacity of $\Theta(\sqrt{n})$ bit-meters/sec in [16].

Consider n nodes located in a disk D of unit area. Let ϕ_j be the location of node j in D, for $j \in [1,n]$. For notational convenience, let $\Phi := (\phi_1, \phi_2, \ldots, \phi_n)$ and $r_{i,j} := |\phi_i - \phi_j|$. Also, let $\mathbf{R} = \{R(i,j) : i,j \in [1,n]\}$ denote a rate vector in the network, where $R(i,j)$ denotes the average rate of information transfer (in bits/transmission) from node i to node j.

Suppose that the bandwidth available to the network is W Hz. Also, suppose that $\frac{N_j^{(0)}}{2}$ is the noise spectral density at node j. First consider the case when the channel has no fading. Let $\mathcal{R}(\Phi)$ be the set of all achievable rate vectors $\mathbf{R}$ in Theorem 4.1 for the node-location vector Φ, the power-constraint vector $\{P_j : j \in [1,n]\}$, and the noise-power vector $\{N_j = N_j^{(0)} W : j \in [1,n]\}$. Now fix a rate vector $\mathbf{R} = \{R(i,j) : i,j \in [1,n]\} \in \mathcal{R}(\Phi)$. Employing by-now standard characterization of transforming rates from bits/transmission to bits/second for a band-limited AWGN channel without fading (see, for instance, [9] and references therein), we obtain that $2WR(i,j)$ bits per second can be sent from node i to node j, for each $i,j \in [1,n]$, with vanishingly small probability of error. Thus, the total bit-meters transported by the network per second is $\sum_{i=1}^n \sum_{j=1}^n 2WR(i,j)\, r_{i,j}$. For the AWGN channel with fading, the above continues to hold except that, since inputs and outputs are now all complex, the rate from node i to node j in bits/second is given by $WR(i,j)$. Hence, we obtain the following:

Theorem 5.1: The transport capacity of a wireless network of n nodes communicating over a Gaussian channel of bandwidth W Hz is lower bounded by

$$\sup_{\mathbf{R} \in \mathcal{R}(\Phi)} \sum_{i=1}^n \sum_{j=1}^n \eta W R(i,j)\, r_{i,j} \text{ bit-meters/sec,} \tag{19}$$

where $\mathcal{R}(\Phi)$ is the achievable rate region given by Theorem 4.1 for the node-location vector Φ with the spatial separation between nodes i and j being $r_{i,j}$, the transmit-power constraint vector $\{P_j : j \in [1,n]\}$, and the noise-power vector $\{N_j = N_j^{(0)} W : j \in [1,n]\}$, and η is 1 (resp., 2) when the channel has fading (resp., no fading).

We next construct a specific wireless network of n nodes where a transport capacity of $\Theta(n)$ bit-meters/sec is achieved.

VI. Feasibility of Transport Capacity of $\Theta(n)$ bit-meters/sec

Our constructive example is motivated by the multiple transmit-receive antenna architecture [11], [12], [29], and the basic idea is as follows. Consider a network of n nodes, where $\frac{n}{2}$ source nodes (for notational simplicity, let n be even) are placed "close" to the origin and $\frac{n}{2}$ destination nodes are placed "close" to $(1,0)$; see Fig. 3. Now since the source nodes are close to each other, they can exchange among themselves all the information that they have to send to their respective destinations in a "short" time. Similarly, the $\frac{n}{2}$ destination

Fig. 3. Placement of nodes.

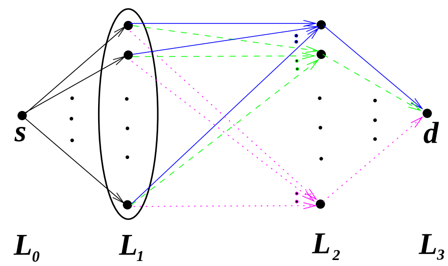

Fig. 4. Flow graphs.

nodes will be able to exchange among themselves all the information that they will receive from the $\frac{n}{2}$ source nodes in a "short" time. Thus, $\frac{n}{2}$ source nodes act as $\frac{n}{2}$ transmit antennas and $\frac{n}{2}$ destination nodes act as $\frac{n}{2}$ receive antennas. Under appropriate conditions on the channel characteristics and specific choices of flow graphs and power allocations, we obtain that $\Theta(n)$ bits/sec can be sent from the source nodes to the destination nodes. Since the source-destination pairs are $\Theta(1)$ meters apart, we obtain that $\Theta(n)$ bit-meters/sec is feasible. We make the construction precise in the following.

Consider a set $\mathcal{N}$ of n nodes communicating over the AWGN channel with fading given by (6) and (8). Here, we consider CSIRT, i.e., the channel gains $\mathcal{H} := \{h_{l',l} : l, l' \in [1, n]\}$ are known to each node in $\mathcal{N}$. Suppose the bandwidth available to the network is W Hz. Further, suppose the average power available to each node in $\mathcal{N}$ for transmission is P and the noise power at each node is N. For the sake of brevity, we discuss here the case where $\mathcal{N}$ has a single source-destination pair (s, d); the extension to the multiple source-destination case is straightforward. The placement of nodes in $\mathcal{N}$ is as given in Fig. 3. Source node s is placed at the origin, while destination node d is placed at $(1, 0)$. Additionally, $(\frac{n}{2} - 1)$ nodes are uniformly placed, together with s, on the circle centered at $(r, 0)$ and of radius r, where

$$r = e^{-\frac{n}{\alpha}}. \tag{20}$$

The remaining $(\frac{n}{2} - 1)$ nodes are similarly placed on the circle centered at $(1-r, 0)$ and of radius r.[4] Now information from source s is sent to destination d over a set of $(\frac{n}{2} - 1)$ flow graphs $\{\mathcal{G}_i\}$, each having $M = 2$ levels. Flow graph $\mathcal{G}_i$ has level L_1 consisting of all $(\frac{n}{2} - 1)$ nodes close to s and level L_2 having only node i from among the nodes close to d, for some arbitrary but fixed ordering of the nodes; see Fig. 4. Let R_i be the rate at which destination d receives information over $\mathcal{G}_i$. Thus, the cumulative rate of information flow from s to d is $R = \sum_{i=1}^{\frac{n}{2}-1} R_i$.

We next apply Theorem 4.1 to obtain the feasibility of $R = \Theta(n)$ bits/transmission. Since s and d are 1 meter apart, Theorem 5.1 would then allow us to show that a transport capacity of $WR = \Theta(Wn)$ bit-meters/sec is feasible. In order to use Theorem 4.1 in the current setup, we treat information flow from s to d over different flow graphs $\mathcal{G}_i$ as information transfer between co-located virtual source-destination pairs (cf. Section IV). With this we next give specific choices of various parameters that will allow the rate-constraint conditions in Theorem 4.1 to hold for $R_i = \Theta(1)$ and thus $R = \Theta(n)$; since Theorem 4.1 actually optimizes over these parameters, the feasible rate it will obtain can only be better. We start by specifying how the transmissions are scheduled in the network. Let Q be a random variable that takes values in $[1, \frac{n}{2}+1]$ with probability distribution: $p(1) = p(2) = \frac{1}{3}$ and $p(i) = \frac{1}{3}\frac{1}{\frac{n}{2}-1}$, for $i = 3, 4, \ldots, \frac{n}{2} + 1$. Now, when $Q = i$, for $i = 1, 2$, only the nodes in level L_{i-1} transmit, while, for $i \geq 3$, only node $L_{2,i-2}$ transmits. Then, the rate-constraint conditions in Theorem 4.1 are satisfied for information flow at rate R from s to d when the following hold:

1) Source s can send information at rate R to nodes in L_1. This is achievable with high probability if

$$R < \min_{j\in[1,\frac{n}{2}-1]} \frac{1}{3} E_{\mathcal{H}} \left[\log_2\left(1 + \frac{P|h_{s,(1,j)}|^2}{N}\right)\right],$$

where $h_{s,(1,j)}$ is the channel gain from source s to the j-th node in level L_1. This is feasible if

$$\begin{aligned} R &< \min_{j\in[1,\frac{n}{2}-1]} \frac{1}{3} E_{\xi} \left[\log_2\left(1 + \frac{P(2r)^{-\alpha}|\xi_{s,(1,j)}|^2}{N}\right)\right] \\ &= \min_{j\in[1,\frac{n}{2}-1]} \Theta(n) \\ &= \Theta(n). \end{aligned} \tag{21}$$

2) Nodes in L_1 can simultaneously send independent information at rate R_i to node i in L_2, for $i \in [1, \frac{n}{2} - 1]$. To achieve this, nodes in L_1 perform *zero-forcing beamforming* [6]. Let H denote the matrix of channel gains from the nodes in L_1 to those in L_2. Then, almost surely H has full rank, as the channel fading coefficients between all pairs of nodes in L_1 and L_2 are statistically independent. Thus, H^{-1} exists almost surely. Now, when $Q = 2$, the nodes in L_1 transmit the signal vector $X_1 = (X_{1,1}, X_{1,2}, \ldots, X_{1,\frac{n}{2}-1})^T$, where $X_1 = 0$ for the zero-probability event that H does not have full rank, while otherwise

$$X_1 = \sqrt{P(\frac{n}{2} - 1)} \frac{H^{-1}\Psi}{||H^{-1}\Psi||},$$

where $\Psi = (\Psi_1, \Psi_2, \ldots, \Psi_{\frac{n}{2}-1})^T \sim \mathcal{N}_{\mathbb{C}}(0, I)$. Then, $E[||X_1||^2 \,|\, Q = 2] = P(\frac{n}{2} - 1)$, and thus $E[||X_1||^2] = \frac{1}{3}P(\frac{n}{2} - 1)$, which, by near symmetry, ensures that the constraint on the average transmission power at the nodes in L_1 is satisfied. Furthermore, when $Q = 2$, the

[4]Note that nodes in $\mathcal{N}$ are located within the unit-area disk centered at $(\frac{1}{2}, 0)$.

signal vector received by the nodes in L_2 is

$$\begin{aligned} Y_2 &= HX_1 + Z_2 \\ &= \frac{\sqrt{P(\frac{n}{2}-1)}}{\|H^{-1}\Psi\|}\Psi + Z_2 \quad \text{a.s.} \end{aligned}$$

Thus, rates $R_i, i \in [1, \frac{n}{2}-1]$, are simultaneously feasible from nodes in L_1 to node i in L_2, respectively, with high probability if

$$\begin{aligned} R_i &< \frac{1}{3} E_{H,\psi}\left[\log_2\left(1 + \frac{\frac{P(\frac{n}{2}-1)}{\|H^{-1}\Psi\|^2}}{N}\right)\right] \\ &= \frac{1}{3} E_{H,\psi}\left[\log_2\left(1 + \frac{P}{N}\frac{1}{\frac{1}{\frac{n}{2}-1}\|H^{-1}\Psi\|^2}\right)\right]. \end{aligned} \quad (22)$$

Now, recall that $\Psi \sim \mathcal{N}_{\mathbb{C}}(0, I)$ and the entries of H are independent complex Gaussian with mean zero and variance close to unity for large n (since r is given by (20), $(1-4r)^{-\alpha} \sim 1$, for large n). Thus, $(\frac{1}{\frac{n}{2}-1}\|H^{-1}\Psi\|^2)^{-1}$ is nearly distributed as $F_{2,\,2(\frac{n}{2}-1)}$, which has mean $(1+\frac{1}{\frac{n}{2}-2})$ and variance $(1-\frac{1}{\frac{n}{2}-1})^{-2}(1-\frac{2}{\frac{n}{2}-1})^{-1}$ (see Theorem 3.3.28 [15] and its generalization to complex vectors, for instance, in [19]). Furthermore, it is easy to show that, for $\eta \sim F_{2,2K}$, $K > 1$, and any $\rho > 0$, $E_\eta[\log_2(1+\rho\eta)] > e^{-1}\log_2(1+\rho)$. Hence, (22) holds if

$$R_i < \frac{1}{3}\left(e^{-1}\log_2\left(1+\frac{P}{N}\right) - o(1)\right).$$

Therefore, rate R is feasible from s to d with high probability if

$$\begin{aligned} R &= \sum_{i=1}^{\frac{n}{2}-1} R_i \\ &< (\frac{n}{2}-1)\frac{1}{3}\left(e^{-1}\log_2\left(1+\frac{P}{N}\right) - o(1)\right) \\ &= \Theta(n). \end{aligned} \quad (23)$$

3) Lastly, destination d should be able to successfully decode all the information that is being sent over flow graphs $\{\mathcal{G}_i\}$. This is achievable with high probability if

$$\begin{aligned} R_i &< \frac{1}{3}\frac{1}{\frac{n}{2}-1} E_\xi\left[\log_2\left(1 + \frac{P(2r)^{-\alpha}\left|\xi_{(2,i),d}\right|^2}{N}\right)\right] \\ &= \Theta(1), \end{aligned}$$

where the last step follows from (20). Thus, rate R is feasible with high probability if

$$R = \sum_{i=1}^{\frac{n}{2}-1} R_i < \Theta(n). \quad (24)$$

Hence, from (21)-(24), we determine that a rate of $\Theta(n)$ bits/transmission is achievable from s to d. Since s and d are 1 meter apart, we obtain from Theorem 5.1 that a transport capacity of $\Theta(Wn)$ bit-meters/sec is feasible.

The above improvement in the transport capacity is for a specific class of networks. Whether such a favorable scaling law for the transport capacity holds for a network with randomly located nodes is a challenging open question.

VII. Conclusions

We have constructed an information-theoretic scheme for obtaining an achievable rate region in a network of arbitrary size and topology. Many well-known capacity-defining achievable rate results can be derived as special cases of our scheme. A few such examples are the physically degraded and reversely-degraded relay channels, the Gaussian multiple access channel, and the Gaussian broadcast channel. The proposed scheme also leads to a better achievable bound on the transport capacity of wireless networks than previously obtained. In fact, the scheme allows a feasible transport capacity of $\Theta(n)$ bit-meters/sec in a specific wireless network of n nodes located in a region of unit area, as compared to the best possible transport capacity of $\Theta(\sqrt{n})$ bit-meters/sec shown earlier [16] for less sophisticated receiver operation. Even though the improvement is shown for a specific class of networks, a clear implication is that one may obtain sizable gains by designing and employing more sophisticated multi-user coding schemes in at least some large wireless networks. Quantifying the gains in a general network is an important subject for future work.

Appendix I
Some Definitions and Background Results

In this appendix, we recall some standard definitions and results from the literature.

Let $\{X_1, X_2, \ldots, X_k\}$ denote a finite collection of discrete random variables with some fixed joint distribution, $p(x_1, x_2, \ldots, x_k)$, $(x_1, x_2, \ldots, x_k) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$. Let S be a subset of $\{X_1, X_2, \ldots, X_k\}$. Consider b independent copies of S, $\{S_1, S_2, \ldots, S_b\}$. Let $\mathbf{S}$ denote the vector $(S_1, S_2, \ldots, S_b)$. Thus, for $\mathbf{s} = (s_1, s_2, \ldots, s_b)$,

$$P(\mathbf{S} = \mathbf{s}) = \prod_{i=1}^{b} P(S_i = s_i).$$

Then, by the law of large numbers,

$$-\frac{1}{b}\log_2 p(\mathbf{S}) = -\frac{1}{b}\sum_{i=1}^{b}\log_2 p(S_i) \to H(S),$$

where the convergence takes place simultaneously with probability 1 for all 2^k subsets, $S \subseteq \{X_1, X_2, \ldots, X_k\}$.

Next recall the definition of jointly ϵ-typical sequences [9]: The set $A_\epsilon^{(b)}(X_1, X_2, \ldots, X_k)$ of ϵ-typical b-sequences $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) \in \mathcal{X}_1^b \times \mathcal{X}_2^b \times \cdots \times \mathcal{X}_k^b$ is defined by

$$A_\epsilon^{(b)}(X_1, X_2, \ldots, X_k) := \Big\{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k) : \left| -\frac{1}{b}\log_2 p(\mathbf{s}) - H(S)\right| < \epsilon, \forall S \subseteq \{X_1, X_2, \ldots, X_k\}\Big\}.$$

Let $A_\epsilon^{(b)}(S)$ denote the restriction of $A_\epsilon^{(b)}(X_1, X_2, \ldots, X_k)$

to the coordinates of S. For example, if $S = (X_1, X_2)$,

$$A_\epsilon^{(b)}(S) = \Big\{(\mathbf{x}_1, \mathbf{x}_2) : \left|-\frac{1}{b}\log_2 p(\mathbf{x}_1) - H(X_1)\right| < \epsilon, \left|-\frac{1}{b}\log_2 p(\mathbf{x}_2) - H(X_2)\right| < \epsilon, \left|-\frac{1}{b}\log_2 p(\mathbf{x}_1, \mathbf{x}_2) - H(X_1, X_2)\right| < \epsilon\Big\}.$$

We recall the following well known fact (see for instance [9]).

Lemma 1.1: For any $\epsilon > 0$, for sufficiently large b,

(i) $P(A_\epsilon^{(b)}(S)) \geq 1 - \epsilon, \quad \forall S \subseteq \{X_1, X_2, \ldots, X_k\}$.

(ii) $\mathbf{s} \in A_\epsilon^{(b)}(S) \Rightarrow |-\frac{1}{b}\log_2 p(\mathbf{s}) - H(S)| < \epsilon$.

(iii) $(1-\epsilon)2^{b(H(S)-\epsilon)} \leq |A_\epsilon^{(b)}(S)| \leq 2^{b(H(S)+\epsilon)}$, where $|A_\epsilon^{(b)}(S)|$ denotes the cardinality of the set $A_\epsilon^{(b)}(S)$.

To bound the probability of error, we will use another well known result on the probability that conditionally independent sequences are jointly typical. Let $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$ be three subsets of $\{X_1, X_2, \ldots, X_k\}$. Let $(\hat{S}^{(1)}, \hat{S}^{(2)}, \hat{S}^{(3)})$ be such that $\hat{S}^{(1)}$ and $\hat{S}^{(2)}$ are conditionally independent given $\hat{S}^{(3)}$ but otherwise share the same pairwise marginals of $(S^{(1)}, S^{(2)}, S^{(3)})$ as shown below. Then the probability of joint typicality of $(\hat{S}^{(1)}, \hat{S}^{(2)}, \hat{S}^{(3)})$ is as follows (see [9]).

Lemma 1.2: Let

$$P(\mathbf{S}^{(1)} = \mathbf{s}_1, \mathbf{S}^{(2)} = \mathbf{s}_2, \mathbf{S}^{(3)} = \mathbf{s}_3) = \prod_{i=1}^{b} p(s_{1i}, s_{2i}, s_{3i})$$

and

$$P(\hat{\mathbf{S}}^{(1)} = \mathbf{s}_1, \hat{\mathbf{S}}^{(2)} = \mathbf{s}_2, \hat{\mathbf{S}}^{(3)} = \mathbf{s}_3) = \prod_{i=1}^{b} p(s_{1i}|s_{3i})\, p(s_{2i}|s_{3i})\, p(s_{3i}).$$

Then for b such that $P(A_\epsilon^{(b)}(S^{(1)}, S^{(2)}, S^{(3)})) \geq 1 - \epsilon$,

$$\begin{aligned}(1-\epsilon)2^{-b(I(S^{(1)};S^{(2)}|S^{(3)})+7\epsilon)} &\leq P((\hat{\mathbf{S}}^{(1)}, \hat{\mathbf{S}}^{(2)}, \hat{\mathbf{S}}^{(3)}) \in A_\epsilon^{(b)}(S^{(1)}, S^{(2)}, S^{(3)})) \\ &\leq 2^{-b(I(S^{(1)};S^{(2)}|S^{(3)})-7\epsilon)}.\end{aligned}$$

We next discuss the proof of Theorem 2.1, which obtains an achievable rate in a network of nodes communicating over a general V-DMC and having a single source-destination pair.

APPENDIX II
PROOF OF THEOREM 2.1

The proof is a generalization of the one given in [8] for the *single* relay channel. As Theorem 2.1 deals with a network of arbitrary size and topology, its proof discussed next is rather involved. A prior study of the much simpler case in [8] may aid in understanding this proof.

The basic approach of the proof is similar to [8]. Relay nodes at a particular level in a given flow graph perform *successive cancellation* to decode the information transmitted by the nodes in the previous levels, including the source node. They then employ *random conditional codebooks* and *random partitions* to encode their own information as well as aid the transmissions by the nodes in the subsequent levels, in order to jointly relay information from the source node to the nodes in the levels even further down, including the destination node. Complication in the proof arises due to *multiple* levels of relaying, with each level possibly having multiple nodes, all of which need to perform many levels of decoding and encoding jointly in order to achieve the rate given in Theorem 2.1. This requires that the random codebooks, the random partitions, and the encoding and decoding procedures for various nodes in the network are carefully chosen. We make this precise in the following.

Fix a flow graph $\mathcal{G}(s, d)$. For this flow graph, fix $x_F \in \mathcal{X}_F$ and a probability mass function $p(x_0, x_1, \ldots, x_M \,|\, x_F)$, for $(x_0, x_1, \ldots, x_M) \in \mathcal{X}_0 \times \mathcal{X}_1 \times \cdots \times \mathcal{X}_M$. Consider a rate R_0 satisfying (5) for some $\{(R_1, R_2, \ldots, R_M) : R_0 \geq R_1 \geq \cdots \geq R_M \geq 0\}$. As in [8], we use *block Markov encoding* to achieve rate R_0. Consider B blocks of transmission, each of b symbols. A sequence of $(B-M)$ messages $w_i \in [1, 2^{bR_0}], i = 1, 2, \ldots, (B-M)$, will be sent from source s to destination d in bB transmissions. As $B \to \infty$, for fixed b and M, the rate $\frac{R_0(B-M)}{B}$ is arbitrarily close to R_0.

Generation of random codebooks

In each b-block $i = 1, 2, \ldots, B$, node j in level L_m of $\mathcal{G}(s, d)$ uses the same random codebook, denoted by $\mathcal{C}_{m,j}$, to determine the codeword to transmit in block i, for $j \in [1, n_m]$, $m \in [0, M]$. Now, all the random codebooks employed by nodes in a particular level L_m, denoted by $\mathcal{C}_m := \{\mathcal{C}_{m,j} : j \in [1, n_m]\}$, are generated according to a specific *joint* conditional distribution, as follows:

- First construct the codebooks $\mathcal{C}_M$ for nodes in level L_M by generating at random 2^{bR_M} independent and identically distributed b-sequences $\mathbf{x}_M(s_M) = (x_{M1}(s_M), x_{M2}(s_M), \ldots, x_{Mb}(s_M))$, $s_M \in [1, 2^{bR_M}]$, drawn according to

$$p\Big(\mathbf{x}_M(s_M)\Big| x_F\Big) = \prod_{l=1}^{b} p\Big(x_{Ml}(s_M)\Big| x_F\Big).$$

- Next construct the codebooks $\mathcal{C}_{M-1}$ for L_{M-1} as follows. Consider a node at L_{M-1}. It selects a codeword to transmit that depends not only on the index of its own level, denoted by s_{M-1}, but also on the index s_M transmitted by the higher level L_M. Specifically, $\mathcal{C}_{M-1}$ is obtained as follows. For each $\mathbf{x}_M(s_M), s_M \in [1, 2^{bR_M}]$, generate $2^{bR_{M-1}}$ conditionally independent b-sequences, denoted by $\mathbf{x}_{M-1}(s_{M-1}|s_M)$, $s_{M-1} \in [1, 2^{bR_{M-1}}]$, drawn according to

$$p\Big(\mathbf{x}_{M-1}(s_{M-1}|s_M)\,\Big|\,\mathbf{x}_M(s_M), x_F\Big) = \prod_{l=1}^{b} p\Big(x_{(M-1)l}(s_{M-1}|s_M)\,\Big|\,x_{Ml}(s_M), x_F\Big).$$

- Next, proceeding recursively, and supposing that $C_M, C_{M-1}, \ldots, C_{m+1}$ have been generated, we use the following procedure to construct the codebooks C_m for L_m. As at L_{M-1}, a node at level L_m selects a codeword to transmit that depends not only on the index s_m but also on the indices $s_{m+1}, s_{m+2}, \dots, s_M$ transmitted by higher levels. This way the nodes at level L_m not only send the index of their own level but also "aid" the transmissions of the nodes in higher levels L_{m+k}, $k \in [1, M-m]$. More specifically, let $s_{[k,M]}$ denote the vector $(s_k, s_{k+1}, \dots, s_M)$, then $\mathcal{C}_m$ is obtained as follows. For each $\big(\mathbf{x}_{m+1}(s_{m+1}|s_{[m+2,M]}), \quad \mathbf{x}_{m+2}(s_{m+2}|s_{[m+3,M]}), \dots, \mathbf{x}_{M-1}(s_{M-1}|s_M), \mathbf{x}_M(s_M)\big)$, $s_l \in [1, 2^{bR_l}]$, $l \in [m+1, M]$, generate 2^{bR_m} conditionally independent b-sequences, $\mathbf{x}_m(s_m|s_{[m+1,M]})$, $s_m \in [1, 2^{bR_m}]$, drawn according to

$$
\begin{aligned}
&p\Big(\mathbf{x}_m(s_m|s_{[m+1,M]}) \,\Big|\, \mathbf{x}_{m+1}(s_{m+1}|s_{[m+2,M]}), \dots, \\
&\qquad \mathbf{x}_{M-1}(s_{M-1}|s_M), \mathbf{x}_M(s_M), x_F\Big) \\
&= \prod_{l=1}^{b} p\Big(x_{ml}(s_m|s_{[m+1,M]}) \,\Big|\, x_{(m+1)l}(s_{m+1}|s_{[m+2,M]}), \\
&\qquad \dots, x_{(M-1)l}(s_{M-1}|s_M), x_{Ml}(s_M), x_F\Big).
\end{aligned}
$$

Random partitions

We next construct a nested set of random partitions $\mathcal{S}^0, \mathcal{S}^1, \dots, \mathcal{S}^M$ of $[1, 2^{bR_0}]$. These nested random partitions will allow us to recursively use the random binning argument, introduced in [27] and used in [8] for the single-relay channel, to show that R_0 is achievable.

Let $\mathcal{S}^0 = [1, 2^{bR_0}]$. Having obtained $\mathcal{S}^{m-1} = \{S_1^{m-1}, S_2^{m-1}, \dots, S_{2^{bR_{m-1}}}^{m-1}\}$, we generate a random partition $\mathcal{S}^m = \{S_1^m, S_2^m, \dots, S_{2^{bR_m}}^m\}$ of $\mathcal{S}^{m-1}$ by randomly and independently assigning $S_k^{m-1}, k \in [1, 2^{bR_{m-1}}]$, to S_l^m according to a uniform distribution over the indices $l = 1, 2, \dots, 2^{bR_m}$. That is, all the indices in the set S_k^{m-1} are made members of the set S_l^m.

Encoding

The above random codebooks and nested random partitions are employed to do encoding as follows. Let $w_i \in [1, 2^{bR_0}]$ be the new index to be sent from source s in block i, for destination d. As indicated during the codebook generation step, node $L_{m,j}$ at level L_m transmits in block i a codeword from its codebook $\mathcal{C}_{m,j}$ that depends not only on its estimate of $s_{m;i}$, the index that L_m has to send, but also on its estimates of $\{s_{m+k;i} : k \in [1, M-m]\}$, the indices that higher levels $\{L_{m+k} : k \in [1, M-m]\}$ send in block i.

To make the above precise, we will use the following notation:

1) $\hat{s}_{m;i}^{m,j}$:= Estimate by node $L_{m,j}$ of $s_{m;i}$, the index that level L_m needs to send in block i.
2) $\hat{s}_{m+k;i}^{m,j}$:= "k-step forward" estimate by node $L_{m,j}$ of $s_{m+k;i}$, the index to be sent by level L_{m+k} in block i. This estimate is obtained by $L_{m,j}$ using "set inclusion," defined in the sequel, at the end of block $(i-k-1)$, for $k \in [1, M-m]$.
3) $\hat{s}_{m-k;i}^{m,j}$:= "k-step backward" estimate by node $L_{m,j}$ of $s_{m-k;i}$, the index sent by level L_{m-k} in block i. This estimate is made by $L_{m,j}$ using joint typicality at the end of block $(i+k-1)$, for $k \in [1, m]$ (details will be given in the sequel).

In order to initialize the encoding procedure, as well as the decoding procedure described in the sequel, $\{w_{-(M-1)}, w_{-(M-2)}, \dots, w_{-1}, w_0\}$ are independently and uniformly chosen from $[1, 2^{bR_0}]$, and are made known to all the nodes $\{L_{m,j} : m \in [0, M+1], j \in [1, n_m]\}$ before the beginning of block 1.

Now the indices $\{\hat{s}_{m;i}^{m,j} \in [1, 2^{bR_m}] : m \in [0, M], j \in [1, n_m]\}$ are obtained as follows:

- $\hat{s}_{0;i}^{0,1} = w_i$, i.e., the source node simply chooses the symbol w_i,
- For $m \in [1, M]$, $j \in [1, n_m]$, $\hat{s}_{m;i}^{m,j}$ is the index of the cell in partition $\mathcal{S}^m$ which contains $S_{\hat{s}_{m-1;i-1}^{m,j}}^{m-1}$ (the cell in $\mathcal{S}^{m-1}$ indexed by $\hat{s}_{m-1;i-1}^{m,j}$). How the node $L_{m,j}$ obtains the one-step backward estimate $\hat{s}_{m-1;i-1}^{m,j}$ at the end of block $(i-1)$ will be specified in the sequel.

Next, node $L_{m,j}$ obtains the forward estimates $\{\hat{s}_{m+k;i}^{m,j}\}$, based on $\{\hat{s}_{m;i-k}^{m,j}\}$, using set inclusion, as follows. For each $k \in [1, M-m]$, $L_{m,j}$ estimates $s_{m+k;i}$ as $\hat{s}_{m+k;i}^{m,j} = s$, where $s \in [1, 2^{bR_{m+k}}]$ is such that $S_{\hat{s}_{m;i-k}^{m,j}}^m \subseteq S_s^{m+k}$. Note that $\hat{s}_{m+k;i}^{m,j}$ depends only on $\hat{s}_{m;i-k}^{m,j}$, which in turn depends only on $\hat{s}_{m-1;i-k-1}^{m,j}$. Thus the estimate $\hat{s}_{m+k;i}^{m,j}$ can be made by node $L_{m,j}$ at the end of block $(i-k-1)$.

Let

$$
\hat{s}_{[m+1,M];i}^{m,j} \quad := \quad \Big(\hat{s}_{m+1;i}^{m,j}, \hat{s}_{m+2;i}^{m,j}, \dots, \hat{s}_{M;i}^{m,j}\Big).
$$

Then, having obtained the estimates above, which one should note can be made at the end of block $(i-1)$, node $L_{m,j}$ transmits in block i the codeword $\mathbf{x}_{m,j}(\hat{s}_{m;i}^{m,j}|\ \hat{s}_{[m+1,M];i}^{m,j}) \in \mathcal{C}_{m,j}$.

Decoding

The decoding will be defined in an iterative fashion. Suppose that at the end of block $(i-1)$, node $L_{m,j}$ has estimates $(\hat{s}_{m-1;1}^{m,j}, \hat{s}_{m-1;2}^{m,j}, \dots, \hat{s}_{m-1;i-1}^{m,j})$, $(\hat{s}_{m-2;1}^{m,j}, \hat{s}_{m-2;2}^{m,j}, \dots, \hat{s}_{m-2;i-2}^{m,j}), \dots,$ $(\hat{s}_{0;1}^{m,j}, \hat{s}_{0;2}^{m,j}, \dots, \hat{s}_{0;i-m}^{m,j})$. Decoding at node $L_{m,j}$ at the end of block i then proceeds as follows:

- $L_{m,j}$ first obtains the one-step backward estimate $\hat{s}_{m-1;i}^{m,j}$ of $s_{m-1;i}$, the index sent by level L_{m-1} in block i, as follows. Using the forward estimates $\hat{s}_{m;i}^{m,j}, \hat{s}_{m+1;i}^{m,j}, \dots, \hat{s}_{M;i}^{m,j}$ and upon receiving the signal $\mathbf{y}_{m,j;i}$, node $L_{m,j}$ estimates $\hat{s}_{m-1;i}^{m,j} = s$ *iff* there exists a *unique* index $s \in [1, 2^{bR_{m-1}}]$ such that $s \in \mathcal{L}_{m,j,1;i}$, where

$$
\begin{aligned}
\mathcal{L}_{m,j,1;i} \ := \ \Big\{ s \in [1, 2^{bR_{m-1}}] : \ \Big(&\mathbf{x}_{m-1}(s|\hat{s}_{[m,M];i}^{m,j}), \\
&\mathbf{x}_m(\hat{s}_{m;i}^{m,j}|\hat{s}_{[m+1,M];i}^{m,j}), \dots, \mathbf{x}_M(\hat{s}_{M;i}^{m,j}), \\
&x_F, \mathbf{y}_{m,j;i}\Big) \text{ are jointly } \epsilon\text{-typical} \Big\}. \quad (25)
\end{aligned}
$$

In the usual fashion of information theory, an error event captures situations where there does not exist any such estimate s, or there is more than one such s. This contingency will be taken into account in the probability of error calculations in the sequel.

- Next $L_{m,j}$ iteratively decodes the indices sent by levels $L_{m-k}, k = 2, 3, \ldots, m$, as follows. Having obtained backward estimates $\hat{s}^{m,j}_{m-1;i}, \hat{s}^{m,j}_{m-2;i-1}, \ldots, \hat{s}^{m,j}_{m-k+1;i-k+2}$, node $L_{m,j}$ estimates $\hat{s}^{m,j}_{m-k;i-k+1} = s$ as the index sent by L_{m-k} in block $i-k+1$ *iff* there exists a *unique* $s \in [1, 2^{bR_{m-k}}]$ such that $s \in S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}} \cap \mathcal{L}_{m,j,k;i}$, where

$$\begin{aligned}&\mathcal{L}_{m,j,k;i}\\ &:= \Big\{ s \in [1, 2^{bR_{m-k}}] : \Big(\mathbf{x}_{m-k}(s \,|\, \hat{s}^{m,j}_{[m-k+1,M];i-k+1}),\\ &\quad \mathbf{x}_{m-k+1}(\hat{s}^{m,j}_{m-k+1;i-k+1} \,|\, \hat{s}^{m,j}_{[m-k+2,M];i-k+1}),\\ &\quad \ldots, \mathbf{x}_M(\hat{s}^{m,j}_{M;i-k+1}), x_F, \mathbf{y}_{m,j;i-k+1}\Big) \text{ are}\\ &\quad \text{jointly } \epsilon\text{-typical}\Big\}. \quad (26)\end{aligned}$$

As above, an error occurs when such a unique s does not exist.

Probability of error calculations

Let w_i be the index to be sent from source $L_{0,1}$ in block i, for destination $L_{M+1,1}$. Define the "ideal" indices to be sent by levels $\{L_m : m \in [0, M]\}$ as follows:

- $\bar{s}_{0;i} = w_i$,
- For $m \in [1, M]$, $\bar{s}_{m;i} = \bar{s}$, where $\bar{s} \in [1, 2^{bR_m}]$ is such that $S^{m-1}_{\bar{s}_{m-1;i-1}} \subseteq S^m_{\bar{s}}$.

For the random coding scheme described above, we will declare an error in block i if the following event occurs

$$F(i) \quad := \quad E_0(i) \cup \bigcup_{m=1}^{M+1} \bigcup_{j=1}^{n_m} \bigcup_{k=1}^{m} E_{m,j,k}(i), \qquad (27)$$

where

1) $E_0(i)$ is the event that $\big(\{\mathbf{x}_{m,j;i}(\hat{s}^{m,j}_{m;i} \,|\, \hat{s}^{m,j}_{[m+1,M];i}) : m \in [0, M], j \in [1, n_m]\}, x_F, \{\mathbf{y}_{m',j';i} : m' \in [1, M+1], j' \in [1, n_{m'}]\}\big)$ are not jointly ϵ-typical,
2) $E_{m,j,1}(i)$ is the event that node $L_{m,j}$ fails to obtain one-step backward estimate $\hat{s}^{m,j}_{m-1;i}$ at the end of block i that matches the ideal symbol $\bar{s}_{m-1;i}$, i.e., either $\bar{s}_{m-1;i} \notin \mathcal{L}_{m,j,1;i}$ or there exists $s \neq \bar{s}_{m-1;i}$ such that $s \in \mathcal{L}_{m,j,1;i}$,
3) For $k \in [2, m]$, $E_{m,j,k}(i)$ is the event that node $L_{m,j}$ fails to obtain k-step backward estimate $\hat{s}^{m,j}_{m-k;i-k+1}$ at the end of block i that matches the ideal symbol $\bar{s}_{m-k;i-k+1}$, i.e., either $\bar{s}_{m-k;i-k+1} \notin S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}} \cap \mathcal{L}_{m,j,k;i}$ or there exists $s \neq \bar{s}_{m-k;i-k+1}$ such that $s \in S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}} \cap \mathcal{L}_{m,j,k;i}$.

Now let

$$F_0(i-1) \quad := \quad \bigcup_{l=1}^{i-1} F(l). \qquad (28)$$

Thus $F_0^c(i-1)$ is the event that all the decoding steps at all nodes till the end of block $(i-1)$ are error free.

We will now show that $P(F(i) \cap F_0^c(i-1))$ is small. This will lead to a recursive proof that $P(\cup_{i=1}^B F(i))$ is small. We begin as follows.

Lemma 2.1: For b sufficiently large,

$$P(E_0(i) \cap F_0^c(i-1)) \quad \leq \quad \epsilon. \qquad (29)$$

Proof: The event $F_0^c(i-1)$ implies that the following hold:

$$\hat{s}^{m,j}_{m-k;l-k+1} \quad = \quad \bar{s}_{m-k;l-k+1}, \qquad (30)$$

for each $m \in [1, M+1]$, $j \in [1, n_m]$, $k \in [1, m]$, $l \in [1, i-1]$. Furthermore, $\big(\{\mathbf{x}_{m,j;l}(\hat{s}^{m,j}_{m;l} | \hat{s}^{m,j}_{[m+1,M];l}) : m \in [0, M], j \in [1, n_m]\}, x_F, \{\mathbf{y}_{m',j';l} : m' \in [1, M+1], j' \in [1, n_{m'}]\}\big)$ are jointly ϵ-typical, for each $l \in [1, i-1]$.

Now $\hat{s}^{m,j}_{m+k;i}$ is a forward estimate that depends only on $\hat{s}^{m,j}_{m;i-k}$, which in turn depends only on $\hat{s}^{m,j}_{m-1;i-k-1}$ made at the end of block $(i-k-1)$. Hence $\hat{s}^{m,j}_{[m+1,M];i}$ depends only on the estimates $\{\hat{s}^{m,j}_{m-1;l} : l \leq i-2\}$ made by the end of block $(i-2)$.

Hence

$$F_0^c(i-1) \quad \Rightarrow \quad \hat{s}^{m,j}_{[m+1,M];i} \;=\; \bar{s}_{[m+1,M];i}.$$

Moreover, by (30), we get that

$$\begin{aligned}F_0^c(i-1) \quad &\Rightarrow \quad \hat{s}^{m,j}_{m-1;i-1} \;=\; \bar{s}_{m-1;i-1}\\ &\Rightarrow \quad \hat{s}^{m,j}_{m;i} \;=\; \bar{s}_{m;i},\end{aligned}$$

since the dependence of $\hat{s}^{m,j}_{m;i}$ on $\hat{s}^{m,j}_{m-1;i-1}$ is in the same way as $\bar{s}_{m;i}$ depends on $\bar{s}_{m-1;i-1}$.

Thus

$$\begin{aligned}&P(E_0(i) \cap F_0^c(i-1))\\ &\leq P\Big(\Big(\{\mathbf{x}_{m,j;i}(\bar{s}_{m;i} \,|\, \bar{s}_{[m+1,M];i}) : m \in [0, M], j \in [1, n_m]\},\\ &\qquad x_F, \{\mathbf{y}_{m',j';i} : m' \in [1, M+1], j' \in [1, n_{m'}]\}\Big) \text{ are not}\\ &\qquad \text{jointly } \epsilon\text{-typical}\Big).\end{aligned}$$

Hence, from Lemma 1.1(i), we get that

$$P(E_0(i) \cap F_0^c(i-1)) \quad \leq \quad \epsilon,$$

when b is sufficiently large. This completes the proof of Lemma 2.1. ∎

Next let

$$F_{m,j,k-1}(i) \;:=\; \begin{cases} E_0(i), & k = 1,\\ E_0(i) \cup \bigcup_{l=1}^{k-1} E_{m,j,l}(i), & k \in [2, m].\end{cases}$$

Hence $F^c_{m,j,k-1}(i)$ is the event that all the decoding steps conducted at $L_{m,j}$ at the end of block i for layers $\{L_{m-k+1}, \ldots, L_{m-2}, L_{m-1}\}$ are error free, as well as the joint typicality property holds.

We first show that the conditional probability of error in obtaining the one-step backward estimate, $P\big(E_{m,j,1}(i) \cap F^c_{m,j,0}(i) \cap F_0^c(i-1)\big)$, is small.

Lemma 2.2: If

$$R_{m-1} \quad < \quad I(X_{m-1}; Y_{m,j} \,|\, X_{[m,M]}, x_F) - 7\epsilon,$$

then for sufficiently large b

$$P\Big(E_{m,j,1}(i) \cap F^c_{m,j,0}(i) \cap F^c_0(i-1)\Big) \quad \leq \quad \epsilon.$$

Proof: Recall from (25) that $\mathcal{L}_{m,j,1;i}$ is the set of all $s \in [1, 2^{bR_{m-1}}]$ such that $\big(\mathbf{x}_{m-1}(s|\hat{s}^{m,j}_{[m,M];i}), \mathbf{x}_m(\hat{s}^{m,j}_{m;i}|\hat{s}^{m,j}_{[m+1,M];i}), \ldots, \mathbf{x}_M(\hat{s}^{m,j}_{M;i}), x_F, \mathbf{y}_{m,j;i}\big)$ are jointly ϵ-typical. Let $E_{m,j,1}(i) = E'_{m,j,1}(i) \cup E''_{m,j,1}(i)$, where

- $E'_{m,j,1}(i)$ is the event that $\bar{s}_{m-1;i} \notin \mathcal{L}_{m,j,1;i}$, and
- $E''_{m,j,1}(i)$ is the event that there exists $s \neq \bar{s}_{m-1;i}$ such that $s \in \mathcal{L}_{m,j,1;i}$.

As in Lemma 2.1, the event $F^c_0(i-1)$ implies that $E'_{m,j,1}(i)$ is the event that $\big(\mathbf{x}_{m-1}(\bar{s}_{m-1;i}| \bar{s}_{[m,M];i}), \mathbf{x}_m(\bar{s}_{m;i}|\bar{s}_{[m+1,M];i}), \ldots, \mathbf{x}_M(\bar{s}_{M;i}), x_F, \mathbf{y}_{m,j;i}\big)$ are not jointly ϵ-typical.

Also, $F^c_0(i-1)$ implies that $F^c_{m,j,0}(i) = E^c_0(i)$ is the event that $\big(\{x_{m',j';i}(\bar{s}_{m';i}|\bar{s}_{[m'+1,M];i}) : m' \in [0,M], j' \in [1, n_{m'}]\}, x_F, \{\mathbf{y}_{m'',j'';i} : m'' \in [1, M+1], j'' \in [1, n_{m''}]\}\big)$ are jointly ϵ-typical. In particular, the event $F^c_0(i-1) \cap F^c_{m,j,0}(i)$ implies that $\big(\mathbf{x}_{m-1}(\bar{s}_{m-1;i}|\bar{s}_{[m,M];i}), \mathbf{x}_m(\bar{s}_{m;i}|\bar{s}_{[m+1,M];i}), \ldots, \mathbf{x}_M(\bar{s}_{M;i}), x_F, \mathbf{y}_{m,j;i}\big)$ are jointly ϵ-typical.

Thus

$$P(E'_{m,j,1}(i) \cap F^c_0(i-1) \cap F^c_{m,j,0}(i)) \quad = \quad 0. \qquad (31)$$

Hence

$$\begin{aligned}
&P(E_{m,j,1}(i) \cap F^c_{m,j,0}(i) \cap F^c_0(i-1)) \\
&\stackrel{(a)}{=} P(E''_{m,j,1}(i) \cap F^c_{m,j,0}(i) \cap F^c_0(i-1)) \\
&\stackrel{(b)}{=} P\Big(\Big\{\exists\, s \in [1, 2^{bR_{m-1}}] \setminus \{\bar{s}_{m-1;i}\} \text{ s.t. } s \in \mathcal{L}_{m,j,1;i}\Big\} \cap F^c_{m,j,0}(i) \cap F^c_0(i-1)\Big) \\
&\stackrel{(c)}{\leq} \sum_{\substack{s \neq \bar{s}_{m-1;i} \\ s \in [1, 2^{bR_{m-1}}]}} P\Big(\Big(\mathbf{x}_{m-1}(s|\bar{s}_{[m,M];i}), \mathbf{x}_m(\bar{s}_{m;i}|\bar{s}_{[m+1,M];i}), \ldots, \mathbf{x}_M(\bar{s}_{M;i}), x_F, \\
&\qquad \mathbf{y}_{m,j;i}\big(\{\mathbf{x}_{m'}(\bar{s}_{m';i}|\bar{s}_{[m'+1,M];i}) : m' \in [0,M]\}, x_F\big)\Big) \text{ are jointly } \epsilon\text{-typical}\Big) \\
&\stackrel{(d)}{\leq} 2^{bR_{m-1}} \cdot 2^{b(-I(X_{m-1}; Y_{m,j} \,|\, X_{[m+M]}, x_F) + 7\epsilon)},
\end{aligned}$$

where (a) uses (31), (b) is from the definition of $E''_{m,j,1}(i)$, (c) follows from an argument similar to Lemma 2.1, and (d) holds due to Lemma 1.2 since $\mathbf{x}_{m-1}(s|\bar{s}_{[m,M];i})$ and $\mathbf{x}_{m-1}(\bar{s}_{m-1;i}|\bar{s}_{[m,M];i})$ are conditionally independent given $(\mathbf{x}_m(\bar{s}_{m;i}|\bar{s}_{[m+1,M];i}), \ldots, \mathbf{x}_M(\bar{s}_{M;i}), x_F)$ for each $s \neq \bar{s}_{m-1;i}$.

Thus if

$$R_{m-1} \quad < \quad I(X_{m-1}; Y_{m,j} \,|\, X_{[m,M]}, x_F) - 7\epsilon,$$

then for sufficiently large b,

$$P\Big(E_{m,j,1}(i) \cap F^c_{m,j,0}(i) \cap F^c_0(i-1)\Big) \quad \leq \quad \epsilon,$$

which completes the proof of Lemma 2.2. ∎

We next show that the conditional probability of error in obtaining the k-step backward estimate, $P(E_{m,j,k}(i) \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1))$, can also be made small for each $k \in [2, m]$.

Lemma 2.3: If

$$R_{m-k} \; < \; R_{m-k+1} + I(X_{m-k}; Y_{m,j}|X_{[m-k+1,M]}, x_F) - 7\epsilon,$$

then for sufficiently large b

$$P\Big(E_{m,j,k}(i) \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \quad \leq \quad \epsilon,$$

for $k = 2, 3, \ldots, m$.

Proof: Recall that $\mathcal{L}_{m,j,k;i}$ is the set of all $s \in [1, 2^{bR_{m-k}}]$ such that $\big(\mathbf{x}_{m-k}(s|\hat{s}^{m,j}_{[m-k+1,M];i-k+1}), \mathbf{x}_{m-k+1}(\hat{s}^{m,j}_{m-k+1;i-k+1}|\hat{s}^{m,j}_{[m-k+2,M];i-k+1}), \ldots, \mathbf{x}_M(\hat{s}^{m,j}_{M;i-k+1}), x_F, \mathbf{y}_{m,j;i-k+1}\big)$ are jointly ϵ-typical. Let $E_{m,j,k}(i) = E'_{m,j,k}(i) \cup E''_{m,j,k}(i)$, where

- $E'_{m,j,k}(i)$ is the event that $\bar{s}_{m-k;i-k+1} \notin S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}} \cap \mathcal{L}_{m,j,k;i}$, and
- $E''_{m,j,k}(i)$ is the event that there exists $s \neq \bar{s}_{m-k;i-k+1}$ such that $s \in S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}} \cap \mathcal{L}_{m,j,k;i}$.

Now

$$\begin{aligned}
&P\Big(E'_{m,j,k}(i) \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \\
&\leq P\Big(\{\bar{s}_{m-k;i-k+1} \notin \mathcal{L}_{m,j,k;i}\} \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \\
&\quad + P\Big(\{\bar{s}_{m-k;i-k+1} \notin S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}}\} \\
&\qquad \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big). \qquad (32)
\end{aligned}$$

We next evaluate the two terms on the RHS of (32) separately. For the first term, arguing along lines similar to (31), we get that

$$\begin{aligned}
&P\Big(\{\bar{s}_{m-k;i-k+1} \notin \mathcal{L}_{m,j,k;i}\} \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \\
&\quad = 0. \qquad (33)
\end{aligned}$$

Now consider the second term in (32)

$$\begin{aligned}
&P\Big(\{\bar{s}_{m-k;i-k+1} \notin S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}}\} \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \\
&\stackrel{(a)}{\leq} P\Big(\bar{s}_{m-k;i-k+1} \notin S^{m-k+1}_{\bar{s}_{m-k+1;i-k+2}}\Big) \\
&\stackrel{(b)}{=} 0, \qquad (34)
\end{aligned}$$

where (a) follows from the definitions of $F_{m,j,k-1}(i)$ and $F_0(i-1)$ and an argument similar to Lemma 2.1, and the equality (b) follows from the definition of $\bar{s}_{m-k+1;i-k+2}$. Substituting (33) and (34) in (32), we get

$$P\Big(E'_{m,j,k}(i) \cap F^c_{m,j,k-1}(i) \cap F^c_0(i-1)\Big) \quad = \quad 0. \quad (35)$$

Hence we obtain the bound given at the top of the next page, where (a) uses (35), (b) is from the definition of $E''_{m,j,k}(i)$, (c) follows from an argument similar to Lemma 2.1, and (d) follows from the fact that each $s \in [1, 2^{bR_{m-k}}]$ is independently and uniformly assigned to S^{m-k+1}_l, $l \in [1, 2^{bR_{m-k+1}}]$, and from Lemma 1.2, as, for each $s \neq \bar{s}_{m-k;i-k+1}$, $\mathbf{x}_{m-k}(s|\bar{s}_{[m-k+1,M];i-k+1})$ and

$$
\begin{aligned}
&P\Big(E_{m,j,k}(i)\cap F^c_{m,j,k-1}(i)\,\cap\,F^c_0(i-1)\Big)\\
&\overset{(a)}{=}\quad P\Big(E''_{m,j,k}(i)\,\cap\,F^c_0(i-1)\cap F^c_{m,j,k-1}(i)\Big)\\
&\overset{(b)}{=}\quad P\bigg(\Big\{\exists\, s\in[1,2^{bR_{m-k}}]\setminus \bar{s}_{m-k;i-k+1}\ \text{ s.t. }\ s\in\mathcal{L}_{m,j,k;i}\cap S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}}\Big\}\,\cap\,F^c_0(i-1)\,\cap\,F^c_{m,j,k-1}(i)\bigg)\\
&=\ P\bigg(\bigcup_{\substack{s\neq\bar{s}_{m-k;i-k+1}\\ s\in[1,2^{bR_{m-k}}]}}\bigg(\Big(\Big\{s\in S^{m-k+1}_{\hat{s}^{m,j}_{m-k+1;i-k+2}}\Big\}\cap F^c_0(i-1)\cap F^c_{m,j,k-1}(i)\Big)\cap\Big(\Big\{s\in\mathcal{L}_{m,j,k;i}\Big\}\cap F^c_0(i-1)\cap F^c_{m,j,k-1}(i)\Big)\bigg)\bigg)\\
&\overset{(c)}{\le}\ P\bigg(\bigcup_{\substack{s\neq\bar{s}_{m-k;i-k+1}\\ s\in[1,2^{bR_{m-k}}]}}\Big\{s\in S^{m-k+1}_{\bar{s}_{m-k+1;i-k+2}}\Big\}\cap\Big\{\Big(\mathbf{x}_{m-k}(s|\bar{s}_{[m-k+1,M];i-k+1}),\mathbf{x}_{m-k+1}(\bar{s}_{m-k+1;i-k+1}|\bar{s}_{[m-k+2,M];i-k+1}),\\
&\qquad\ldots,\mathbf{x}_M(\bar{s}_{M;i-k+1}),x_F,\mathbf{y}_{m,j;i-k+1}\big(\{\mathbf{x}_{m'}(\bar{s}_{m';i-k+1}|\bar{s}_{[m'+1,M];i-k+1}):m'\in[0,M]\},x_F\big)\Big)\text{ are jointly }\epsilon\text{-typical}\Big\}\bigg)\\
&\le\ \sum_{\substack{s\neq\bar{s}_{m-k;i-k+1}\\ s\in[1,2^{bR_{m-k}}]}}P\Big(\Big\{s\in S^{m-k+1}_{\bar{s}_{m-k+1;i-k+2}}\Big\}\Big)\\
&\qquad\cdot P\bigg(\Big\{\Big(\mathbf{x}_{m-k}(s|\bar{s}_{[m-k+1,M];i-k+1}),\mathbf{x}_{m-k+1}(\bar{s}_{m-k+1;i-k+1}|\bar{s}_{[m-k+2,M];i-k+1}),\ \ldots,\ \mathbf{x}_M(\bar{s}_{M;i-k+1}),\ x_F,\\
&\qquad\mathbf{y}_{m,j;i-k+1}\big(\{\mathbf{x}_{m'}(\bar{s}_{m';i-k+1}|\bar{s}_{[m'+1,M];i-k+1}):m'\in[0,M]\},x_F\big)\Big)\text{ are jointly }\epsilon\text{-typical}\Big\}\ \Big|\ \Big\{s\in S^{m-k+1}_{\bar{s}_{m-k+1;i-k+2}}\Big\}\bigg)\\
&\overset{(d)}{\le}\ \sum_{\substack{s\neq\bar{s}_{m-k;i-k+1}\\ s\in[1,2^{bR_{m-k}}]}}2^{-bR_{m-k+1}}\cdot 2^{-b(I(X_{m-k};Y_{m,j}\,|\,X_{[m-k+1,M]},x_F)-7\epsilon)}\\
&=\ \big(2^{bR_{m-k}}-1\big)\cdot 2^{-bR_{m-k+1}}\cdot 2^{-b(I(X_{m-k};Y_{m,j}\,|\,X_{[m-k+1,M]},x_F)-7\epsilon)}\,.
\end{aligned}
$$

$\mathbf{x}_{m-k}(\bar{s}_{m-k;i-k+1}|\bar{s}_{[m-k+1,M];i-k+1})$ are conditionally independent given $\big(\mathbf{x}_{m-k+1}(\bar{s}_{m-k+1;i-k+1}|\bar{s}_{[m-k+2,M];i-k+1}),\ \ldots,\mathbf{x}_M(\bar{s}_{M;i-k+1}),\ x_F\big)$.

Thus if

$$R_{m-k}\ <\ R_{m-k+1}+I(X_{m-k};Y_{m,j}|X_{[m-k+1,M]},x_F)-7\epsilon,$$

then for sufficiently large b,

$$P\Big(E_{m,j,k}(i)\cap F^c_{m,j,k-1}(i)\,\cap\,F^c_0(i-1)\Big)\quad\le\quad\epsilon,$$

and the proof of Lemma 2.3 is complete. ■

Now we bound the probability of error over the B blocks of size b sent from source s to destination d, for rate R_0 satisfying (5). Let $\mathbf{W}=(W_1,W_2,\ldots,W_{B-M},\emptyset,\ldots,\emptyset)$ be the sequence of indices transmitted by source s in B b-blocks. We assume the indices $\{W_i\}$ are i.i.d. random variables uniformly distributed on $[1,2^{bR_0}]$. Let $\hat{\mathbf{W}}=(\emptyset,\ldots,\emptyset,\hat{W}_1,\hat{W}_2,\ldots,\hat{W}_{B-M})$ be the sequence decoded at destination d at the end of block B, and let $\tilde{\mathbf{W}}:=(\hat{W}_1,\hat{W}_2,\ldots,\hat{W}_{B-M},\emptyset,\ldots,\emptyset)$. Then the probability of error is

$$
\begin{aligned}
&P(\tilde{\mathbf{W}}\neq\mathbf{W})\\
&\overset{(a)}{=}\quad P\left(\bigcup_{i=1}^{B}F(i)\right)\\
&=\quad\sum_{i=1}^{B}P\Big(F(i)\cap F^c_0(i-1)\Big)\\
&\overset{(b)}{=}\quad\sum_{i=1}^{B}P\bigg(\Big(E_0(i)\cup\bigcup_{m=1}^{M+1}\bigcup_{j=1}^{n_m}\bigcup_{k=1}^{m}E_{m,j,k}(i)\Big)\cap F^c_0(i-1)\bigg)\\
&\overset{(c)}{\le}\quad\sum_{i=1}^{B}\bigg(P\Big(E_0(i)\,\cap\,F^c_0(i-1)\Big)\,+\,\sum_{m=1}^{M+1}\sum_{j=1}^{n_m}\sum_{k=1}^{m}\\
&\qquad\qquad P\Big(E_{m,j,k}(i)\,\cap\,F^c_0(i-1)\cap F^c_{m,j,k-1}(i)\Big)\bigg)\\
&\overset{(d)}{\le}\quad B\Big(1+n(M+1)\Big)\epsilon\ =:\ \epsilon',
\end{aligned}
$$

where (a) and (b) follow from the definition of $F(i)$ in (27), (c) uses the definition of $F_{m,j,k-1}(i)$, and (d) uses Lemmas 2.1-2.3 and that n is the total number of nodes in $\mathcal{N}$.

Now, using the standard random coding arguments, we deduce that there exists a selection of codebooks $\{\mathcal{C}^*_{m,j}:m\in[0,M],\,j\in[1,n_m]\}$ for every node in each level of $\mathcal{G}(s,d)$ such that $P(\tilde{\mathbf{W}}\neq\mathbf{W}\,|\,\{\mathcal{C}^*_{m,j}\})\le\epsilon'$. By throwing away the worse half of the $\mathbf{w}$ in $\{1,2,\ldots,2^{bR_0}\}^{B-M}$ and reindexing them, we get that

$$P\Big(\tilde{\mathbf{W}}\neq\mathbf{w}_l\ \Big|\ \{\mathcal{C}^*_{m,j}\},\mathbf{W}=\mathbf{w}_l\Big)\quad\le\quad 2\epsilon',$$

for each $l \in [1, 2^{bR_0(B-M)-1}]$. Thus, for $\epsilon' > 0$, and b sufficiently large, the maximal probability of error $\lambda^{(b)} \leq 2\epsilon'$, for rates $\tilde{R}_0 = \frac{bR_0(B-M)-1}{bB}$, where R_0 satisfies (5). First letting $b \to \infty$, then $B \to \infty$, and finally $\epsilon' \to 0$, we see that $\tilde{R}_0 < R_0$ is achievable. This completes the proof of Theorem 2.1. ■

APPENDIX III
PROOF OF THEOREM 3.1

First consider the AWGN channel without fading (i.e., the channel gains $\{h_{l',l}\}$ are given by (7)). Fix a flow graph $\mathcal{G}(s,d)$. For this flow graph, let $x_F = 0$, and let $\{\beta_{m,j,m+k} \in [0,1),\, m \in [0,M], j \in [1,n_m], k \in [0, M-m]\}$ be such that

$$\sum_{k=0}^{M-m} \beta_{m,j,m+k} \;\;<\;\; 1, \qquad \text{for each } m \in [0,M],\, j \in [1,n_m].$$

Next, let V_m, $m \in [0,M]$, be i.i.d. real-valued Gaussian random variables with mean zero and covariance one. Let

$$X_{m,j} \;\;=\;\; \sum_{k=0}^{M-m} \gamma_{m,j,m+k}\; V_{m+k}, \qquad (36)$$

where

$$\gamma_{m,j,m+k} \;\;:=\;\; \sqrt{\beta_{m,j,m+k}\, P_{m,j}}\;.$$

Now codebooks for nodes in $\mathcal{G}(s,d)$ are generated as follows. Given a rate R_0 satisfying (11) for some $\{(R_1, R_2, \ldots, R_M) : R_0 \geq R_1 \geq \cdots \geq R_M \geq 0\}$, generate i.i.d. b-variate normal random vectors $\mathbf{v}_l(s)$, $s \in [1, 2^{bR_l}]$, $l \in [0,M]$, having mean zero and covariance I_b. Then, $\mathcal{C}_{m,j}$, the codebook for node $L_{m,j}$, is given by $\{\mathbf{x}_{m,j}(s_m | s_{[m+1,M]}) : s_l \in [1, 2^{bR_l}], l \in [m,M]\}$, where

$$\mathbf{x}_{m,j}\left(s_m \middle| s_{[m+1,M]}\right) \;=\; \sum_{k=0}^{M-m} \gamma_{m,j,m+k}\; \mathbf{v}_{m+k}(s_{m+k}). \quad (37)$$

The codewords so generated satisfy the power constraint (9) with high probability, when b is large enough. With these codebooks, the encoding and decoding procedures described in the proof of Theorem 2.1, and employing the standard discretization and quantization arguments used for computing the capacity of general memoryless channels (i.e., channels with not-necessarily finite input and output alphabets; see [13], Ch. 7), we deduce that the rate R_0 given by Theorem 2.1 is achievable with arbitrarily small probability of error whenever (5) holds. Now, from (6), (7), and (36), we obtain

$$\begin{aligned} &I\big(X_m; Y_{m+k,j} \,\big|\, X_{[m+1,M]},\, x_F\big) \\ &\qquad = \;\; C\left(\frac{P^{(R)}_{m+k,j,m}}{N_{m+k,j} + \sum_{k'=1}^{m} P^{(R)}_{m+k,j,m-k'}}\right), \end{aligned}$$

where $C(\cdot)$ is given by (13) and $P^{(R)}_{m+k,j,m-k'}$ is the total power received by node $L_{m+k,j}$ for $V_{m-k'}$. From (6), (7), and (36), we determine that $P^{(R)}_{m+k,j,m-k'}$ is as given in (12).

Next consider the AWGN channel with fading and CSIR. Let $\mathcal{G}(s,d)$ and $\{\beta_{m,j,m+k} \in [0,1), m \in [0,M], j \in [1,n_m], k \in [0, M-m]\}$ be as above. Then, the signals transmitted by nodes and the codewords are as in (36) and (37), respectively, except that $\{V_{m+k}\}$ and $\{\mathbf{v}_{m+k}(s_{m+k})\}$ are now all complex circularly-symmetric Gaussian. Then, as argued in [29], since the channel gains $\mathcal{H}_{m+k,j} := \{h_{(m',j'),(m+k,j)} : j' \in [1, n_{m'}], m' \in [0,M]\}$ are known to the receiving node $L_{m+k,j}$, the channel output for node $L_{m+k,j}$ is $\tilde{Y}_{m+k,j} = (Y_{m+k,j}, \mathcal{H}_{m+k,j})$. Thus, as above, we obtain that the rate R_0 in Theorem 2.1 is achievable whenever (5) holds with $Y_{m+k,j}$ replaced by $\tilde{Y}_{m+k,j}$. Now

$$\begin{aligned} &I\big(X_m; \tilde{Y}_{m+k,j} \,\big|\, X_{[m+1,M]},\, x_F\big) \\ &\quad = \;\; I\big(X_m; \mathcal{H}_{m+k,j} \,\big|\, X_{[m+1,M]},\, x_F\big) \\ &\qquad\quad + \; I\big(X_m; Y_{m+k,j} \,\big|\, X_{[m+1,M]},\, x_F, \mathcal{H}_{m+k,j}\big) \\ &\quad \overset{(a)}{=} \;\; E_{\mathcal{H}_{m+k,j}}\Bigg[I\Bigg(X_m; Y_{m+k,j} \,\Bigg|\, X_{[m+1,M]},\, x_F,\, \mathcal{H}_{m+k,j} \\ &\qquad\quad = \Big\{ h_{(m',j'),(m+k,j)} : j' \in [1, n_{m'}], m' \in [0,M] \Big\} \Bigg) \Bigg] \\ &\quad \overset{(b)}{=} \;\; C\left(\frac{P^{(R)}_{m+k,j,m}}{N_{m+k,j} + \sum_{k'=1}^{m} P^{(R)}_{m+k,j,m-k'}}\right), \end{aligned}$$

where (a) uses the fact that, for $k > 0$, X_m and $\mathcal{H}_{m+k,j}$ are independent and (b) follows from (6) and (8), with $P^{(R)}_{m+k,j,m}$ and $C(\cdot)$ as given in (14) and (15), respectively.

Finally, consider the AWGN channel with fading and CSIRT. The key conceptual difference in obtaining an achievable region here as compared to the CSIR case arises due to the fact that the channel state information is now also available to the transmitting nodes. Consequently, each transmitting node can now do power allocation for aiding various levels in the flow graph depending on the current channel state. More specifically, consider a flow graph $\mathcal{G}(s,d)$ as above. Then, node $L_{m,j}$ transmits

$$X_{m,j} \;\;=\;\; \sum_{k=0}^{M-m} \gamma_{m,j,m+k}(\mathcal{H})\; V_{m+k},$$

where the complex gains $\{\gamma_{m,j,m+k}(\mathcal{H}) \in \mathbb{C},\, k \in [0, M-m]\}$ satisfy (16) in order to meet the power constraint at $L_{m,j}$, and $\{V_m : m \in [0,M]\}$, as above, are i.i.d. with $V_m \sim \mathcal{N}_{\mathbb{C}}(0,1)$. The rest of the argument is now similar to the CSIR case. ■

APPENDIX IV
PROOF OF THEOREM 4.1

The basic idea for showing that a rate vector $(R_1, R_2, \ldots, R_U)$ is achievable in $\mathcal{N}$ is as follows. As in the single s-d pair case, block Markov encoding is used. A codebook for each node j is generated by applying the random-codebook generation procedure of Appendix III to $\mathcal{G}(s_u, d_u)$, for each $u \in \mathcal{U}_j$. Now, in each block i, node j transmits a weighted sum over $u \in \mathcal{U}_j$ of the codewords it would have transmitted if (s_u, d_u) were the only source-destination pair in $\mathcal{N}$, with the weights chosen so as not to violate the constraint on the total power available for transmission to node j. Furthermore, a "scheduling" random variable Q is introduced, which allows nodes to allocate different fractions of their available transmission power at different transmissions to various levels of the flow graphs that they belong. At the end of the transmission of each block i, each node performs *successive cancellation* to iteratively decode indices for the flow graphs it belongs.

We now expand on the above. Given $(R_1, R_2, \dots, R_U)$, let, for each $u \in [1, U]$, $R_{u,0} := R_u$ and $(R_{u,1}, R_{u,2}, \dots, R_{u,M_u})$ be such that $R_u \geq R_{u,1} \geq \cdots \geq R_{u,M_u} \geq 0$. Also, given the probability distribution $\{p(q) : q \in \mathcal{Q}\}$, generate a b-sequence $(q_1, q_2, \dots, q_b)$, where $q_l \in \mathcal{Q}, l \in [1, b]$, are identically and independently drawn according to $\{p(q)\}$.

Codebook generation: The codebook for each node j in $\mathcal{N}$ is generated as follows. For each $q \in \mathcal{Q}$ that the "scheduling" random variable Q can take, each node j allocates a fraction $\beta_{j,u,k}^{(q)}$ of its available power P_j to aid the transmission of level k in $\mathcal{G}(s_u, d_u)$, for each $u \in \mathcal{U}_j$, $k \in [m_j(u), M_u]$. Now, given $\{R_{u,k},\, k \in [0, M_u], u \in [1, U]\}$, generate i.i.d. Gaussian random variables $v_{u,k,l}(s), l \in [1, b], s \in [1, 2^{bR_{u,k}}]$, having mean zero and variance one. Then the codebook for node j, $\mathcal{C}_j$, is given by $\{\mathbf{x}_j(\{s_{u,k} : u \in \mathcal{U}_j, k \in [m_j(u), M_u]\}) = (\, x_{j,l}(\{s_{u,k} : u \in \mathcal{U}_j, k \in [m_j(u), M_u]\})\,)_{l \in [1,b]} : s_{u,k} \in [1, 2^{bR_{u,k}}]\}$, where

$$x_{j,l}\Big(\{s_{u,k} : u \in \mathcal{U}_j,\, k \in [m_j(u),\, M_u]\}\Big) = \sum_{u \in \mathcal{U}_j} \sum_{k=m_j(u)}^{M_u} \gamma_{j,u,k}^{(q_l)}\, v_{u,k,l}(s_{u,k}).$$

Above $\gamma_{j,u,k}^{(q)} = \sqrt{\beta_{j,u,k}^{(q)} P_j}$ for the AWGN channel without fading or with fading and CSIR, and $E_{\mathcal{H}}[|\gamma_{j,u,k}^{(q)}(\mathcal{H})|^2] = \beta_{j,u,k}^{(q)} P_j$ for the AWGN channel with fading and CSIRT. The codewords so generated satisfy the power constraint of node j with high probability when

$$\sum_{q \in \mathcal{Q}} p(q) \sum_{u \in \mathcal{U}_j} \sum_{k=m_j(u)}^{M_u} \beta_{j,u,k}^{(q)} \quad < \quad 1$$

and b is large enough.

Random partitions: For each $u \in [1, U]$, obtain nested random partitions $\{\mathcal{S}^{u,k} : k \in [0, M_u]\}$ of $[1, 2^{bR_{u,0}}]$, using the random-partitioning procedure given in Appendix II.

Encoding: In b-block i, each node j transmits the codeword $\mathbf{x}_j(\{\hat{s}_{u,k;i}^j : k \in [m_j(u), M_u], u \in \mathcal{U}_j\})$, where, for each $u \in \mathcal{U}_j$, forward estimates $\{\hat{s}_{u,k;i}^j : k \in [m_j(u), M_u]\}$ are obtained by applying the encoding procedure given in Appendix II to $\mathcal{G}(s_u, d_u)$.

Decoding: Each node decodes using *successive cancellation* [5], [9], [30]. Thus each node j defines an order in which it successively decodes the signals received from different levels of the different flow graphs that it belongs. While decoding a particular level of a particular flow graph, the signal components corresponding to already decoded levels are removed from the received signal and the signal components corresponding to levels with higher order act as noise.

More specifically, for each $q \in \mathcal{Q}$, let $\sigma_j^{(q)}(u, k) \in \{0, 1, 2, \dots\}$ denote the step in which node j decodes level k of flow graph $\mathcal{G}(s_u, d_u)$, where $\sigma_j^{(q)}(u, k) = 0$ denotes that the information transmitted by level k of flow graph $\mathcal{G}(s_u, d_u)$ is known to j without any decoding operation (for instance, $\sigma_j^{(q)}(u, k) = 0$, $k \in [m_j(u), M_u]$, $u \in \mathcal{U}_j$). A restriction on $\sigma_j^{(q)}$ is that, for each $u \in \mathcal{U}_j$, the order it assigns to decoding the signals transmitted by different levels within flow graph $\mathcal{G}(s_u, d_u)$ should correspond to the decoding order described in Appendix II, i.e., for each $u \in \mathcal{U}_j$, $\sigma_j^{(q)}(u, m_j(u) - 1) < \sigma_j^{(q)}(u, m_j(u) - 2) < \cdots < \sigma_j^{(q)}(u, 0)$. Then, from straightforward extensions of the probability of error calculations for the single source-destination pair case given in Appendix II and the U-user multiple access channel in Section 14.3 of [9], we deduce that each node j will be able to decode all the required indices with arbitrarily small probability of error, if, for each $u \in \mathcal{U}_j$, $\{R_{u,k} : k \in [0, m_j(u) - 1]\}$ satisfy

$$R_{u,k} \quad < \quad \sum_{q \in \mathcal{Q}} p(q)\, C\Bigg(\frac{P_{j,u,k}^{(R)(q)}}{N_j + \sum\limits_{\substack{\{(u',k') : \sigma_j^{(q)}(u',k') \\ > \sigma_j^{(q)}(u,k)\}}} P_{j,u',k'}^{(R)(q)}}\Bigg),$$

for $k = m_j(u) - 1$, and

$$R_{u,k} \,<\, R_{u,k+1} + \sum_{q \in \mathcal{Q}} p(q) C\Bigg(\frac{P_{j,u,k}^{(R)(q)}}{N_j + \sum\limits_{\substack{\{(u',k') : \sigma_j^{(q)}(u',k') \\ > \sigma_j^{(q)}(u,k)\}}} P_{j,u',k'}^{(R)(q)}}\Bigg),$$

for each $k \in [0, m_j(u) - 2]$, where $P_{j,u,k}^{(R)(q)}$ is the total power received by node j corresponding to the signal component transmitted by level k of flow graph $\mathcal{G}(s_u, d_u)$ for when the scheduling random variable Q takes value q, and it is given by

$$P_{j,u,k}^{(R)(q)} \quad = \quad \Bigg| \sum_{\substack{j' \neq j \\ j' \in [1,n]}} \gamma_{j',u,k}^{(q)}\, h_{j',j} \Bigg|^2.$$

Finally, by employing time-sharing over the rate vectors $(R_1, R_2, \dots, R_U)$ that are achievable using the above encoding-decoding procedure, we obtain a convex achievable rate region. This completes the proof of Theorem 4.1. ■

ACKNOWLEDGMENT

The authors are grateful to the Associate Editor and the anonymous reviewers for their many helpful comments. They also wish to thank Bert Hochwald and Gerhard Kramer for some insightful discussions.

REFERENCES

[1] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Intl. Symp. Information Theory*, pp. 23-52, Prague, 1971.

[2] R. Ahlswede, N. Cai, S. R. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 4, pp. 1204–1216, July 2000.

[3] I. A. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey of sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–115, Aug. 2002.

[4] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 1, pp. 197–207, Jan. 1973.

[5] P. Bergmans and T. M. Cover, "Cooperative broadcasting," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 3, pp. 317–324, May 1974.

[6] E. Biglieri, J. Proakis, S. Shamai (Shitz), “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. Inform. Theory*, vol. IT-44, no. 6, pp. 2619–2692, Oct. 1998.

[7] T. M. Cover, “Comments on broadcast channels,” *IEEE Trans. Inform. Theory*, vol. IT-44, no. 6, pp. 2524–2530, Oct. 1998.

[8] T. M. Cover and A. El Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inform. Theory*, vol. IT-25, no. 5, pp. 572–584, Sept. 1979.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[10] F. H. Eskali, K. F. Petty, and P. P. Variaya, “Dynamic channel allocation for vehicle-to-vehicle communication in automated highway systems,” in *Proc. IEEE Conf. Intelligent Transportation System*, 1997, pp. 58–63.

[11] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[12] G. J. Foschini and M. J. Gans, “On limits of wireless communications in a fading environment when using multiple antennas,” *Wireless Personal Communications*, vol. 6, no. 3, p. 311, March 1998.

[13] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[14] R. G. Gallager, “Capacity and coding for degraded broadcast channels,” *Problemy Peredaci Informaccii*, 10(3): 3–14, 1974.

[15] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. New York: Chapman & Hall/CRC, 1999.

[16] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Trans. Inform. Theory,* vol. IT-46, no. 2, pp. 388–404, March 2000.

[17] J. C. Haartsen, “The Bluetooth radio system,” *IEEE Pers. Commun.*, vol. 7, no. 1, pp. 28–36, Feb. 2000.

[18] B. Hochwald and T. L. Marzetta, “Capacity of a mobile multiple-antenna communication link in a Rayleigh flat-fading environment,” *IEEE Trans. Inform. Theory,* vol. IT-45, no. 1, pp. 139–157, Jan. 1999.

[19] B. Hochwald and S. Vishwanath, “Space-time multiple access: linear growth in the sum-rate,” in *Proc. 40th Annual Allerton Conf. Communications, Control, & Computing*, Allerton, IL, Oct. 2002.

[20] IEEE Computer Society LAN MAN Standards Committee, “Wireless LAN medium access control (MAC) and physical layer (PHY) specifications.” IEEE Standard 802.11-1999. The Institute of Electrical and Electronics Engineers, New York, NY, 1999.

[21] H. Liao, “Multiple access channels,” Ph.D. Thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[22] E. C. van der Meulen, “Three-terminal communication channels,” *Advanced Applied Probability*, vol. 3, pp. 120–154, 1971.

[23] E. C. van der Meulen, “Some reflections on the interference channel,” in *Communications and Cryptography: Two Sides of One Tapestry*, R. E. Blahut, D. J. Costello, U. Maurer, and T. Mittelholzer, Eds. Boston, MA: Kluwer, 1994, pp. 409-421.

[24] K. J. Negus, A. P. Stephens, and J. Lansford, “HomeRF: Wireless networking for the connected home,” *IEEE Pers. Commun.*, vol. 7, no. 1, pp. 20–27, Feb. 2000.

[25] B. Schein and R. Gallager, “The Gaussian parallel relay network,” in *Proc. IEEE Intl. Symp. Information Theory*, Sorrento, Italy, June 25–30, 2000.

[26] S. Shamai (Shitz) and T. L. Marzetta, “Multiuser capacity in block fading with no channel state information,” *IEEE Trans. Inform. Theory*, vol. IT-48, no. 4, pp. 938–942, April 2002.

[27] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, no. 4, pp. 471–480, July 1973.

[28] G. L. Stüber, *Principles of Mobile Communication*. Norwell, MA: Kluwer Academic, 2001.

[29] I. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585-595, 1999.

[30] A. Wyner, “Recent results in Shannon theory,” *IEEE Trans. Inform. Theory*, vol. IT-20, no. 1, pp. 2–10, Jan. 1974.