An Achievable Rate for the Multiple-Level Relay Channel

Liang-Liang Xie, *Member, IEEE,* and P. R. Kumar, *Fellow, IEEE*

***Abstract*—For the multiple-level relay channel, an achievable rate formula, and a simple coding scheme to achieve it, are presented. Generally, higher rates can be achieved with this coding scheme in the multiple-level relay case than previously known. For a class of degraded channels, this achievable rate is shown to be the exact capacity. An application of the coding scheme to the allcast problem is also discussed.**

***Index Terms*—Channel with feedback, degraded channel, multiple-relay channel, multiuser information theory, network information theory.**

I. INTRODUCTION

THE relay channel was introduced by van der Meulen [1], [2]. The simplest case, shown in Fig. 1, is the three-node scenario where node 1 functions purely as a relay to help the information transmission from node 0 to node 2. An immediate application of this framework, for instance, is in wireless communications, where a node is placed between the source node and the destination node, in order to shorten the distance of a hop, which has implications in terms of the amount of traffic carried, interference, power consumption, etc. In [1], a special discrete memoryless relay channel is even constructed for which no reliable information transmission is possible without the help of the relay node.

The simplest discrete memoryless one-relay channel is depicted in Fig. 1, where nodes 0, 1, and 2 are the source, the relay, and the destination, respectively. This channel can be denoted by $(\mathcal{X}_0 \times \mathcal{X}_1, p(y_1, y_2|x_0, x_1), \mathcal{Y}_1 \times \mathcal{Y}_2)$, where $\mathcal{X}_0$, $\mathcal{X}_1$ are the transmitter alphabets of nodes 0 and 1, respectively, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are the receiver alphabets of nodes 1 and 2, respectively, and a collection of probability distributions $p(\cdot, \cdot|x_0, x_1)$ on $\mathcal{Y}_1 \times \mathcal{Y}_2$, one for each $(x_0, x_1) \in \mathcal{X}_0 \times \mathcal{X}_1$. The interpretation is that x_0 is the input to the channel from the source node 0, y_2 is the output of the channel to the destination node 2, and y_1 is the output received by the relay node 1. After processing y_1, the relay node 1 sends the input x_1 chosen as a function of its past parameters

$$x_1(t) = f_t(y_1(t-1), y_1(t-2), \ldots), \qquad \text{for every } t \tag{1}$$

where $f_t(\cdot)$ can be any causal function. Note that a one-step time delay is assumed in (1) to account for the signal processing time by the relay.

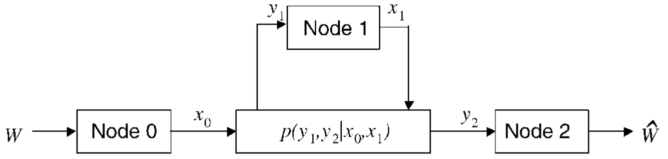

Fig. 1. The one-relay channel.

For the one-relay channel described above, to date, the highest achievable rates proved are still those obtained in [3], where one most remarkable conclusion is that the following rate is achievable:

$$R < \max_{p(x_0,x_1)} \min\{I(X_0;Y_1|X_1), I(X_0,X_1;Y_2)\}. \tag{2}$$

It is worth noting that the coding scheme to achieve the above rate is *not* simply multihop. At first, information goes from the source to the relay, and then from the relay to the destination. However, the destination needs to take into account both the inputs by the source and the relay in order to achieve (2). One can imagine that in the wireless relay channel example mentioned earlier, the destination node can make use of the signal coming directly from the source node, even though it may be relatively weaker than that from the relay node due perhaps to its greater distance from the source.

Moreover, [3] also proved that if $X_0 \to (X_1, Y_1) \to Y_2$ forms a Markov chain, i.e.,

$$p(y_2|y_1, x_0, x_1) = p(y_2|y_1, x_1) \tag{3}$$

then the right-hand side (RHS) of (2) is the capacity of this physically degraded relay channel. However, the capacity for the general nondegraded case is still unknown. A recent study [4] of the so-called Gaussian parallel relay channel showed an interesting result: The asymptotically optimal coding scheme dramatically depends on the relative locations of the nodes, or equivalently, on the relative amplitudes of their signal-to-noise ratios (SNRs). To some extent, this result excludes the possibility of the existence of a unifying optimal coding scheme.

Manuscript received December 1, 2003; revised November 28, 2004. This work was supported in part by USARO under Contracts DAAD19-00-1-0466 and DAAD19-01010-465, by DARPA under Contracts N00014-01-1-0576 and F33615-01-C-1905, by AFOSR under Contract F49620-02-1-0217, by DARPA/AFOSR under Contract F49620-02-1-0325, and by the National Science Foundation under Contracts NSF ANI 02-21357 and CCR-0325716. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004. This work was conducted while L.-L. Xie was visiting the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign.

L.-L. Xie is with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100080, China (e-mail: xie@iss.ac.cn).

P. R. Kumar is with the University of Illinois at Urbana-Champaign, Coordinated Science Laboratory, Urbana, IL 61801-2307, USA (e-mail: prkumar@uiuc.edu).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.844066

0018-9448/$20.00 © 2005 IEEE

Up to now, at least three coding schemes have been developed that are capable of achieving (2). (The most recent survey on relay channels can be found in [5].) The original coding scheme designed in [3] uses several complex techniques: block Markov superposition encoding, random partitioning (binning), and successive decoding. This scheme even uses codebooks of different sizes. Later on, a much simpler coding scheme was developed by Carleial [6] in the study of multiple-access channel with generalized feedback (MAC-GF), which includes the one-relay channel as a special case. This new scheme still uses block Markov superposition encoding, but avoids random partitioning, and all codebooks are of the same size. The key new idea lies in the decoding: Unlike the sequential manner in [3], it is a *simultaneous* typicality check of two consecutive blocks. But the paper [6] itself did not point out that this was a new scheme for achieving (2). The third scheme achieving (2) is the backward decoding introduced in [7]. When MAC-GF is concerned, the backward decoding is more powerful in achieving higher rates than Carleial's scheme as was shown in [8]. But for relay channels, they achieve the same rates. Moreover, since the backward decoding starts the decoding process only after all the blocks have been received, it incurs a substantial decoding delay.[1]

Actually, in wireless networks there can be more than one relay node. For instance, the Gaussian parallel relay channel considered in [4] consists of two-relay nodes. The general framework would be that there are multiple levels of relays and each level consists of one or more nodes. This general multiple-level relay channel was studied in [9], where the coding scheme of [3] was extended, and an achievable rate formula in a recursive constraint form was proved. The same coding scheme was applied to a special physically degraded Gaussian multiple-relay channel in [10], where a specific formula was obtained and was shown to achieve the capacity for that channel.

In [11], we proposed a new coding scheme for the Gaussian multiple-level relay channel and obtained a new achievable rate formula. Although it coincides with [3] in giving the same achievable rate formula for the one-relay case, it is easier to extend to the multiple-level relay case and generally achieves higher rates than those proved in [9]. It was discovered later that this scheme is similar to Carleial's scheme. In the decoding, they both employ *simultaneous* typicality check of multiple blocks. But the encoding part of our scheme in [11] is substantially simpler due to the special character of the Gaussian framework.

In the current paper, we present the corresponding results for the discrete memoryless case. Without the additive property of Gaussian channels, the scheme we develop here is more complex and is essentially an extension of Carleial's scheme to a multistage format. The new achievable rate formula proved is neat (see (9)) and seems a natural extension of (2) to the multiple-relay case. Also, it is generally higher than that proved in [9].

The advantages of the new coding scheme of [11] for the multiple-level relay channels have also been recognized in [12], where the corresponding achievable rate formula for the discrete memoryless case is also stated. The paper [12] also goes on to obtain the capacity of some relay channels under fading, which is the first significant capacity result for such channels, and one which may possibly constitute a breakthrough in the field.

The coding scheme for the discrete memoryless case presented in this paper follows in a similar style to that of the Gaussian case presented in [11], although it is more complex in the construction of codebooks as previously noted. Actually, as is well known ([13, Ch. 7]), we can always use the coding scheme presented here for the Gaussian case. But due to the special character of the Gaussian framework, a simpler coding scheme could be chosen as was done in [11]. For example, for a Gaussian channel with $M-1$ relays, we only need generate M^2 random matrices of size $2^{TR} \times T$ for codebooks as shown in [11]; but here for the discrete memoryless case, $M\sum_{i=0}^{M-1}(2^{TR})^i$ random matrices of the same size are needed.

Another feature in our coding scheme worth emphasizing is that for the $(M-1)$-level relay channel, the codebooks of any M consecutive blocks should be independent of each other, so that in the decoding process when the simultaneous typicality check of M consecutive blocks is carried out, the decoding errors arising from different blocks are independent of each other. The necessity for this independence can be better understood with the following observation. Consider the following two additive white Gaussian noise (AWGN) channels with the same capacity $C = \frac{1}{2}\log(1+P/N)$:

$$\begin{cases} Y_1 = X_1 + Z_1 \\ Y_2 = X_2 + Z_2 \end{cases}$$

where inputs X_1 and X_2 are of the same power constraint P, and noise Z_1 and Z_2 are of the same variance N. The efficient usage of these two channels together is to let them transmit independently and then combine the rates afterward. In this way, we can achieve any rate up to $2C = \log(1+P/N)$. But if we use the same codebook for them and set $X_1 = X_2$, the best we can achieve is only up to $\frac{1}{2}\log(1+2P/N) < 2C$.

In this paper, we focus our discussion on discrete memoryless channels. Besides proving a new achievable rate formula, we also discuss relay systems with feedback and degraded versions of these systems. Some corresponding results for the Gaussian case have been proved in [11, Theorems 3.11–3.12], which include the formula in [10] as a special case.

II. Models of Multiple-Level Relay Channels

We begin with a definition of the discrete memoryless multiple-level relay channel. Consider a channel with $M+1$ nodes. Let the source node be denoted by 0, the destination node by M, and let the other $M-1$ nodes be denoted sequentially as $1, 2, \ldots, M-1$ in arbitrary order. Assume each node $i \in \{0, 1, \ldots, M-1\}$ sends $x_i(t) \in \mathcal{X}_i$ at time t, and each node $k \in \{1, 2, \ldots, M\}$ receives $y_k(t) \in \mathcal{Y}_k$ at time t, where the finite sets $\mathcal{X}_i$ and $\mathcal{Y}_k$ are the corresponding input and output alphabets for the corresponding nodes. The channel dynamics is

[1] We are grateful to G. Kramer for keeping us abreast of a thorough historical search on coding schemes for relay channels, leading to the above categorization of three coding schemes including Carleial's.

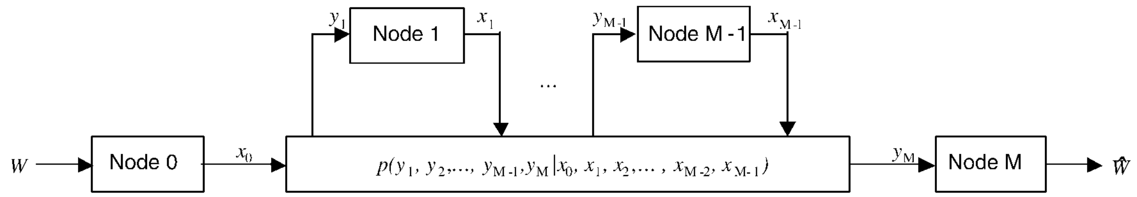

Fig. 2. The multiple-level relay channel.

described by the following conditional probability mass function:

$$p(y_1, y_2, \ldots, y_M|x_0, x_1, \ldots, x_{M-1}) \tag{4}$$

for all

$$(x_0, \ldots, x_{M-1}) \in \mathcal{X}_0 \times \cdots \times \mathcal{X}_{M-1}$$

and

$$(y_1, \ldots, y_M) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_M.$$

Finally, we assume a one-step time delay at every relay node to account for the signal processing time, so that for all $i \in \{1, 2, \ldots, M-1\}$

$$x_i(t) = f_{i,t}(y_i(t-1), y_i(t-2), \ldots), \qquad \text{for all } t$$

where $f_{i,t}$ can be any causal function. Fig. 2 depicts this scenario.

Remark 2.1: Note that for simplicity in the above channel formulation we have not allowed any output y_0 from the channel to the source node 0, or any input x_M to the channel from the destination node M. The involvement of y_0 and x_M is equivalent to allowing feedback in the channel since it allows the modeling of systems where all nodes can both hear as well as transmit. Both these seemingly more general formulations are, however, actually covered by our formulation where node 0 cannot hear, and node M cannot transmit. To see this, simply consider a situation where node $M-1$ serves as a surrogate for node M in that node M simply hears $y_M = (x_{M-1}, y_{M-1})$. Then nodes $M-1$ and M have exactly the same information. Thus, $M-1$ serves as a transmission-capable surrogate for node M, which is simply a dummy node. In the same vein, suppose that node 1 is a surrogate for node 0 in that node 0 can transmit W directly (or close to that through a high-bandwidth link) which only node 1 can hear, but none of the other nodes. Then node 1 serves as listening-capable surrogate for node 0, which is the only outlet that node 0 has to the rest of the nodes, and node 0 simply becomes a dummy. Thus, the system with nodes $\{0, 1, 2, \ldots, M-2, M-1, M\}$ where node 0 cannot hear and node M cannot transmit, is simply the same as the system with nodes $\{1, 2, \ldots, M-2, M-1\}$, where node 1 is the source which is also hearing capable, and node $M-1$ is the destination which is transmission capable.

The following definitions of codes and achievable rates are standard.

Definition 2.1: A $(2^{TR}, T, \lambda_T)$ code for a discrete memoryless multiple-level relay channel consists of the following.

1) A random variable W with $P(W = k) = \frac{1}{2^{TR}}$, for every $k \in \{1, 2, \ldots, 2^{TR}\}$.

2) An encoding function $F_0 : \{1, 2, \ldots, 2^{TR}\} \to \mathcal{X}_0^T$ for the source node 0, and relay functions $f_{i,t} : \mathcal{Y}_i^{t-1} \to \mathcal{X}_i$, $t = 2, \ldots, T$ for all the relay nodes $i \in \{1, 2, \ldots, M-1\}$, such that

$$\boldsymbol{x}_0 := [x_0(1), x_0(2), \ldots, x_0(T)] = F_0(W) \tag{5}$$

$$x_i(t) = f_{i,t}(y_i(t-1), \ldots, y_i(1)), \quad \text{with } x_i(1) \text{ any element in } \mathcal{X}_i. \tag{6}$$

3) A decoding function $g : \mathcal{Y}_M^T \to \{1, 2, \ldots, 2^{TR}\}$ for the destination node M.

4) The maximal probability of error

$$\lambda_T := \max_{k \in \{1,2,\ldots,2^{TR}\}} \text{Prob}\{g(\boldsymbol{y}_M) \neq k | W = k\} \tag{7}$$

where $\boldsymbol{y}_M := [y_M(1), y_M(2), \ldots, y_M(T)]$.

Definition 2.2: A rate $R > 0$ is said to be *achievable* if there exists a sequence of $(2^{TR}, T, \lambda_T)$ codes such that the maximal probability of error λ_T tends to zero as $T \to \infty$.

We will also consider the *degraded* version of our channel.

Definition 2.3: A discrete memoryless multiple-level relay channel is said to be *degraded* if

$$\begin{aligned} &p(y_{k+1}, \ldots, y_M | y_k, x_0, \ldots, x_{k-1}, x_k, \ldots, x_{M-1}) \\ &\quad = p(y_{k+1}, \ldots, y_M | y_k, x_k, \ldots, x_{M-1}), \\ &\qquad\qquad \text{for } k = 1, \ldots, M-1. \end{aligned} \tag{8}$$

Equivalently, (8) means that

$$(X_0, \ldots, X_{k-1}) \to (Y_k, X_k, \ldots, X_{M-1}) \to (Y_{k+1}, \ldots, Y_M)$$

forms a Markov chain for every $k = 1, \ldots, M-1$. In the case of $M = 2$, (8) reduces to (3).

Remark 2.2: If node M can transmit, then we simply extend the conditioning on both sides of (8) to include x_M. We note then that a degraded system remains degraded under the embedding procedure of Remark 2.1, and conversely. So our formulation allows the treatment of degraded systems with feedback.

III. Main Results

We state the main theorem of this paper.

Theorem 3.1: For the discrete memoryless multiple-level relay channel defined above, the following rate is achievable:

$$R < \max_{p(x_0,\ldots,x_{M-1})} \min_{1 \le k \le M} I(X_0, \ldots, X_{k-1}; Y_k | X_k, \ldots, X_{M-1}). \tag{9}$$

Remark 3.1: Generally, (9) achieves higher rates than the results given in [9]. To see this, first consider the two-level relay channel example ($M = 3$) given in [9], which shows that a rate

R_0 is achievable if there exist R_1, R_2, and some $p(x_0, x_1, x_2)$ such that

$$R_0 < I(X_0; Y_1|X_1, X_2)$$

and

$$R_1 < I(X_1; Y_2|X_2)$$
$$R_0 < I(X_0; Y_2|X_1, X_2) + R_1$$

and

$$R_2 < I(X_2; Y_3)$$
$$R_1 < I(X_1; Y_3|X_2) + R_2$$
$$R_0 < I(X_0; Y_3|X_1, X_2) + R_1.$$

It is easy to check that this will reduce to (9) if and only if $I(X_1; Y_2|X_2) = I(X_1, X_2; Y_3)$, which generally does not hold. The same reasoning applies to the general case in [9, Theorem 2.1].

For the degraded discrete memoryless multiple level relay channel, the following theorem shows that the RHS of (9) is actually the capacity.

Theorem 3.2: The capacity of the degraded discrete memoryless multiple-level relay channel is

$$C = \max_{p(x_0,\dots,x_{M-1})} \min_{1\le k\le M} I(X_0, \dots, X_{k-1}; Y_k|X_k, \dots, X_{M-1}). \quad (10)$$

Remark 3.2: In the case of one relay, i.e., $M = 2$, Theorem 3.2 reduces to [3, Theorem 1].

To illustrate the case of feedback, we now turn to the more general setting where the destination node M also has an input $x_M \in \mathcal{X}_M$ to the channel. Then it follows immediately from Theorem 3.1 that the following rate is achievable:

$$R < \max_{x_M\in\mathcal{X}_M} \max_{p(x_0,\dots,x_{M-1})} \min_{1\le k\le M} I(X_0, \dots, X_{k-1}; Y_k|X_k, \dots, X_{M-1}, x_M). \quad (11)$$

However, we can achieve higher rates than the RHS of (11) as stated in the following theorem.

Theorem 3.3: For the discrete memoryless multiple-level relay channel defined above with an additional input $x_M \in \mathcal{X}_M$ from the destination node M, the following rate is achievable:

$$R < \max_{p(x_0,\dots,x_M)} \min_{1\le k\le M} I(X_0, \dots, X_{k-1}; Y_k|X_k, \dots, X_M). \quad (12)$$

Remark 3.3: Noting the embedding procedure of Remark 2.1, and the preservation of the degraded property under embedding as noted in Remark 2.2, it follows that (12) also achieves the capacity for the degraded case where node M can also transmit, i.e., x_M also exists.

TABLE I
THE CHANNEL DYNAMICS $p(y_1, y_2|x_0, x_1, x_2)$ OF THE ONE-RELAY CHANNEL IN EXAMPLE 1

$(x_0, x_1, x_2)\backslash(y_1, y_2)$	(0,0)	(0,1)	(1,0)	(1,1)
(0,0,0)	1	0	0	0
(0,1,0)	0	1	0	0
(1,0,0)	1	0	0	0
(1,1,0)	0	1	0	0
(0,0,1)	1	0	0	0
(0,1,1)	1	0	0	0
(1,0,1)	0	0	1	0
(1,1,1)	0	0	1	0

Next, we show that (12) generally achieves larger rates than (11). First, it is obvious that (12) is at least as large as (11), since we can always choose $\text{Prob}(X_M = x_M) = 1$ for any $x_M \in \mathcal{X}_M$. Second, the following example shows that the RHS of (12) can indeed be larger.

Example 1: Consider a simple channel with $M = 2$, and

$$\mathcal{X}_0 = \mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_1 = \mathcal{Y}_2 = \{0, 1\}.$$

The corresponding $p(y_1, y_2|x_0, x_1, x_2)$'s are shown in Table I.

It is easy to check that

$$\text{Prob}(Y_1 = 0|x_2 = 0) = 1 \quad \text{and} \quad \text{Prob}(Y_2 = 0|x_2 = 1) = 1.$$

Hence, $I(X_0; Y_1|X_1, x_2 = 0) = 0$ for every $p(x_0, x_1)$, and $I(X_0, X_1; Y_2|x_2 = 1) = 0$ for every $p(x_0, x_1)$. Therefore,

$$\max_{x_2\in\{0,1\}} \max_{p(x_0,x_1)} \min\{I(X_0; Y_1|X_1, x_2), I(X_0, X_1; Y_2|x_2)\} = 0. \quad (13)$$

Moreover, it is easy to check that $\text{Prob}(Y_1 = X_0|x_2 = 1) = 1$ and $\text{Prob}(Y_2 = X_1|x_2 = 0) = 1$. Hence,

$$\begin{aligned} I(X_0; Y_1|X_1, X_2) &= I(X_0; Y_1|X_1, x_2 = 0)\text{Prob}(X_2 = 0) \\ &\quad + I(X_0; Y_1|X_1, x_2 = 1)\text{Prob}(X_2 = 1) \\ &= H(X_0|X_1, x_2 = 1)\text{Prob}(X_2 = 1) \\ I(X_0, X_1; Y_2|X_2) &= I(X_0, X_1; Y_2|x_2 = 0)\text{Prob}(X_2 = 0) \\ &\quad + I(X_0, X_1; Y_2|x_2 = 1)\text{Prob}(X_2 = 1) \\ &= H(X_1|x_2 = 0)\text{Prob}(X_2 = 0). \end{aligned}$$

Therefore, we get (14) at the bottom of the page, where the equality in (14) can be achieved by letting

$$p(x_0, x_1, x_2) = p(x_0)p(x_1)p(x_2)$$

and

$$\text{Prob}(X_0 = 0) = \text{Prob}(X_1 = 0) = \frac{1}{2}.$$

Finally, from (13) and (14), (12) is seen to be strictly larger than (11) in this example. □

An intuitive interpretation of (9) or (12) comes directly from the coding scheme used in the achievability proof of

$$\begin{aligned} &\max_{p(x_0,x_1,x_2)} \min\{I(X_0; Y_1|X_1, X_2), I(X_0, X_1; Y_2|X_2)\} \\ &\quad= \max_{p(x_0,x_1,x_2)} \min\big\{H(X_0|X_1, x_2 = 1)\text{Prob}(X_2 = 1), H(X_1|x_2 = 0)\text{Prob}(X_2 = 0)\big\} \\ &\quad\le \max_{p(x_2)} \min\{\text{Prob}(X_2 = 1), \text{Prob}(X_2 = 0)\} = \frac{1}{2} \end{aligned} \quad (14)$$

Theorem 3.1. We can imagine that there is an information flow from the source node 0 to the destination node M along the path $0 \to 1 \to \cdots \to M$. Each node i decodes the information one-step (actually one time block in the coding) before the next node $i+1$. Hence, by the time the information reaches node k, all the upstream nodes (i.e., nodes with smaller index than k) have already obtained the same information and can therefore cooperate. This results in the achievability of any rate

$$R < I(X_0, \ldots, X_{k-1}; Y_k | X_k, \ldots, X_M)$$

where the conditioning is due to the same reason: the downstream nodes (with larger index than k) get no more information than node k, and therefore their inputs are predictable by node k.

Since the order of the nodes, except the source, can be arbitrarily chosen, by the above interpretation of the coding scheme, it follows that to increase (9) one should assign a smaller index to nodes with better "receiving capability." (Note that it is even not necessary to set the destination to be node M.) However, generally, receiving capability is not easily comparable (unless in the degraded case). Of course, we can always try all the permutations to maximize (9) or (12).

What makes the maximization problem even more complicated is the following possibility. We can arrange the nodes into groups, with each group consisting of one or more nodes. The information flow then is along a path formed out of the groups, but in any one group all the nodes have the same level (i.e., they decode the information at the same time). Put into mathematical terms, we have the following theorem.

Theorem 3.4: For a discrete memoryless multiple-level relay channel with source node 0, destination node M, and the other nodes arranged into $M-1$ groups with each group k consisting of n_k nodes, $k = 1, \ldots, M-1$, the following rate is achievable:

$$R < \max_{p(\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)} \min_{1 \le k \le M} \min_{1 \le i \le n_k} I(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_{k-1}; Y_{k,i} | \boldsymbol{X}_k, \ldots, \boldsymbol{X}_M) \quad (15)$$

where boldface characters are used to denote vectors for each group: e.g., $\boldsymbol{x}_k := (x_{k,1}, x_{k,2}, \ldots, x_{k,n_k})$, $k = 0, 1, \ldots, M$, with $x_{k,i}$ denoting the input of the ith node in group k. Note that we have set group $0 := \{\text{node } 0\}$ and group $M := \{\text{node } M\}$.

For an application of this group relaying in the Gaussian channel case, we refer the reader to [11, Theorem 3.12], where it is used to study wireless networks under low attenuation.

Finally, a remark on the allcast problem: Suppose the task is for the source node 0 to send the same information to all the other nodes $1, 2, \ldots, M$. As we will see in the proof of Theorem 3.1 in Section IV, this task is implicitly achieved by the following relaying scheme. The upstream nodes decode the information before the downstream nodes. Once the final node M gets the information, all the other nodes have already obtained the information. Hence, the rate (9), (12), or (15) is also an achievable rate for the allcast problem.

IV. Proof of Theorem 3.1

We will use the now standard "typical sequences" argument to prove achievability. First we summarize some basic properties of typical sequences that will be used later. For more details, see [14, Secs. 8.6 and 14.2].

Let $(Z_1, Z_2, \ldots, Z_m)$ denote a finite collection of discrete random variables with some fixed joint distribution $p(z_1, z_2, \ldots, z_m)$ for

$$(z_1, z_2, \ldots, z_m) \in \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \times \mathcal{Z}_m.$$

Definition 4.1: The set $A_\epsilon^{(T)}$ of ϵ-typical T-sequences $(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m)$ is defined by

$$A_\epsilon^{(T)}(Z_1, Z_2, \ldots, Z_m) := \left\{ (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m) : \left| -\frac{1}{n} \log \operatorname{Prob}(\boldsymbol{s}) - H(S) \right| < \epsilon, \quad \forall\, S \subseteq \{Z_1, Z_2, \ldots, Z_m\} \right\}$$

where each $\boldsymbol{z}_i = (z_{i,1}, z_{i,2}, \ldots, z_{i,T})$ is a T-vector, $i = 1, 2, \ldots, m$, and $\boldsymbol{s}$ is defined as follows: If $S = (Z_{i_1}, Z_{i_2}, \ldots, Z_{i_\ell})$, then $\boldsymbol{s} = (\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2}, \ldots, \boldsymbol{z}_{i_\ell})$ and

$$\operatorname{Prob}(\boldsymbol{s}) = \operatorname{Prob}(\boldsymbol{z}_{i_1}, \boldsymbol{z}_{i_2}, \ldots, \boldsymbol{z}_{i_\ell}) = \prod_{t=1}^{T} p(z_{i_1,t}, z_{i_2,t}, \ldots, z_{i_\ell,t}).$$

Lemma 4.1: For any $\epsilon > 0$, the following hold for sufficiently large T.

i) Let a T-sequence $(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m)$ be generated according to

$$\prod_{t=1}^{T} p(z_{1,t}, z_{2,t}, \ldots, z_{m,t}).$$

Then

$$\operatorname{Prob}\left((\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m) \in A_\epsilon^{(T)}(Z_1, Z_2, \ldots, Z_m) \right) \ge 1 - \epsilon.$$

ii) Let a T-sequence $(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m)$ be generated according to

$$\prod_{t=1}^{T} p(z_{1,t} | z_{2,t}, \ldots, z_{m-1,t}) \cdot p(z_{m,t} | z_{2,t}, \ldots, z_{m-1,t}) \cdot p(z_{2,t}, \ldots, z_{m-1,t}).$$

Then

$$\operatorname{Prob}\left((\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m) \in A_\epsilon^{(T)}(Z_1, Z_2, \ldots, Z_m) \right) < 2^{-T(I(Z_1; Z_m | Z_2, \ldots, Z_{m-1}) - 6\epsilon)}.$$

Lemma 4.1 i) and ii) follow immediately from [14, Theorems 14.2.1 and 14.2.3], respectively.

The coding scheme used here is different from that of [3]. For example, the Slepian–Wolf partitioning is no longer used. This new coding scheme is easier to extend to the multiple-relay case, and generally achieves larger rates (compared with [9]), although it coincides with [3] in the one relay case.

To make the proof easier to follow and allow a better comparison with the coding scheme in [3], first we give the proof for the one-relay case ($M=2$), which contains all the essential ideas. Then we present the straightforward extension to the general case ($M \geq 2$).

1. One-relay case ($M=2$)

We consider any fixed $p(x_0, x_1)$. Throughout the proof, we will use the following marginal or conditional probability mass functions obtained from $p(x_0, x_1)$ and $p(y_1, y_2; x_0, x_1)$:

$$p(x_1), p(x_0|x_1), p(y_1|x_0, x_1), p(y_2|x_0, x_1), p(y_2|x_1)$$

where, for simplicity, the subscripts to distinguish these p functions are omitted since the exact meaning is obvious from the arguments.

We still use a block coding argument. We consider B blocks of transmission, each of T transmission slots. A sequence of $B-1$ indices, $w(b) \in \{1, \dots, 2^{TR}\}$, $b = 1, 2, \dots, B-1$ will be sent over in TB transmission slots. (Note that as $B \to \infty$, the rate $TR(B-1)/TB$ is arbitrarily close to R for any T.)

Generation of Codebooks: The joint codebook is still generated in a backward manner. But here one significant difference is that all the codebooks are of the same length 2^{TR}. (No more 2^{TR_0} as in the proof of [3].)

1) Generate at random 2^{TR} independent and identically distributed (i.i.d.) T-sequences in $\mathcal{X}_1^T$, each drawn according to

$$\text{Prob}(\boldsymbol{x}_1) = \prod_{t=1}^{T} p(x_{1,t}).$$

Index them as $\boldsymbol{x}_1(w_1)$, $w_1 \in \{1, 2, \dots, 2^{TR}\}$. This is the random codebook for node 1.

2) For each $\boldsymbol{x}_1(w_1)$, generate 2^{TR} conditionally independent T-sequences $\boldsymbol{x}_0(w_0|w_1)$, $w_0 \in \{1, 2, \dots, 2^{TR}\}$, drawn independently according to

$$\text{Prob}(\boldsymbol{x}_0|\boldsymbol{x}_1(w_1)) = \prod_{t=1}^{T} p(x_{0,t}|x_{1,t}(w_1)).$$

This defines the joint codebook for nodes 0, 1 as

$$\mathcal{C}_0 := \{\boldsymbol{x}_0(w_0|w_1), \boldsymbol{x}_1(w_1)\}. \tag{16}$$

It is apparent from [3] that the reason for this kind of *backward* codebook generation is that the upstream nodes (with smaller index) know what the downstream nodes are going to transmit, and therefore can adjust their own transmission accordingly. The converse is not true because of the unique direction of information flow.

Repeating the above process 1)–2) *independently* once more, we generate another random codebook $\mathcal{C}_1$ similar to $\mathcal{C}_0$ in (16). We will use these two codebooks alternately as follows: In block $b = 1, \dots, B$, the codebook $\mathcal{C}_{(b \bmod 2)}$ is used. Hence, in any two consecutive blocks, codewords from different blocks are independent. This is a property we will use in the analysis of the probability of error (see (30)).

Before the transmission, the joint codebooks $\mathcal{C}_0$, $\mathcal{C}_1$ are revealed to all the nodes 0, 1, 2.

TABLE II
THE ENCODING PROCESS FOR THE ONE-RELAY CHANNEL

block 1	block 2	$\cdots$	block B
$\mathbf{x}_0(w(1)\|1)$	$\mathbf{x}_0(w(2)\|w(1))$	$\cdots$	$\mathbf{x}_0(w(B)\|w(B-1))$
$\mathbf{x}_1(1)$	$\mathbf{x}_1(\hat{w}_1(1))$	$\cdots$	$\mathbf{x}_1(\hat{w}_1(B-1))$

Encoding: At the *beginning* of each block $b \in \{1, \dots, B\}$, node 1 has an estimate (see the decoding section) $\hat{w}_1(b-1)$ of $w(b-1)$, and sends the following T-sequence from the codebook $\mathcal{C}_{(b \bmod 2)}$ in the block:

$$\vec{X}_1(b) := \boldsymbol{x}_1(\hat{w}_1(b-1)). \tag{17}$$

Also, in the same block, node 0 sends the following T-sequence from the same codebook $\mathcal{C}_{(b \bmod 2)}$:

$$\vec{X}_0(b) := \boldsymbol{x}_0(\hat{w}_0(b)|\hat{w}_0(b-1)) \tag{18}$$

where, obviously, the estimate $\hat{w}_0(b) = w(b)$ for every $b \in \{1, \dots, B\}$, since node 0 is the source. Moreover, for the synchronization of all the nodes at the initial time, we set $\hat{w}_i(b_1) = w(b_1) = 1$ for every $b_1 \leq 0$, $i \in \{0, 1, 2\}$. The encoding process is depicted in Table II.

Every node $k \in \{1, 2\}$ thus receives a T-sequence

$$\vec{Y}_k(b) := \vec{Y}_k(\vec{X}_0(b), \vec{X}_1(b)) \tag{19}$$

with probability

$$\text{Prob}\left(\vec{Y}_k(b)|\vec{X}_0(b), \vec{X}_1(b)\right) = \prod_{t=1}^{T} p(\vec{Y}_{k,t}(b)|\vec{X}_{0,t}(b), \vec{X}_{1,t}(b))$$

where $\vec{Y}_{k,t}(b)$ is the tth element of the vector $\vec{Y}_k(b)$, and similar definitions hold for $\vec{X}_{i,t}(b)$, $i = 0, 1$.

Decoding: At the *end* of each block $b \in \{1, \dots, B\}$, decodings at node 1 and node 2 happen simultaneously, but independently.

i) Node 1 declares that $\hat{w}_1(b) = w$ if w is the unique value in $\{1, \dots, 2^{TR}\}$ such that in the block b

$$\{\boldsymbol{x}_0(w|\hat{w}_1(b-1)), \boldsymbol{x}_1(\hat{w}_1(b-1)), \vec{Y}_1(b)\} \in A_\epsilon^{(T)}(X_0, X_1, Y_1). \tag{20}$$

Otherwise, if no unique w as above exists, an error is declared with $\hat{w}_1(b) = 0$.

ii) Node 2 declares that $\hat{w}_2(b-1) = w$ if w is the unique value in $\{1, \dots, 2^{TR}\}$ such that in *both* the blocks b and $b-1$

$$\{\boldsymbol{x}_1(w), \vec{Y}_2(b)\} \in A_\epsilon^{(T)}(X_1, Y_2), \text{ and} \tag{21a}$$

$$\{\boldsymbol{x}_0(w|\hat{w}_2(b-2)), \boldsymbol{x}_1(\hat{w}_2(b-2)), \vec{Y}_2(b-1)\} \in A_\epsilon^{(T)}(X_0, X_1, Y_2). \tag{21b}$$

Otherwise, if no unique w as above exists, an error is declared with $\hat{w}_2(b-1) = 0$. The above scheme implies that the decoding at node 2 has no intention to estimate $w(b)$ at the end of block b.

Analysis of Probability of Error: Denote the event that no decoding error is made in the first b blocks by

$$A_c(b) := \Big\{\hat{w}_k(b_1 - k + 1) = w(b_1 - k + 1), \text{ for all } b_1 \in \{1, \dots, b\} \text{ and } k \in \{1, 2\}\Big\}$$

and let its probability be

$$P_c(b) := \mathrm{Prob}\,(A_c(b))$$

with $P_c(0) := 1$.

Then the probability that some decoding error is made at some node $k \in \{1, 2\}$ in some block $b \in \{1, \dots, B\}$ is

$$\begin{aligned} P_e :=& \mathrm{Prob}(\hat{w}_k(b-k+1) \neq w(b-k+1), \\ &\qquad\qquad \text{for some } k \in \{1,2\}, b \in \{1,\dots,B\}) \\ =& \sum_{b=1}^{B} \mathrm{Prob}(\hat{w}_k(b-k+1) \neq w(b-k+1) \\ &\qquad \text{for some } k \in \{1,2\} | A_c(b-1)) \cdot P_c(b-1) \\ \leq& \sum_{b=1}^{B}\sum_{k=1}^{2} \mathrm{Prob}(\hat{w}_k(b-k+1) \\ &\qquad \neq w(b-k+1) | A_c(b-1)) \cdot P_c(b-1) \\ =:& \sum_{b=1}^{B}\sum_{k=1}^{2} P_{e,k}(b) \cdot P_c(b-1) \end{aligned} \tag{22}$$

where

$$P_{e,k}(b) := \mathrm{Prob}(\hat{w}_k(b-k+1) \neq w(b-k+1) | A_c(b-1)).$$

Hence, $P_{e,k}(b)$ is the probability that a decoding error happens at node k in block b, conditioned on the event that no decoding error was made in the previous $b-1$ blocks.

Next, we calculate $P_{e,k}(b)$, $k \in \{1, 2\}$. Since $A_c(b-1)$ is presumed to hold, for every node $k \in \{1, 2\}$ we have

$$\hat{w}_k(b_1 - k + 1) = w(b_1 - k + 1), \qquad \text{for } 1 \leq b_1 \leq b - 1.$$

Hence, by (17)–(19), the decoding rule (20) is equivalent to

$$\begin{aligned} \{&\boldsymbol{x}_0(w|w(b-1)), \boldsymbol{x}_1(w(b-1)), \\ &\vec{Y}_1(\boldsymbol{x}_0(w(b)|w(b-1)), \boldsymbol{x}_1(w(b-1)))\} \in A_\epsilon^{(T)}(X_0, X_1, Y_1) \end{aligned} \tag{23}$$

and the decoding rule (21a) and (21b) is equivalent to

$$\begin{aligned} \{&\boldsymbol{x}_1(w), \vec{Y}_2(\boldsymbol{x}_0(w(b)|w(b-1)), \boldsymbol{x}_1(w(b-1)))\} \\ &\in A_\epsilon^{(T)}(X_1, Y_2) \end{aligned} \tag{24a}$$

and

$$\begin{aligned} \{&\boldsymbol{x}_0(w|w(b-2)), \boldsymbol{x}_1(w(b-2)), \\ &\vec{Y}_2(\boldsymbol{x}_0(w(b-1)|w(b-2)), \boldsymbol{x}_1(w(b-2)))\} \\ &\in A_\epsilon^{(T)}(X_0, X_1, Y_2). \end{aligned} \tag{24b}$$

Now, let

$$\begin{aligned} \mathcal{W}_1(b) &:= \{w \in \{1, \dots, 2^{TR}\} : w \text{ satisfies (23)}\} \\ \mathcal{W}_{2,0}(b) &:= \{w \in \{1, \dots, 2^{TR}\} : w \text{ satisfies (24a)}\} \\ \mathcal{W}_{2,1}(b) &:= \{w \in \{1, \dots, 2^{TR}\} : w \text{ satisfies (24b)}\} \\ \mathcal{W}_2(b) &:= \mathcal{W}_{2,0}(b) \bigcap \mathcal{W}_{2,1}(b). \end{aligned}$$

Then, $P_{e,k}(b)$ is the probability that $w(b-k+1) \notin \mathcal{W}_k(b)$, or some $w' \in \mathcal{W}_k(b)$ but $w' \neq w(b-k+1)$, conditioned on the event that no decoding error was made in the previous $b-1$ blocks. Thus, we get the equation at the bottom of the page. Hence, by (22)

$$\begin{aligned} P_e \leq& \sum_{b=1}^{B}\sum_{k=1}^{2} \Big[\mathrm{Prob}\,(w(b-k+1) \notin \mathcal{W}_k(b)) \\ &+ \mathrm{Prob}\,(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1))\Big]. \end{aligned} \tag{25}$$

Now, for node 1 with (23), for T sufficiently enough, applying Lemma 4.1 i) with $(Z_1, \dots, Z_m) = (X_0, X_1, Y_1)$, we have

$$\mathrm{Prob}\,(w(b) \notin \mathcal{W}_1(b)) < \epsilon \tag{26}$$

and for any $w' \neq w(b)$, applying Lemma 4.1 ii), we have

$$\mathrm{Prob}\,(w' \in \mathcal{W}_1(b)) < 2^{-T(I(X_0;Y_1|X_1)-6\epsilon)}.$$

And also for node 2 with (24a) and (24b), for T sufficiently enough, applying Lemma 4.1, we have for $j = 0, 1$

$$\mathrm{Prob}\,(w(b-1) \notin \mathcal{W}_{2,j}(b)) < \epsilon$$

and for any $w' \neq w(b-1)$

$$\mathrm{Prob}\,(w' \in \mathcal{W}_{2,0}(b)) < 2^{-T(I(X_1;Y_2)-6\epsilon)} \tag{27}$$

$$\mathrm{Prob}\,(w' \in \mathcal{W}_{2,1}(b)) < 2^{-T(I(X_0;Y_2|X_1)-6\epsilon)}. \tag{28}$$

Hence,

$$\begin{aligned} &\mathrm{Prob}\,(w(b-1) \notin \mathcal{W}_2(b)) \\ &\qquad \leq \sum_{j=0}^{1} \mathrm{Prob}\,(w(b-1) \notin \mathcal{W}_{2,j}(b)) \leq 2\epsilon \end{aligned} \tag{29}$$

and

$$\begin{aligned} &\mathrm{Prob}\big(\text{some } w' \in \mathcal{W}_2(b) \text{ but } w' \neq w(b-1)\big) \\ &\leq \sum_{\substack{w' \in \{1,\dots,2^{TR}\} \\ w' \neq w(b-1)}} \mathrm{Prob}\,(w' \in \mathcal{W}_2(b)) \\ &= \sum_{\substack{w' \in \{1,\dots,2^{TR}\} \\ w' \neq w(b-1)}} \prod_{j=0}^{1} \mathrm{Prob}\,(w' \in \mathcal{W}_{2,j}(b)) \qquad (30) \\ &\leq (2^{TR} - 1) 2^{-T(I(X_0,X_1;Y_2)-12\epsilon)} \qquad (31) \end{aligned}$$

where (30) follows from the independence between the codebooks of any two consecutive blocks, and (31) follows from (27), (28), and the following equation:

$$I(X_0, X_1; Y_2) = I(X_1; Y_2) + I(X_0; Y_2|X_1).$$

For any R satisfying (9), by choosing ϵ small enough, we can make T large enough such that for any $\varepsilon_1 > 0$, we get (32) at the bottom of the following page.

$$\begin{aligned} P_{e,k}(b) =& \mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_k(b), \text{ or some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1) \big| A_c(b-1)) \\ \leq& \mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_k(b) \big| A_c(b-1)) + \mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1) \big| A_c(b-1)) \\ \leq& \frac{\mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_k(b))}{P_c(b-1)} + \frac{\mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1))}{P_c(b-1)}. \end{aligned}$$

TABLE III
THE ENCODING PROCESS FOR THE TWO-RELAY CHANNEL

block 1	block 2	block 3	$\cdots$	block B
$\mathbf{x}_0\big(w(1)\vert 1,1\big)$	$\mathbf{x}_0\big(w(2)\vert w(1),1\big)$	$\mathbf{x}_0\big(w(3)\vert w(2),w(1)\big)$	$\cdots$	$\mathbf{x}_0\big(w(B)\vert w(B-1),w(B-2)\big)$
$\mathbf{x}_1\,(1\vert 1)$	$\mathbf{x}_1\big(\hat{w}_1(1)\vert 1\big)$	$\mathbf{x}_1\big(\hat{w}_1(2)\vert\hat{w}_1(1)\big)$	$\cdots$	$\mathbf{x}_1\big(\hat{w}_1(B-1)\vert\hat{w}_1(B-2)\big)$
$\mathbf{x}_2(1)$	$\mathbf{x}_2(1)$	$\mathbf{x}_2(\hat{w}_2(1))$	$\cdots$	$\mathbf{x}_2(\hat{w}_2(B-2))$

Hence, by (25), (26), (29), and (32),

$$\begin{aligned}P_e \le& \sum_{b=1}^{B}(3\epsilon + 2\varepsilon_1)\\ \le& 3B\epsilon + 2B\varepsilon_1\end{aligned}$$

which can be made arbitrarily small by letting $T \to \infty$.

Finally, the argument on choosing one good codebook from many random codebooks and throwing away the worse half of its codewords is standard.

2. Multiple-relay case $(M \ge 2)$

We consider B blocks of transmission, each of T transmission slots. A sequence of $B - M + 1$ indices, $w(b) \in \{1, \dots, 2^{TR}\}$, $b = 1, 2, \dots, B - M + 1$ will be sent over in TB transmission slots. (Note that as $B \to \infty$, the rate $TR(B - M + 1)/TB$ is arbitrarily close to R for any T.)

Generation of codebooks.

1) Generate at random 2^{TR} i.i.d. T-sequences in $\mathcal{X}_{M-1}^T$, each drawn according to

$$\text{Prob}(\boldsymbol{x}_{M-1}) = \prod_{t=1}^{T} p(x_{M-1,t}).$$

Index them as $\boldsymbol{x}_{M-1}(w_{M-1})$, $w_{M-1} \in \{1, 2, \dots, 2^{TR}\}$. This is the random codebook for node $M - 1$.

2) For each $\boldsymbol{x}_{M-1}(w_{M-1})$, generate 2^{TR} conditionally independent T-sequences $\boldsymbol{x}_{M-2}(w_{M-2}|w_{M-1})$, $w_{M-2} \in \{1, 2, \dots, 2^{TR}\}$, drawn independently according to

$$\begin{aligned}&\text{Prob}(\boldsymbol{x}_{M-2}|\boldsymbol{x}_{M-1}(w_{M-1}))\\ &\qquad = \prod_{t=1}^{T} p(x_{M-2,t}|x_{M-1,t}(w_{M-1})).\end{aligned}$$

This defines the joint codebook for nodes $M-2$, $M-1$: $\{\boldsymbol{x}_{M-2}(w_{M-2}|w_{M-1}), \boldsymbol{x}_{M-1}(w_{M-1})\}$.

3) Continue the process 2) sequentially for nodes $i = M - 3, M - 4, \dots, 0$, as follows: For each

$$\begin{aligned}&\{\boldsymbol{x}_{i+1}(w_{i+1}|w_{i+2}, \dots, w_{M-1}),\\ &\quad \boldsymbol{x}_{i+2}(w_{i+2}|w_{i+3}, \dots, w_{M-1}), \dots, \boldsymbol{x}_{M-1}(w_{M-1})\}\\ &\qquad =: \boldsymbol{x}_{i+1}^{M-1}(w_{i+1}, \dots, w_{M-1})\end{aligned}$$

generate 2^{TR} conditionally independent T-sequence $\boldsymbol{x}_i(w_i|w_{i+1}, \dots, w_{M-1})$, drawn independently according to

$$\begin{aligned}&\text{Prob}\left(\boldsymbol{x}_i|\boldsymbol{x}_{i+1}^{M-1}(w_{i+1}, \dots, w_{M-1})\right)\\ &\qquad = \prod_{t=1}^{T} p\left(x_{i,t}|x_{i+1,t}^{M-1,t}(w_{i+1}, \dots, w_{M-1})\right).\end{aligned}$$

Finally, we get a joint codebook for all the transmitter nodes $0, 1, \dots, M - 1$ as

$$\begin{aligned}\mathcal{C}_0 = \Big\{&\boldsymbol{x}_0(w_0|w_1, \dots, w_{M-1}),\\ &\boldsymbol{x}_1(w_1|w_2, \dots, w_{M-1}), \dots, \boldsymbol{x}_{M-1}(w_{M-1})\Big\}\end{aligned}$$

with each $w_i \in \{1, 2, \dots, 2^{TR}\}$, for $i = 0, 1, \dots, M - 1$.

Repeating the above process 1)–3) *independently* $M - 1$ times, we generate another $M - 1$ random codebooks $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_{M-1}$ similar to $\mathcal{C}_0$. We will use these M codebooks in a sequential way as follows: In block $b = 1, \dots, B$, the codebook $\mathcal{C}_{(b \bmod M)}$ is used. Hence, in any M consecutive blocks, codewords from different blocks are independent. This is a property we will use in the analysis of the probability of error (see (41)).

Before the transmission, all the joint codebooks $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_{M-1}$ are revealed to all the nodes $0, 1, \dots, M$.

Encoding: At the *beginning* of each block $b \in \{1, \dots, B\}$, every node $i \in \{0, \dots, M-1\}$ has estimates (see the decoding section) $\hat{w}_i(b - k + 1)$ of $w(b - k + 1)$, $k \ge i + 1$ (with $\hat{w}_0(b_1) = w(b_1)$ for $1 \le b_1 \le b$), and sends the following T-sequence from the codebook $\mathcal{C}_{(b \bmod M)}$ in the block

$$\vec{X}_i(b) := \boldsymbol{x}_i(\hat{w}_i(b-i)|\hat{w}_i(b-i-1), \dots, \hat{w}_i(b-M+1)) \quad (33)$$

where we set $\hat{w}_i(b_1) = w(b_1) := 1$ for every $b_1 \le 0$. As an example, the encoding process for two relays $(M = 3)$ is depicted in Table III.

Every node $k \in \{1, 2, \dots, M\}$ thus receives a T-sequence

$$\vec{Y}_k(b) := \vec{Y}_k(\vec{X}_0(b), \dots, \vec{X}_{M-1}(b)) \quad (34)$$

with probability

$$\begin{aligned}&\text{Prob}(\vec{Y}_k(b)|\vec{X}_0(b), \dots, \vec{X}_{M-1}(b))\\ &\qquad = \prod_{t=1}^{T} p(\vec{Y}_{k,t}(b)|\vec{X}_{0,t}(b), \dots, \vec{X}_{M-1,t}(b))\end{aligned}$$

$$\begin{cases}\text{Prob}\,(\text{some } w' \in \mathcal{W}_1(b) \text{ but } w' \ne w(b)) \le (2^{TR} - 1)2^{-T(I(X_0;Y_1|X_1)-6\epsilon)} < \varepsilon_1\\ \text{Prob}\,(\text{some } w' \in \mathcal{W}_2(b) \text{ but } w' \ne w(b-1)) \le (2^{TR} - 1)2^{-T(I(X_0,X_1;Y_2)-12\epsilon)} < \varepsilon_1.\end{cases} \quad (32)$$

where $\vec{X}_{k,t}(b)$ is the tth element of the vector $\vec{Y}_k(b)$, and similar definitions hold for $\vec{X}_{i,t}(b)$, $i = 0, \dots, M-1$.

Decoding: At the *end* of each block $b \in \{1, \dots, B\}$, every node $k \in \{1, \dots, M\}$ (for $b - k + 1 \geq 1$) declares

$$\hat{w}_k(b - k + 1) = w$$

if w is the unique value in $\{1, \dots, 2^{TR}\}$ such that in *all* the blocks $b - j$, $j = 0, 1, \dots, k-1$ we get (35) at the bottom of the page. Otherwise, if no unique w as above exists, an error is declared with $\hat{w}_k(b - k + 1) = 0$.

Analysis of Probability of Error: Denote the event that no decoding error is made in the first b blocks by

$$A_c(b) := \{\hat{w}_k(b_1 - k + 1) = w(b_1 - k + 1), \\ \text{for all } b_1 \in \{1, \dots, b\} \text{ and } k \in \{1, \dots, M\}\}$$

and let its probability be

$$P_c(b) := \mathrm{Prob}(A_c(b))$$

with $P_c(0) := 1$.

Then the probability that some decoding error is made at some node $k \in \{1, \dots, M\}$ in some block $b \in \{1, \dots, B\}$ is

$$\begin{aligned} P_e :=& \,\mathrm{Prob}\left(\hat{w}_k(b-k+1) \neq w(b-k+1),\right. \\ & \left. \text{for some } k \in \{1, \dots, M\}, b \in \{1, \dots, B\}\right) \\ =& \sum_{b=1}^{B} \mathrm{Prob}\left(\hat{w}_k(b-k+1) \neq w(b-k+1)\right. \\ & \left. \text{for some } k \in \{1, \dots, M\} \middle| A_c(b-1)\right) \cdot P_c(b-1) \\ \leq& \sum_{b=1}^{B} \sum_{k=1}^{M} \mathrm{Prob}\left(\hat{w}_k(b-k+1) \neq w(b-k+1) \middle| A_c(b-1)\right) \\ & \cdot P_c(b-1) \\ =:& \sum_{b=1}^{B} \sum_{k=1}^{M} P_{e,k}(b) \cdot P_c(b-1) \end{aligned} \tag{36}$$

where

$$P_{e,k}(b) := \mathrm{Prob}(\hat{w}_k(b - k + 1) \neq w(b - k + 1) | A_c(b - 1)).$$

Hence, $P_{e,k}(b)$ is the probability that a decoding error happens at node k in block b, conditioned on the event that no decoding error was made in the previous $b - 1$ blocks.

Next, we calculate $P_{e,k}(b)$, $k \in \{1, \dots, M\}$. Since $A_c(b-1)$ is presumed to hold, for every node $k \in \{1, \dots, M\}$ we have

$$\hat{w}_k(b_1 - k + 1) = w(b_1 - k + 1), \qquad \text{for } 1 \leq b_1 \leq b - 1.$$

Hence, by (33) and (34), the decoding rule (35) is equivalent to the following. Each node $k \in \{1, \dots, M\}$ (when $b - k + 1 \geq 1$) declares $\hat{w}_k(b - k + 1) = w$ if w is the unique value in $\{1, \dots, 2^{2R}\}$ such that the joint typicality check (37) holds simultaneously for all the blocks $b - j$, for $j = 0, 1, \dots, k - 1$

$$\begin{aligned} &\{\boldsymbol{x}_{k-1-j}(w | w(b-k), \dots, w(b-j-M+1)), \\ &\quad \boldsymbol{x}_{k-j}(w(b-k) | w(b-k-1), \dots, \\ &\qquad w(b-j-M+1)), \dots, \\ &\qquad \boldsymbol{x}_{M-1}(w(b-j-M+1)), \\ &\quad \vec{Y}_k(\boldsymbol{x}_0(w(b-j) | w(b-j-1), \dots, \\ &\quad w(b-j-M+1)), \dots, \boldsymbol{x}_{M-1}(w(b-j-M+1)))\} \\ &\qquad \in A_\epsilon^{(T)}(X_{k-1-j}, X_{k-j}, \dots, X_{M-1}, Y_k). \end{aligned} \tag{37}$$

Let

$$\mathcal{W}_{k,j}(b) := \{w \in \{1, \dots, 2^{TR}\} : w \text{ satisfies (37)}\}$$
$$\mathcal{W}_k(b) := \bigcap_{j=0}^{k-1} \mathcal{W}_{k,j}(b).$$

Then, $P_{e,k}(b)$ is the probability that $w(b - k + 1) \notin \mathcal{W}_k(b)$, or some $w' \in \mathcal{W}_k(b)$ but $w' \neq w(b - k + 1)$, conditioned on the event that no decoding error was made in the previous $b - 1$ blocks. Thus,

$$\begin{aligned} &P_{e,k}(b) \\ &\quad = \mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_k(b), \\ &\qquad\qquad \text{or some } w' \in \mathcal{W}_k(b) \\ &\qquad\qquad \text{but } w' \neq w(b-k+1) | A_c(b-1)) \\ &\quad \leq \mathrm{Prob}\left(w(b-k+1) \notin \mathcal{W}_k(b) | A_c(b-1)\right) \\ &\qquad + \mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \\ &\qquad\qquad \text{but } w' \neq w(b-k+1) | A_c(b-1)) \\ &\quad \leq \frac{\mathrm{Prob}\left(w(b-k+1) \notin \mathcal{W}_k(b)\right)}{P_c(b-1)} \\ &\qquad + \frac{\mathrm{Prob}\left(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1)\right)}{P_c(b-1)}. \end{aligned}$$

Hence, by (36)

$$\begin{aligned} P_e \leq& \sum_{b=1}^{B} \sum_{k=1}^{M} [\mathrm{Prob}\left(w(b-k+1) \notin \mathcal{W}_k(b)\right) \\ & + \mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1))]. \end{aligned} \tag{38}$$

Now, by Lemma 4.1 with

$$(Z_1, \dots, Z_m) = (X_{k-1-j}, \dots, X_{M-1}, Y_k)$$

for T large enough, we have for $j = 0, 1, \dots, k - 1$

$$\mathrm{Prob}(w(b - k + 1) \notin \mathcal{W}_{k,j}(b)) < \epsilon$$

$$\begin{aligned} &\left\{\boldsymbol{x}_{k-1-j}\left(w | \hat{w}_k(b-k), \dots, \hat{w}_k(b-j-M+1)\right), \dots, \boldsymbol{x}_{M-1}(\hat{w}_k(b-j-M+1)),\ \vec{Y}_k(b-j)\right\} \\ &\qquad\qquad \in A_\epsilon^{(T)}(X_{k-1-j}, \dots, X_{M-1}, Y_k). \end{aligned} \tag{35}$$

and for any $w' \neq w(b-k+1)$

$$\mathrm{Prob}(w' \in \mathcal{W}_{k,j}(b)) < 2^{-T(I(X_{k-1-j}, Y_k | X_{k-j}, \ldots, X_{M-1}) - 6\epsilon)}. \tag{39}$$

Hence,

$$\begin{aligned} &\mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_k(b)) \\ &\quad \leq \sum_{j=0}^{k-1} \mathrm{Prob}(w(b-k+1) \notin \mathcal{W}_{k,j}(b)) \\ &\quad \leq \sum_{j=0}^{k-1} \epsilon = k\epsilon \leq M\epsilon, \end{aligned} \tag{40}$$

and

$$\begin{aligned} &\mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1)) \\ &\quad \leq \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w(b-k+1)}} \mathrm{Prob}(w' \in \mathcal{W}_k(b)) \\ &\quad = \sum_{\substack{w' \in \{1,\ldots,2^{TR}\} \\ w' \neq w(b-k+1)}} \prod_{j=0}^{k-1} \mathrm{Prob}(w' \in \mathcal{W}_{k,j}(b)) \qquad (41) \\ &\quad \leq (2^{TR}-1) 2^{-T(I(X_0,\ldots,X_{k-1}; Y_k | X_k, \ldots, X_{M-1}) - 6k\epsilon)} \qquad (42) \end{aligned}$$

where (41) follows from the independence among the codebooks of any M consecutive blocks, and (42) follows from (39) and the following equation:

$$I(X_0,\ldots,X_{k-1}; Y_k | X_k, \ldots, X_{M-1}) = \sum_{j=0}^{k-1} I(X_{k-1-j}, Y_k | X_{k-j}, \ldots, X_{M-1}).$$

For any R satisfying (9), by choosing ϵ small enough, we can make T large enough such that for any $\varepsilon_1 > 0$ and all $k \in \{1, \ldots, M\}$

$$\begin{aligned} &\mathrm{Prob}(\text{some } w' \in \mathcal{W}_k(b) \text{ but } w' \neq w(b-k+1)) \\ &\quad \leq (2^{TR}-1) 2^{-T(I(X_0,\ldots,X_{k-1}; Y_k | X_k, \ldots, X_{M-1}) - 6k\epsilon)} \\ &\quad < \varepsilon_1. \end{aligned} \tag{43}$$

Hence, by (38), (40), and (43),

$$\begin{aligned} P_e &\leq \sum_{b=1}^{B} \sum_{k=1}^{M} (M\epsilon + \varepsilon_1) \\ &\leq BM^2\epsilon + BM\varepsilon_1 \end{aligned}$$

which can be made arbitrarily small by letting $T \to \infty$.

Finally, the argument on choosing one good codebook from many random codebooks and throwing away the worse half of the codewords is standard. □

V. Proof of Theorem 3.2

The achievability is proved in Theorem 3.1. The converse follows immediately from the max-flow min-cut theorem for general multiple-node networks stated in [14, Theorem 14.10.1], where the node set S is chosen to be $\{0\}, \{0,1\}, \ldots, \{0,1,\ldots,M-1\}$ sequentially, and with the following equations:

$$\begin{aligned} &I(X_0,\ldots,X_{k-1}; Y_k, \ldots, Y_M | X_k, \ldots, X_{M-1}) \\ &\quad = I(X_0,\ldots,X_{k-1}; Y_k | X_k, \ldots, X_{M-1}) \\ &\qquad + \sum_{j=k+1}^{M} I(X_0,\ldots,X_{k-1}; Y_j | Y_k, \ldots, \\ &\qquad\qquad Y_{j-1}, X_k, \ldots, X_{M-1}) \\ &\quad = I(X_0,\ldots,X_{k-1}; Y_k | X_k, \ldots, X_{M-1}), \\ &\qquad\qquad \text{for } k = 1, \ldots, M \end{aligned}$$

where the last equation follows from

$$\begin{aligned} &\sum_{j=k+1}^{M} I(X_0,\ldots,X_{k-1}; Y_j | Y_k, \ldots, Y_{j-1}, X_k, \ldots, X_{M-1}) \\ &\quad = \sum_{j=k+1}^{M} H(Y_j | Y_k, \ldots, Y_{j-1}, X_k, \ldots, X_{M-1}) \\ &\qquad - \sum_{j=k+1}^{M} H(Y_j | X_0, \ldots, X_{k-1}, Y_k, \ldots, Y_{j-1}, \\ &\qquad\qquad X_k, \ldots, X_{M-1}) \\ &\quad = H(Y_{k+1}, \ldots, Y_M | Y_k, X_k, \ldots, X_{M-1}) \\ &\qquad - H(Y_{k+1}, \ldots, Y_M | Y_k, X_0, \ldots, X_{M-1}) \\ &\quad = 0, \qquad \text{for } k = 1, \ldots, M-1 \end{aligned}$$

with the last equation following immediately from (8) and the definition of conditional entropy. □

VI. Proof of Theorem 3.3

We add a "virtual" node $M+1$ to the channel with output $y_{M+1} \equiv y_M$, but with no input. Hence, this node will not affect the dynamics of the channel in any way. Then by Theorem 3.1, the following rate is achievable from node 0 to node $M+1$:

$$R < \max_{p(x_0,\ldots,x_M)} \min_{1 \leq k \leq M+1} I(X_0, \ldots, X_{k-1}; Y_k | X_k, \ldots, X_M). \tag{44}$$

By the coding scheme stated in the proof of Theorem 3.1, the rate above is also achieved from node 0 to node M. (This is also obvious since $y_{M+1} \equiv y_M$ and there is no x_{M+1}.)

Now by (44), to prove Theorem 3.1, we only need to show that the following inequality always holds:

$$I(X_0, \ldots, X_M; Y_{M+1}) \geq I(X_0, \ldots, X_{M-1}; Y_M | X_M).$$

Fortunately, by the construction that $Y_{M+1} = Y_M$, it follows immediately that

$$\begin{aligned} &I(X_0, \ldots, X_M; Y_{M+1}) \\ &\quad = I(X_0, \ldots, X_M; Y_M) \\ &\quad = I(X_M; Y_M) + I(X_0, \ldots, X_{M-1}; Y_M | X_M) \\ &\quad \geq I(X_0, \ldots, X_{M-1}; Y_M | X_M). \end{aligned}$$ □

Finally, the proof of Theorem 3.4 is similar to that of Theorem 3.1 and is omitted.

VII. Concluding Remarks

We have developed a simple coding scheme for the multiple-level relay problem. This allows us to develop a simple expression for an achievable rate that is generally higher than that in [9]. For degraded channels, our result achieves the capacity. Also, we generalize this result to the case where the destination is allowed to transmit. The achievable rate that is established is higher than that established when the destination simply "facilitates" the channel by sending a constant signal.

References

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

[2] ——, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Dept. Statist., Univ. Calif., Berkeley, 1968.

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

[4] B. Schein and R. Gallager, "The Gaussian parallel relay network," in *Proc. 2000 IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.

[5] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, submitted for publication.

[6] A. B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 841–850, Nov. 1982.

[7] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 3, pp. 313–327, May 1985.

[8] C.-M. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inf. Theory*, vol. 35, no. 6, pp. 1160–1165, Nov. 1989.

[9] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[10] A. Reznik, S. Kulkarni, and S. Verdú, "Capacity and optimal resource allocation in the degraded Gaussian relay channel with multiple relays," in *Proc. 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002.

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

[12] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc. 41th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003.

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

[14] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.