Degraded Gaussian Multirelay Channel: Capacity and Optimal Power Allocation

Alex Reznik, *Member, IEEE*, Sanjeev R. Kulkarni, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—We determine the capacity region of a degraded Gaussian relay channel with multiple relay stages. This is done by building an inductive argument based on the single-relay capacity theorem of Cover and El Gamal. For an arbitrary distribution of noise powers, we derive the optimal power distribution strategy among the transmitter and the relays and the best possible improvement in signal-to-noise ratio (SNR) that can be achieved from using a given number of relays. The time-division multiplexing operation of the relay channel in the wideband regime is analyzed and it is shown that time division does not achieve minimum energy per bit.**

***Index Terms*—Capacity, optimal resource allocation, relay channel, wideband channels.**

I. Introduction

THE relay channel was introduced by van der Muelen in his work on multiterminal networks [9]–[11]. The most thorough analysis to date was provided in 1979 by Cover and El Gamal [2]. In particular, they determined the capacity region for the physically degraded version of the channel.

Until recently, little has been done to extend these results to channels with multiple relays. However, a renewed interest in *ad hoc* networks and network information theory has sparked new research on relay channels. One set of recent results in this area is in [5] where Gupta and Kumar demonstrate an achievable rate region result for a fairly general communication network, of which a degraded relay channel is a special case. The results in [5] hold for both the discrete and the additive white Gaussian noise memoryless models. A followup paper by Xie and Kumar [13] establishes an explicit achievable rate expression for the degraded Gaussian channel with multiple relays which, in general, exceeds the rate in [5].

In this paper, we concentrate our attention on a Gaussian physically degraded relay channel with multiple relay stages. Fig. 1 depicts such a channel with K relay stages. As shown in Fig. 1, the channel consists of a transmitter, whose output at transmission time i is $x_{0,i}$ and K relays whose outputs at transmission time i are $x_{1,i}$ through $x_{K,i}$. The input to relay k at transmission time i is $y_{k-1,i}$ with $y_{K,i}$ received by the receiver. The channel is physically degraded in the sense of [3] since

$$\begin{aligned} y_{k,i} &= x_{k,i} + z_{k,i} + y_{k-1,i}, \qquad 1 \le k \le K \\ y_{0,i} &= x_{0,i} + z_{0,i}. \end{aligned} \tag{1}$$

At each transmission stage, the signal is corrupted by independently generated Gaussian random variables Z_0 through Z_K with $Z_k \sim \mathcal{N}(0, N_k)$. Denote the transmitter power by P_0 and the power of relay k by P_k. Let β_k denote the ratio between the relay power and the transmitter power, thus,

$$\beta_k \stackrel{\text{def}}{=} \frac{P_k}{P_0}. \tag{2}$$

To simplify notation we will sometimes use β_0 which is always equal to 1. We also define

$$\nu_k \stackrel{\text{def}}{=} \frac{N_k}{\sum_{j=0}^{k-1} N_j}. \tag{3}$$

The goal of the paper is to determine the capacity of this channel for any given set of $P_0, \dots, P_K$ and $N_0, \dots, N_K$. Additionally, we seek to determine $\beta_1, \dots, \beta_K$ such that the capacity is maximized. We are to do this under the constraint that only a finite total amount of power P_T is available, i.e., under the constraint that

$$1 + \sum_{k=1}^{K} \beta_k \le \frac{P_T}{P_0}.$$

The rest of the paper is structured as follows. In Section II, we summarize the single-relay degraded Gaussian channel results obtained by Cover and El Gamal in [2].

In Section III, we build upon the achievability results of [2] and use an inductive argument to determine the capacity region. Thus, we give an alternative derivation for the achievable rate determined in [13]. With our inductive proof, we demonstrate how the coding strategy can be built recursively on the basis of the bin-coding argument utilized by [2]. By means of another inductive argument we then prove the converse to the capacity theorem as well.

For the capacity-achieving communication strategy, we determine how power should be distributed between the transmitter

Manuscript received February 19, 2003; revised August 2, 2004. this work was supported in part by ODDR&E MURI through the Army Research Office under Grant DAAD19-00-1-0466, Draper Laboratory under IR&D 6002 Grant DL-H-546263, and the National Science Foundation under Grant CCR-0312413. The material in this paper was presented in part at the 40th Annual Allerton Conference on Communications, Control and Computing, October 2002 [7].

A. Reznik is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA and Incubation Center, InterDigital Communications Corp., King of Prussia, PA 19406 USA (e-mail: Alex.Reznik@interdigital.com).

S. R. Kulkarni and S. Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: kulkarni@princeton.edu; verdu@princeton.edu).

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

Digital Object Identifier 10.1109/TIT.2004.838373

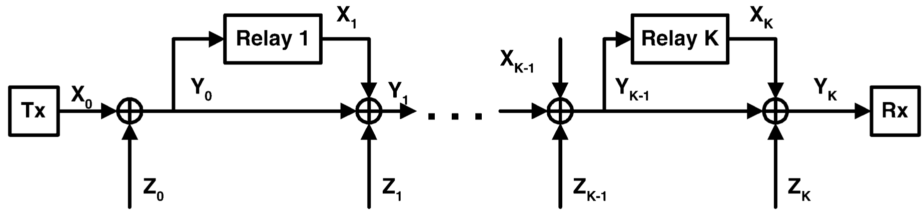

Fig. 1. Physically degraded Gaussian relay channel with K relay stages.

and the relays so that the resulting capacity is maximized under a total power constraint. It is shown that the capacity achieved by optimally distributing the total available power P_T between the transmitter and the relays is given by

$$\boldsymbol{C}_K = C\left(\frac{P_T}{\sum_{k=0}^{K}\frac{N_k}{k+1}}\right). \tag{4}$$

These results are presented in Section IV.

Finally, in Section V, we address the operation of the degraded relay channels in the wideband regime. Here, it is straightforward to show that time-division multiplexing (TDM) cannot achieve $\frac{E_b}{N_0}_{\min}$ achieved by the optimal communications scheme.

II. The Single-Relay Channel

We begin by restating the result of [2], changing the notation of [2] to that used in this paper.

Theorem 1 (Single-Relay Capacity ([2])): The capacity $\boldsymbol{C}_1$ of the single-relay degraded Gaussian channel is given by

$$\boldsymbol{C}_1 = \max_{0\le\alpha_1\le 1}\min\left\{C\left(\frac{1+\beta_1+2\sqrt{(1-\alpha_1)\beta_1}}{1+\nu_1}\frac{P_0}{N_0}\right),\; C\left(\alpha_1\frac{P_0}{N_0}\right)\right\} \tag{5}$$

where $C(x) \stackrel{\text{def}}{=} \frac{1}{2}\log(1+x)$.

Moreover, let α_1^* denote the value of α that achieves the optimum in (5). Then, if $\beta_1 < \nu_1$

$$\boldsymbol{C}_1 = C\left(\alpha_1^*\frac{P_0}{N_0}\right) = C\left(\alpha_1^*\frac{P_0(1+\nu_1)}{N_0+N_1}\right) \tag{6}$$

with $\alpha_1^* < 1$; otherwise, $\alpha_1^* = 1$ and

$$\boldsymbol{C}_1 = C\left(\frac{P_0}{N_0}\right). \tag{7}$$

Proof: The ideas we use to prove out multirelay results in Section III rely heavily on the techniques introduced in [2] to prove the single-relay result. Therefore, it is useful to reproduce the proof of [2] in a condensed form.

We begin with the proof of *achievability* of the rate given above. For $0 \le \alpha_1 \le 1$ let

$$X_1 \sim \mathcal{N}(0, \alpha_{1,1}P) \quad \text{and} \quad \hat{X}_0 \sim \mathcal{N}(0, \alpha_1 P)$$

with $\hat{X}_0, X_1$ independent and let $X_0 = \sqrt{1-\alpha_1}X_1 + \hat{X}_0$.

The proof uses block-Markov coding. Let

$$\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$$

be the set of messages to be transmitted. Let

$$\mathcal{S}_1 = \{S_{1,1}, S_{1,2}, \ldots, S_{1,2^{nR_1}}\}$$

be a partition of $\mathcal{W}$ generated in a uniform and random fashion independently from everything else.

We have two random codebooks:

- $\hat{\boldsymbol{X}}_0(w)$ independent and identically distributed (i.i.d.) $\sim \mathcal{N}_n(0, \alpha_1 P)$, $w \in \mathcal{W}$
- $\boldsymbol{X}_1(s_1)$ i.i.d. $\sim \mathcal{N}_n(0, (1-\alpha_1)P)$, $s_1 \in \mathcal{S}_1$.

Finally, for transmission block i, $s_{1,i}$ is chosen so that $s_{1,i} = S_{1,j}$ where $w_{i-1} \in S_{1,j}$. It is then shown in [2] that the receiver can decode the message w_{i-1} at the end of transmission interval i.

To prove the *converse* result we begin by defining

$$\mathcal{I}_0 = I(X_0; Y_k | X_1) \tag{8}$$

and

$$\mathcal{I}_1 = I(X_0, X_1; Y_k). \tag{9}$$

Then from the cutset bound ([3, Ch. 14]) and the degradedness of the channel it follows that

$$\boldsymbol{C}_1 \le \sup_{p(x_0,x_1)} \min[\mathcal{I}_0, \mathcal{I}_1]. \tag{10}$$

Using convexity arguments we can then show that

$$\mathcal{I}_0 \le \mathsf{E}C\left(\frac{\mathsf{Var}(X_0|X_1)}{N_0}\right) \le C\left(\frac{\mathsf{E}\,\mathsf{Var}(X_0|X_1)}{N_0}\right) \tag{11}$$

and

$$\mathcal{I}_1 \le \mathsf{E}C\left(\frac{\mathsf{Var}(X_0+X_1)}{N_0+N_1}\right) \le C\left(\frac{\mathsf{E}\,\mathsf{Var}(X_0+X_1)}{N_0+N_1}\right). \tag{12}$$

The proof then reduces to showing that the inequalities in (11) and (12) can be made tight by defining

$$\alpha_1 \stackrel{\text{def}}{=} \frac{1}{P_0}\left(1 - \mathsf{E}\mathsf{E}^2(X_0|X_1)\right). \tag{13}$$

The details of the computation are provided in [2] □

The value of α_1^* in (6) may be obtained explicitly. As stated, if $\beta_1 \ge \nu_1$ then $\alpha_1^* = 1$. Otherwise, α_1^* is obtained by solving

$$\alpha_1^* = \frac{1+\beta_1+2\sqrt{(1-\alpha_1^*)\beta_1}}{1+\nu_1}. \tag{14}$$

Doing this we obtain

$$\alpha_1^* = \frac{1-\beta_1+\nu_1+\nu_1\beta_1 \pm 2\sqrt{\nu_1^2\beta_1 - \nu_1\beta_1^2 + \nu_1\beta_1}}{(1+\nu_1)^2} = \frac{\left(\sqrt{1+\nu_1-\beta_1} \pm \sqrt{\nu_1\beta_1}\right)^2}{(1+\nu_1)^2}. \tag{15}$$

In the context of our problem only one of the two solutions is valid. As long as $0 < \beta_1 \leq \nu_1$, this solution is the one that lies in the range $(\frac{1}{1+\nu_1}, 1]$ since $\alpha_1^* = \frac{1}{1+\nu_1}$ is the result obtained if the relay is not used at all. This is the solution

$$\alpha_1^* = \frac{\left(\sqrt{1+\nu_1-\beta_1}+\sqrt{\nu_1\beta_1}\right)^2}{(1+\nu_1)^2}. \tag{16}$$

This solution is always less than 1 if $\beta_1 < \nu_1$. If $\beta_1 = \nu_1$, it reduces to $\alpha_1^* = 1$ from which we can conclude that α_1^* is a continuous function of β_1.

We can also solve (14) for β_1 under the assumption that $\alpha_1^* < 1$. This results in

$$\beta_1 = \left(\sqrt{\alpha_1^*\nu_1} - \sqrt{1-\alpha_1^*}\right)^2. \tag{17}$$

We next pose the optimum power allocation problem for the single-relay channel. We would like to find the value of β_1 such that the capacity of the channel is maximized subject to $1+\beta_1 \leq \frac{P_T}{P_0}$ for some fixed P_T. Since the channel capacity is given by

$$C\left(\alpha_1^*\frac{P_0}{N_0}\right) = C\left(\alpha_1^*(1+\nu_1)\frac{P_0}{N_0+N_1}\right)$$

we can remove the constraint on β_1 by maximizing the improvement in the "effective signal-to-noise ratio (SNR)" delivered by the relay channel. This "effective SNR improvement" is given by

$$J_1(\alpha_1, \beta_1) = \frac{\alpha_1(1+\nu_1)}{1+\beta_1}. \tag{18}$$

The preceding expression is actually an expression in only one variable since we can either use (16) to express α_1 in terms of β_1 or (17) to express β_1 in terms of α_1. We note that such a substitution is only valid if the resulting optimum is such that $\alpha_{1,\text{opt}} < 1$, however, this is true for all $\nu_1 > 0$. In fact, maximizing J_1 we obtain

$$\alpha_{1,\text{opt}} = \frac{4}{4+\nu_1} \tag{19}$$

and

$$\beta_{1,\text{opt}} = \frac{\nu_1}{4+\nu_1} \tag{20}$$

where we note that

$$\beta_{1,\text{opt}} = 1 - \alpha_{1,\text{opt}}. \tag{21}$$

The relationship (21) becomes important when we consider the general resource optimization problem for multiple relays.

III. The Multiple-Relay Channel

We now consider the problem of the multiple-relay channel. This problem has been previously considered in [5] and [13] where achievable rates were found for a channel of which the multiple-relay degraded Gaussian channel is a special case. In this section, we prove explicitly the capacity of the multiple-relay degraded Gaussian channel. It turns out that the achievable rate found in [13] is the capacity of the degraded multirelay channel.

For a specified choice of $\alpha_{i,j}$'s with $0 \leq i \leq j \leq K$ satisfying

$$\sum_{j=0}^{K} \alpha_{0,j} = 1 \tag{22}$$

and

$$\sum_{j=i}^{K} \alpha_{i,j} = \beta_i, \qquad \forall\, 1 \leq i \leq K \tag{23}$$

define

$$\mathcal{R}_k(\bar{\alpha}) = C\left(P_0 \frac{\sum_{j=0}^{k}\left(\sum_{i=0}^{j}\sqrt{\alpha_{i,j}}\right)^2}{\sum_{j=0}^{k} N_j}\right) \tag{24}$$

and

$$\mathcal{C}_K(\bar{\alpha}) = \min_{0\leq k\leq K} \mathcal{R}_k(\bar{\alpha}) \tag{25}$$

where we use $\bar{\alpha}$ as a shorthand for $\{\alpha_{i,j}\}_{0\leq i\leq j\leq K}$. We then have the following theorem.

Theorem 2 (Multirelay Capacity): The capacity of the multiple-relay degraded Gaussian channel with K relays is given by

$$\boldsymbol{C}_K = \sup_{\{\alpha_{i,j}\}} \mathcal{C}_K(\bar{\alpha}) \tag{26}$$

with $\mathcal{C}_K$ as defined by (24) and (25).

Proof: We prove both the achievability and the converse parts of the theorem by induction. In both cases, the single-relay result of [2] as presented in Theorem 1 serves as the initial step in the induction. Indeed, using the notation presented above, we have for the single-relay channel: $K = 1$; α_1 of Theorem 1 is $\alpha_{0,0}$; $1-\alpha_1$ of Theorem 1 is $\alpha_{0,1}$; $\alpha_{1,1} = \beta_1$; additionally

$$\begin{aligned}
\mathcal{R}_0(\bar{\alpha}) &= C\left(\alpha_1\frac{P_0}{N_0}\right)\\
\mathcal{R}_1(\bar{\alpha}) &= C\left(\frac{1+\beta_1+2\sqrt{(1-\alpha_1)\beta_1}}{1+\nu_1}\frac{P_0}{N_0}\right)\\
&= C\left(\frac{\left(\sqrt{\alpha_1}\right)^2+\left(\sqrt{1-\alpha_1}+\sqrt{\beta_1}\right)^2}{1+\nu_1}\frac{P_0}{N_0}\right).
\end{aligned}$$

To prove achievability, we need to specify our coding strategy. We simply extend the method used in [2]. The resulting coding strategy is similar to the one proposed in [13], although our method for generating it is recursive and it builds directly on the coding strategy used in [2]. This is unlike [13], where the coding strategy is specified directly. The resulting coding strategy can also be thought of as a multiple-layered block-Markov coding approach where transmitter k uses $K-k$ layers of block-Markov coding for data transmission.

Achievability: For the induction step of the proof of achievability, assume that the theorem holds for $K-1$ relays. Fix some appropriate choice of $\alpha_{i,j}$'s. Let $\mathcal{C}_{K-1}(\bar{\alpha})$ be the rate achievable in this channel and with this choice of $\alpha_{i,j}$'s and assume that this rate is achievable using a codebook such that the output of transmitter k is given by a random variable $\widetilde{X}_k \sim \mathcal{N}(0, \beta_k P_0)$.

Now consider adding another relay at the end of the last stage. One way to do this is to turn the final receiver into a relay and add a new receiver after this relay. Thus, we can think of this operation as simply adding a new transmitter (indexed K) and a new receiver (indexed K).

Let $\mathcal{W} = \{1, 2, \ldots, 2^{nR}\}$ be the set of messages to be transmitted. Let $\mathcal{S}_K = \{S_{K,1}, S_{K,2}, \ldots, S_{K,2^{nR_K}}\}$ be a partition of $\mathcal{W}$ generated in a uniform and random fashion independently from everything else. Define a random codebook $\boldsymbol{X}_K(s_K)$ i.i.d. $\sim \mathcal{N}_n(0, \alpha_{K,K} P_0)$, $s_K \in \{1, \ldots, 2^{nR_K}\}$, where we note that $\alpha_{K,K} \equiv \beta_K$. For transmission block i, $s_{K,i}$ is chosen so that $s_{K,i} = j$ where $w_{i-K} \in S_{K,j}$. Now for $0 \leq k \leq K-1$ define

$$X_k = \sqrt{1 - \alpha_{k,K}}\widetilde{X}_k + \sqrt{\frac{\alpha_{k,K}}{\alpha_{K,K}}} X_K$$

where $\alpha_{k,K}$ is the proportion of the power that transmitter k allocated to the newly added receiver K.

In [2], it was necessary to assume that at the start of transmission block i the relay (receiver) has successsrully decoded messages $w_1, \ldots, w_{i-1}$. Extending this assumption, we assume that at the start of transmission block i, the receiver k has successsrully decoded messages $w_1, \ldots, w_{i-k}$. In particular, at transmission block i, all receivers up to and including receiver $K-1$ know w_{i-K}. This assumption should be thought of as part of the induction hypothesis. Alternatively, one may also assume that receivers rely on other decoding techniques, e.g., the backward decoding of [14] to achieve reliable communication through this instance of time.

Thus, receivers 1 through $K-1$ can successsrully remove the contribution from X_K to the received signal. Thus, the rate $\mathcal{C}_{K-1}(\bar{\alpha})$ as defined by (24) and (25) is achievable from the point of view of communicating to receivers 1 through $K-1$. We note that by adding nonzero $\alpha_{i,K}$'s we reduced $\mathcal{C}_{K-1}(\bar{\alpha})$ since $\sum_{j=i}^{K-1} \alpha_{i,j}$ is reduced from β_i to $\beta_i - \alpha_{i,K}$.

Suppose now that, under the assumption that reliable communication is achieved to receivers 1 through $K-1$, it is possible to communicate to receiver K at a rate $\mathcal{R}_K(\bar{\alpha})$ as defined by (24). We then note that because the same information is being communicated to all the receivers, the rate

$$\min(\mathcal{C}_{K-1}(\bar{\alpha}), \mathcal{R}_K(\bar{\alpha})) = \min_{1 \leq k \leq K} \mathcal{R}_k(\bar{\alpha})$$

is achievable since we can communicate reliably at this rate to all the receivers. Finally, taking a supremum over the choices of $\alpha_{i,j}$'s we obtain the desired capacity rate.

It remains to show that, assuming that reliable communication to all other receivers is attained, it is indeed possible to communicate to receiver K at a rate $\mathcal{R}_K(\bar{\alpha})$. To do this, we examine Y_K, the received signal at the ultimate receiver, receiver K. It follows from our recursive codebook construction that

$$Y_K = \sum_{k=0}^{K} Z_k + P_0 \sum_{k=0}^{K} \left(\sum_{i=0}^{k} \sqrt{\alpha_{i,k}} \right) U_k \tag{27}$$

where $U_k \sim \mathcal{N}(0,1)$ and U_k's are independent from each other and from Z_k's. Specifically, U_k, $1 \leq k \leq K$ represents the codebook used to encode the random variable S_k where the $s_{k,i}$ (the realization of S_k at transmission block i) carries information regarding which set in the partition $\mathcal{S}_k$ of $\mathcal{W}$ the message w_{i-k} belonged to. U_0 represents the encoding of W—i.e., the messages themselves. Note that we can now clearly identify the meaning of $\alpha_{k,j}$: it is the power applied to U_j at transmitter k.

We now proceed analogously to [2], [5], and [13]. We use successive interference cancellation to obtain and decode information carried by U_k's starting from U_K and proceeding in the order of descending subscript. Then, the information carried by U_k can be reliably received by receiver K if its rate is not greater than

$$\widehat{R}_k(\bar{\alpha}) = C \left(\frac{P_0 \left(\sum_{i=0}^{k} \sqrt{\alpha_{i,k}} \right)^2}{\sum_{k=0}^{K} N_k + P_0 \sum_{j=0}^{k-1} \left(\sum_{i=0}^{j} \sqrt{\alpha_{i,j}} \right)^2} \right). \tag{28}$$

Thus, the message received by receiver K at transmission block i carries independent information regarding w_{i-k} at a rate no greater than $\widehat{R}_k$. Thus, at transmission block i the information available about the message w_{i-K} is equal to

$$\sum_{k=0}^{K} \widehat{R}_k(\bar{\alpha}) = \mathcal{R}_K(\bar{\alpha})$$

as required.

After B transmission blocks, we can therefore achieve a rate of $\mathcal{R}_K(\bar{\alpha}) \frac{B-K}{B}$ that approaches $\mathcal{R}_K(\bar{\alpha})$ as $B \to \infty$. This shows that under the assumption that reliable communication to receivers 0 through $K-1$ is achieved, we can achieve reliable communication to receiver K at a rate $\mathcal{R}_K$; and this completes the proof of the achievability part of the theorem.

Converse: We begin by defining

$$\mathcal{I}_k = I\left(X_0, \ldots, X_k; Y_k | X_{k+1}, \ldots, X_K\right). \tag{29}$$

Then from the cutset bound ([3, Ch. 14)]) and the degradedness of the channel it follows that

$$\boldsymbol{C}_K \leq \sup_{p(x_0, \ldots, x_K)} \min_{0 \leq k \leq K} \mathcal{I}_k. \tag{30}$$

Then we have

$$\mathcal{I}_k \leq \mathsf{E} C \left(\frac{\mathsf{Var}\left(X_0 + \cdots + X_k | X_{k+1}, \ldots, X_K\right)}{\sum_{j=0}^{k} N_j} \right) \tag{31}$$

$$\leq C \left(\frac{\mathsf{E}\,\mathsf{Var}\left(X_0 + \cdots + X_k | X_{k+1}, \ldots, X_K\right)}{\sum_{j=0}^{k} N_j} \right). \tag{32}$$

Thus, we need to show that there exists some choice of real values $\{\alpha_{i,j}\}$ defined for $0 \leq i \leq j \leq K$ and satisfying constraints (22) and (23) such that

$$\mathsf{E}\,\mathsf{Var}\left(X_0 + \cdots + X_k | X_{k+1}, \ldots, X_K\right) \leq P_0 \sum_{j=0}^{k} \left(\sum_{i=0}^{j} \sqrt{\alpha_{i,j}} \right)^2. \tag{33}$$

If we can show this for some appropriate set $\{\alpha_{i,j}\}$, then we have found a set of $\{\alpha_{i,j}\}$ such that

$$\min_{0 \leq k \leq K} \mathcal{I}_k \leq \mathcal{C}_{\boldsymbol{K}}(\bar{\alpha}) \tag{34}$$

which would prove the converse.

To define an appropriate set $\{\alpha_{i,j}\}$ we follow the approach used in [2]. Specifically, we define

$$\alpha_{i,K} \stackrel{\text{def}}{=} \frac{1}{P_0} \mathsf{E}\mathsf{E}^2(X_i | X_K). \tag{35}$$

The rest of the α's will be defined implicitly based on our induction argument; at this point, we only note that from (22) and (23) these must satissry the constraint

$$\sum_{j=i}^{K-1} \alpha_{i,j} = \frac{1}{P_0}\left(\mathsf{E}X_i^2 - \mathsf{E}\mathsf{E}^2(X_i|X_K)\right), \qquad 0 \le i \le K-1. \tag{36}$$

Moreover, we note that in the case of a single-relay channel, the definition in (35) reduces to the definition of $(1-\alpha)$ in the proof of the converse part of Theorem 1 in [2]; and the left-hand side of (36) reduces to α in the proof of the converse part of Theorem 1 in [2]. Thus, we can indeed use the single-relay proof of [2] as the initial step in an induction proof of the overall theorem.

For the inductive step, let us assume that the converse holds for a $(K-1)$ relay channel. Thus, we assume that for every choice of the transmitter output distribution and noise powers there exists a choice of real values $\{\widetilde{\alpha}_{i,j}\}$ for $0 \le i \le j \le K-1$ satisfying (35) and satisfying (36) for any choice of real values β_j, $1 \le j \le K-1$ such that

$$\mathsf{E}\,\mathsf{Var}\left(X_0 + \cdots + X_k | X_{k+1}, \ldots, X_{K-1}\right) \le P_0 \sum_{j=0}^{k}\left(\sum_{i=0}^{j}\sqrt{\widetilde{\alpha}_{i,j}}\right)^2 \tag{37}$$

holds for all $0 \le k \le K-1$.

Now consider a K-relay channel. Define, as we did throughout, β_i to be the power of relay i, normalized to the transmitter power P_0.

First we show that (33) must hold for $0 \le k \le K-1$. Our argument is constructed as follows. We fix $\{\alpha_{i,K}\}$'s for all $0 \le i \le K$. Conditioned on such a choice (i.e., conditioned on X_K), we construct an equivalent $K-1$ relay channel with the following two properties:

- The set of $\{\widetilde{\alpha}_{i,j}\}_{0\le i\le j\le K-1}$ satisfying (22) and (23) in the $(K-1)$-relay channel we construct is in one-to-one correspondence with the set of $\{\widetilde{\alpha}_{i,j}\}_{0\le i\le j\le K-1}$ satisfying (36) in the original channel (conditioned on X_K).
- Equation (37) holds for $0 \le k \le K-1$ for the $(K-1)$-relay channel we constructed if and only if (33) holds for $0 \le k \le K-1$ for the original K-relay channel.

Having shown this, we conclude that (33) must hold for $0 \le k \le K-1$, for otherwise we violate the induction hypothesis.

Let X_i denote the output of transmitter i and Z_i denote the noise added at stage i. Fix $\{\alpha_{i,K}\}$, $0 \le i \le K$. The rest of the argument is essentially the following: conditioned on knowledge of X_K and having fixed the resources allocated to the transmission of X_K (these are the $\alpha_{i,K}$'s) the first $(K-1)$ stages of our K-relay channel are equivalent to a $(K-1)$-relay channel for which the converse holds by the induction hypothesis. Thus, we may conclude that (37) holds for the K-relay channel with the additional conditioning on X_K added into the equations for the $(K-1)$-relay channel. This gives us (33) for $0 \le k \le K-1$.

Let $\widetilde{X}_i$ denote the output of transmitter i and $\widetilde{Z}_i$ denote the noise added at stage i in the $(K-1)$-relay channel we are to build. Set $\widetilde{Z}_i \sim Z_i$ for all $0 \le i \le K-1$. For the transmitter outputs, set $\widetilde{X}_i \sim (X_i - \mathsf{E}(X_i|X_K))$.[1] Then for $0 \le i \le K-1$

$$\mathsf{E}\,\widetilde{X}_i^2 = \mathsf{E}\,X_i^2 + \mathsf{E}\,\mathsf{E}^2(X_i|X_K) - 2\mathsf{E}\left(X_i\mathsf{E}\left(X_i|X_K\right)\right) \tag{38}$$

but

$$\begin{aligned}\mathsf{E}\left(X_i\mathsf{E}\left(X_i|X_K\right)\right) &= \mathsf{E}\,\mathsf{E}\left(X_i\mathsf{E}\left(X_i|X_K\right)|X_K\right) &(39)\\ &= \mathsf{E}\left(\mathsf{E}\left(X_i|X_K\right)\mathsf{E}\left(X_i|X_K\right)\right) &(40)\\ &= \mathsf{E}\,\mathsf{E}^2(X_i|X_K). &(41)\end{aligned}$$

Thus,

$$\mathsf{E}\,\widetilde{X}_i^2 = \mathsf{E}\,X_i^2 - \mathsf{E}\,\mathsf{E}^2(X_i|X_K) = P_0\sum_{j=i}^{K-1}\alpha_{i,j}. \tag{42}$$

Thus, once we have selected and fixed $\alpha_{i,K}$'s for all $0 \le i \le K-1$ the remaining valid choices of $\{\alpha_{i,j}\}$ for $0 \le i \le j \le K-1$ are in one-to-one correspondence with every valid choice of $\{\widetilde{\alpha}_{i,j}\}$ for the $(K-1)$-relay channel, where the output power of transmitter i is constrained not to exceed $\mathsf{E}\widetilde{X}_i^2$.

Let us now consider the quantity

$$\mathsf{E}\mathsf{Var}(\widetilde{X}_0 + \cdots + \widetilde{X}_k | \widetilde{X}_{k+1}, \ldots, \widetilde{X}_{K-1})$$

for some $0 \le k \le K-1$. To simplify notation in the next few lines, we define

$$A_k \stackrel{\text{def}}{=} X_0 + \cdots + X_k \tag{43}$$

and

$$B(X_K) \stackrel{\text{def}}{=} (\widetilde{X}_{k+1}, \ldots, \widetilde{X}_{K-1}) \tag{44}$$

where we note explicitly that B depends on X_K. Then

$$\widetilde{X}_0 + \cdots + \widetilde{X}_k = \sum_{i=0}^{k} X_i - \mathsf{E}(X_i|X_K) = A_k - \mathsf{E}(A_k|X_K) \tag{45}$$

and

$$\begin{aligned}&\mathsf{Var}\left(\widetilde{X}_0 + \cdots + \widetilde{X}_k | B(X_K)\right)\\ &\quad = \mathsf{Var}\left(A_k - \mathsf{E}(A_k|X_K)|B(X_K)\right) &(46)\\ &\quad = \mathsf{Var}\left(A_k - \bar{A}_k(X_K)|B(X_K)\right) &(47)\\ &\quad = \mathsf{Var}\left(A_k|B(X_K), X_K\right) &(48)\end{aligned}$$

where $\bar{A}_k(X_K)$ denotes the conditional mean of A on X_K.

Thus,

$$\begin{aligned}&\mathsf{E}\,\mathsf{Var}\left(\widetilde{X}_0 + \cdots + \widetilde{X}_k | \widetilde{X}_{k+1}, \ldots, \widetilde{X}_{K-1}\right)\\ &\quad = \mathsf{E}\,\mathsf{Var}\left(X_0 + \cdots + X_k | \widetilde{X}_{k+1}, \ldots, \widetilde{X}_{K-1}, X_K\right) &(49)\\ &\quad = \mathsf{E}\,\mathsf{Var}\left(X_0 + \cdots + X_k | X_{k+1}, \ldots, X_{K-1}, X_K\right) &(50)\end{aligned}$$

where the outer expectations are taken with respect to the collection of random variables that are conditioned upon in the inner expectations (i.e., the variances).

Recall the induction hypothesis: for every choice of $\{\alpha_{i,j}\}$ (37) holds. From this and the above equation we can conclude

[1]We note that in the expression of the form $\mathsf{E}\,(\cdot|\cdot)$ the expectation is always with respect to the collection of the arguments to the left of the vertical bar and the result is a function of the arguments to the right of the vertical bar.

that for any fixed choice of $\{\alpha_{i,K}\}$ as given by (35) and any choice of $\{\alpha_{i,j}\}$ $0 \le i \le j \le K-1$ satisfying (36) (33) holds for $0 \le k \le K-1$.

We pause here to return to the definition of $\alpha_{i,j}$'s used in the converse. We had previously only defined these explicitly for $j = K$ in (35). The inductive proof presented above allows us to understand what the definition would be for $j < K$. We simply need to consider the j-relay subchannel and define $\alpha_{i,j}$ from that channel analogously to (35). We can illustrate this by defining

$$\alpha_{i,K-1} \stackrel{\text{def}}{=} \frac{1}{P_0} \mathsf{E}\,\mathsf{E}^2(\widetilde{X}_i|X_{K-1}) = \frac{1}{P_0} \mathsf{E}\,\mathsf{E}^2\left(X_i - \mathsf{E}(X_i|X_K)|X_{K-1}\right). \tag{51}$$

As the value of j becomes lower, the nesting of conditional expectations as shown in (51) grows and this makes the explicit definition of $\alpha_{i,j}$ for all j more cumbersome.

To complete the proof of the converse it remains to show that

$$\mathsf{E}\,\mathsf{Var}(X_0 + \cdots + X_K) \le P_0 \sum_{j=0}^{K} \left(\sum_{i=0}^{j} \sqrt{\alpha_{i,j}} \right)^2 \tag{52}$$

where we note that the outer expectation is in this case redundant since the variance is unconditional.

To show (52), take

$$\widehat{X} \stackrel{\text{def}}{=} X_0 + \cdots + X_{K-1}. \tag{53}$$

Then, from (33) applied for $k = K-1$, we have

$$\mathsf{E}\,\mathsf{Var}\,(\widehat{X}|X_K) \le P_0 \sum_{j=0}^{K-1} \left(\sum_{i=0}^{j} \sqrt{\alpha_{i,j}} \right)^2. \tag{54}$$

It therefore suffices to show that

$$\mathsf{Var}\,(\widehat{X} + X_K) \le \mathsf{E}\,\mathsf{Var}\,(\widehat{X}|X_K) + P_0 \left(\sum_{i=0}^{K} \sqrt{\alpha_{i,K}} \right)^2. \tag{55}$$

We then have

$$\begin{aligned}
&\mathsf{E}\,(\widehat{X} + X_K)^2 - \mathsf{E}\,\mathsf{Var}\,(\widehat{X}|X_K) && (56)\\
&= \mathsf{E}\,(\widehat{X}^2) + 2\mathsf{E}\,(\widehat{X}X_K) + \mathsf{E}\,\left(X_K^2\right) - \mathsf{E}\,(\widehat{X}^2) + \mathsf{E}\,\mathsf{E}^2(\widehat{X}|X_K) && (57)\\
&= \mathsf{E}\,\left(X_K^2\right) + 2\mathsf{E}\,(X_K\,\mathsf{E}\,\widehat{X}|X_K) + \mathsf{E}\,\mathsf{E}^2(\widehat{X}|X_K) && (58)\\
&= \mathsf{E}\,\left(X_K + \mathsf{E}\,\widehat{X}|X_K\right)^2 = \mathsf{E}\,\mathsf{E}^2\left(\sum_{i=0}^{K} X_i|X_K\right) && (59)\\
&= \sum_{i=0}^{K} \mathsf{E}\,\mathsf{E}^2(X_i|X_K) + 2\sum_{i=0,i<j}^{K} \mathsf{E}\,[(\mathsf{E}\,X_i|X_K)(\mathsf{E}\,X_j|X_K)] && (60)\\
&= P_0 \sum_{i=0}^{K} \alpha_{i,K} + 2\sum_{i=0,i<j}^{K} \mathsf{E}\,[(\mathsf{E}X_i|X_K)(\mathsf{E}\,X_j|X_K)]. && (61)
\end{aligned}$$

Applying Cauchy–Schwartz and then Jensen's inequalities to each term of the second summation above we have (for each term)

$$\begin{aligned}
&\mathsf{E}\,[(\mathsf{E}\,X_i|X_K)(\mathsf{E}\,X_j|X_K)]\\
&\le \left[\mathsf{E}\sqrt{\mathsf{E}^2 X_i|X_K}\right]\left[\mathsf{E}\sqrt{\mathsf{E}^2 X_j|X_K}\right] && (62)\\
&\le \sqrt{\mathsf{E}\,\mathsf{E}^2 X_i|X_K}\sqrt{\mathsf{E}\,\mathsf{E}^2 X_j|X_K} && (63)\\
&= P_0\sqrt{\alpha_{i,K}}\sqrt{\alpha_{j,K}}. && (64)
\end{aligned}$$

Putting all of this together we have

$$\begin{aligned}
&\mathsf{E}(\widehat{X} + X_K)^2 - \mathsf{E}\mathsf{Var}(\widehat{X}|X_K)\\
&\le P_0 \sum_{i=0}^{K} \alpha_{i,K} + 2P_0 \sum_{i=0,i<j}^{K} \sqrt{\alpha_{i,K}}\sqrt{\alpha_{j,K}} && (65)\\
&= P_0 \left(\sum_{i=0}^{K} \sqrt{\alpha_{i,K}} \right)^2 && (66)
\end{aligned}$$

and (55) follows since $\mathsf{Var}\,(\widehat{X} + X_K) \le \mathsf{E}\,(\widehat{X} + X_K)^2$.

This completes the proof of the converse □

We have now expressed the capacity of the multiple-relay degraded Gaussian channel as a max-min problem. Given a general distribution of power between the transmitter and the relays, the optimum in (26) may be very difficult to determine. A general closed-form expression would certainly become too unruly for more than a few relays. However, as we shall see in Section IV, the problem of maximizing the power allocation between relays under a total power constraint lends itself to a simple iterative solution. Exploring this solution permits us to illustrate the benefits inherent in exploiting the relays for communication when these are available.

IV. Optimum Power Allocation for Multiple Relays

In this section, we consider the problem of optimum power allocation between relays in the multiple-relay degraded channel. We start with a K-relay degraded Gaussian channel. We wish to determine $\beta_1, \ldots, \beta_K$ subject to

$$1 + \beta_1 + \cdots + \beta_K \le \frac{P_T}{P_0} \tag{67}$$

such that the capacity of the channel as given by (26) is maximized. In what follows, we fix P_T, the total power available to all the transmitters, and vary the power of each transmitter, including P_0. We begin with the following theorem.

Theorem 3 (Optimum Power Allocation): Consider a K-relay degraded Gaussian channel with the power allocated among the relays is such a way that the capacity of the overall channel is optimized subject to (67). Then the optimum in (26) is achieved by setting

$$\alpha_{0,j} = \alpha_{1,j} = \cdots = \alpha_{j,j} \stackrel{\text{def}}{=} \alpha_j.$$ □

We note that if

$$\alpha_{0,j} = \alpha_{1,j} = \cdots = \alpha_{j,j} \stackrel{\text{def}}{=} \alpha_j$$

then $\beta_k = \sum_{j=k}^{K} \alpha_j$ and, in particular, $\sum_{j=0}^{K} \alpha_j = \beta_0 \equiv 1$. Thus, if Theorem 3 holds, in order to find the optimum power allocation scheme for a K-relay channel, it suffices to determine the values of $\alpha_0, \dots, \alpha_K$ subject to $\sum_{j=0}^{K} \alpha_j = 1$. The values of $\beta_1, \dots, \beta_K$ are then determined directly by $\beta_k = \sum_{j=k}^{K} \alpha_j$. Additionally

$$\beta_k > \beta_{k+1}$$

as long as $\alpha_k > 0 \,\forall\, 0 \le k \le K$. As we shall see, this holds as long as $N_k > 0 \,\forall\, k$.

Proof: The proof of Theorem 3 is an application of the following principle. Let $\{x_m\}$ be a collection of M nonnegative variables. Then the solution to the problem

$$\min \sum_{m=1}^{M} x_m^2 \text{subject to} \sum_{m=1}^{M} x_m \ge K_1$$

or the solution to its dual

$$\max \sum_{m=1}^{M} x_m \text{subject to} \sum_{m=1}^{M} x_m^2 \le K_2$$

is to set $x_m = x, m = 1, \dots, M$ in such a way that the constraint is satissried with equality.

We proceed by induction. We proved the initial step for the single-relay channel in Section II. Recall that we showed that if the power is allocated optimally between the relay and the transmitter, then we have $\beta_1 = 1 - \alpha_1 \Rightarrow \alpha_{1,1} = \alpha_{0,1}$.

For the induction step, assume that the theorem holds for $K - 1$ relays. Let us use the coding scheme utilized in Theorem 2 to achieve capacity. Consider adding one more relay stage as the last relay. As we did in the proof of Theorem 2, this requires transmitter i to allocate $\alpha_{i,K}$ power for the transmission of the codebook X_K (the encoded S_K). Let us assume that $\alpha_{i,K} \equiv \alpha_K \,\forall\, 0 \le i \le K$.

Define

$$\boldsymbol{C}_{K-1}(\alpha_K) \stackrel{\text{def}}{=} \max_{\{\alpha_{i,j}\}:0 \le j \le K-1} \mathcal{C}_{K-1}(\bar{\alpha}). \tag{68}$$

This quantity is the achievable rate to receiver $K - 1$ and we explicitly indicate that the minimum of rates achievable to all the receivers up to $K - 1$ now depends on how much power is allocated for the transmission of X_K. Recall that $x_{K,i}$ (the realization of X_K at transmission block i) depends on w_{i-K}, which is assumed known by receivers $0, \dots, K-1$ at transmission block i. Then, any power allocated for the transmission of X_K is wasted as far as communication to receivers $0, \dots, K-1$ is concerned. For any fixed α_K, we can view the optimization problem defined in (68) as a $(K - 1)$-relay channel capacity problem with total power $P_T - (K+1)\alpha_K$. Thus, by induction assumption, the optimum in (68) is achieved when we set $\alpha_{i,j} = \alpha_j \,\forall\, j \le i \le K, \forall\, 0 \le j \le K-1$.

The rate achievable for transmission of X_K is given by $\mathcal{R}_K(\alpha_K)$ where $\mathcal{R}_K$ is defined by (24) and once again we make the dependence on α_K explicit.

The rate achievable using this scheme is then given by

$$\boldsymbol{C}_K = \max_{\alpha_K} \min \left[\boldsymbol{C}_{K-1}(\alpha_K), \mathcal{R}_K(\alpha_K)\right]. \tag{69}$$

Can we do better than this rate? In order to answer affirmatively we must find a way to increase both $\boldsymbol{C}_{K-1}(\alpha_K)$ and $\mathcal{R}_K(\alpha_K)$. Since we already chose the best α_K, the only option open to us is to violate the assumption that $\alpha_{i,K} \equiv \alpha_K \,\forall\, 0 \le i \le K$.

Recall now that $\mathcal{R}_K = \sum_{k=0}^{K} \widehat{R}_k$ with $\widehat{R}_k$ defined by (28). Among all the $\widehat{R}_k$'s only $\widehat{R}_K$ depends on $\alpha_{i,K}$'s. Using the argument presented at the beginning of this proof, we conclude that the choice of $\alpha_{i,K} \equiv \alpha_K$ maximizes $\widehat{R}_K$ for a given total power utilization. Alternatively, using the second part of the statement we can conclude that the choice of $\alpha_{i,K} \equiv \alpha_K$ minimizes the power used to achieve a fixed $\widehat{R}_K$. Thus, if we violate the condition that $\alpha_{i,K} \equiv \alpha_K$, then either we reduce $\widehat{R}_K$, reducing $\mathcal{R}_K$ in the process; or, if we keep $\widehat{R}_K$ constant, we must reduce the total power available for transmission of $X_0, \dots, X_{K-1}$.

Since, by the inductive assumption, $\boldsymbol{C}_{K-1}$ is always achieved using the optimal allocation of the available power (i.e., other than that already allocated to X_K via $\alpha_{i,K}$), reducing the amount of this power must reduce $\boldsymbol{C}_{K-1}$. We therefore conclude that we cannot find a way to do better than (69). This completes the proof of the theorem. □

To derive an explicit expression for the capacity of a K-relay channel as given by (24)–(26) we need to find a set $\{\alpha_{i,j}\}$ such that $\mathcal{R}_0 = \mathcal{R}_1 = \cdots = \mathcal{R}_K$. While the problem is in general very difficult, under the assumption of optimum power allocation we can use Theorem 3 to simplify it significantly. We begin by rewriting (24) under the assumption that $\alpha_{i,j} = \alpha_j$. This results in

$$\mathcal{R}_k = C\left(P_0 \frac{\sum_{j=0}^{k}(j+1)^2\alpha_j}{\sum_{j=0}^{k} N_j}\right). \tag{70}$$

Setting $\mathcal{R}_k = \mathcal{R}_{k-1}$ we get

$$\frac{\sum_{j=0}^{k}(j+1)^2\alpha_j}{\sum_{j=0}^{k} N_j} = \frac{\sum_{j=0}^{k-1}(j+1)^2\alpha_j}{\sum_{j=0}^{k-1} N_j} \tag{71}$$

which yields

$$\frac{(k+1)^2\alpha_k}{\sum_{j=0}^{k} N_j} = \sum_{j=0}^{k-1}(j+1)^2\alpha_j \left(\frac{1}{\sum_{j=0}^{k-1} N_j} - \frac{1}{\sum_{j=0}^{k} N_j}\right) \tag{72}$$

$$= \sum_{j=0}^{k-1}(j+1)^2\alpha_j \frac{N_k}{\left(\sum_{j=0}^{k-1} N_j\right)\left(\sum_{j=0}^{k} N_j\right)}. \tag{73}$$

Thus,

$$\alpha_k = \frac{1}{(k+1)^2}\nu_k \left(\sum_{j=0}^{k-1}(j+1)^2\alpha_j\right) \tag{74}$$

and one can easily check that if we set $K = 1$ we obtain the solution obtained in Section II by direct solution of the single-relay optimization problem. Additionally, we note that if $N_k > 0$, $\nu_k > 0$ and $\alpha_k > 0$ for all $0 \le k \le K$.

To obtain an explicit solution for α_k, we let

$$s_k \stackrel{\text{def}}{=} \sum_{j=0}^{k}(j+1)^2\alpha_j, \qquad \text{for } k > 0 \text{ and } s_0 \stackrel{\text{def}}{=} \alpha_0. \tag{75}$$

Then from (74) we get

$$s_k = (1+\nu_k)s_{k-1} = \alpha_0 \prod_{j=1}^{k}(1+\nu_j). \tag{76}$$

Expanding this recursion back to $k = 0$ and substituting into (74) we get

$$\alpha_k = \frac{1}{(k+1)^2}\nu_k \prod_{j=1}^{k-1}(1+\nu_j)\alpha_0 = \frac{1}{(k+1)^2}\frac{N_k}{N_0}\alpha_0 \quad (77)$$

where the second equality follows since

$$\prod_{j=1}^{k}(1+\nu_j) = \frac{\sum_{j=0}^{k} N_j}{N_0}. \quad (78)$$

From (77), we can obtain an explicit expression for the optimal α_k's in terms of the relative noise powers, this is simply

$$\alpha_k = \frac{1}{\Gamma_K}\frac{N_k}{(k+1)^2} \quad (79)$$

where the normalization factor Γ_K is given by

$$\Gamma_K = \sum_{k=0}^{K}\frac{N_k}{(k+1)^2}. \quad (80)$$

Using (79), we can now write

$$\beta_k = \frac{1}{\Gamma_K}\sum_{j=k}^{K}\frac{N_j}{(j+1)^2} \quad (81)$$

and

$$\frac{P_T}{P_0} = \frac{1}{\Gamma_K}\sum_{k=0}^{K}\sum_{j=k}^{K}\frac{N_j}{(j+1)^2} = \frac{1}{\Gamma_K}\sum_{k=0}^{K}\frac{N_k}{k+1}. \quad (82)$$

By selecting $\alpha_0,\ldots,\alpha_K$ such that the maximum in (26) is achieved, we are ensuring that $\mathcal{R}_0 = \mathcal{R}_1 = \cdots = \mathcal{R}_K$. Thus, the capacity of the K-relay channel can be written simply as

$$\begin{aligned}\boldsymbol{C}_K &= \mathcal{R}_0 = C\left(\alpha_0\frac{P_0}{N_0}\right)\\ &= C\left(J_1(K,N_0,\ldots,N_K)\frac{P_T}{\sum_{k=0}^{K}N_k}\right)\end{aligned} \quad (83)$$

where $J_1(K,N_0,\ldots,N_K)$ is the maximum SNR improvement obtained by distributing a fixed amount of power over K relays. From (82) and (83) this is given by

$$J_1(K,N_0,\ldots,N_K) \stackrel{\text{def}}{=} \frac{\sum_{k=0}^{K}N_k}{\sum_{k=0}^{K}\frac{N_k}{k+1}}. \quad (84)$$

Up to now we have taken the point of view of distributing fixed total power among K relays. An alternative model, which is applicable in many cases, is that of a fixed-power transmitter and a number of other nodes that allocate a certain amount of the power available to them to relay the message of interest. One might think of this problem in terms of getting the most "bang" for the power expenditure used for relaying. In this case, we have

$$P_T = f[K]P_0, \qquad f[0] = 1. \quad (85)$$

We note that if we fix K, then the problem is fundamentally unchanged. The difference is that we would like to determine the best possible $f[K]$ for a given noise distribution. This can be evaluated directly from (82) as

$$f[K] = \frac{1}{\Gamma_K}\sum_{k=0}^{K}\frac{N_k}{k+1} = \frac{\sum_{k=0}^{K}\frac{N_k}{k+1}}{\sum_{k=0}^{K}\frac{N_k}{(k+1)^2}}. \quad (86)$$

In this case, it makes sense to write the resulting K-relay capacity as

$$\boldsymbol{C}_K = C\left(J_2(K,N_0,\ldots,N_K)\frac{P_0}{\sum_{k=0}^{K}N_k}\right) \quad (87)$$

where $J_2(K,N_0,\ldots,N_K)$ is maximum SNR improvement obtained from using K relays and this is given by

$$J_2(K,N_0,\ldots,N_K) \stackrel{\text{def}}{=} \frac{\sum_{k=0}^{K}N_k}{\sum_{k=0}^{K}\frac{N_k}{(k+1)^2}}. \quad (88)$$

A third point of view on the optimum power allocation is the optimum power allocation when the total available power varies as some function $p[K]$ of the total number of available relays. However, given the number of relays, we are free to distribute the power among the relays. Once again, given the number of available relays, the total available power is fixed. Thus, the fundamental problem is the same and (82) applies.

In this case, both definitions of SNR improvement used above make sense, that is, one may be interested in J_1'—the improvement obtained over pooling all the available power at the transmitter. Alternatively, one may be interested in J_2'—the improvement obtained if only the transmitter, with power $p[0]$ were available for transmission. In this case, $J_1' = J_1$ and J_2' is given by

$$J_2'(K,N_0,\ldots,N_K) = J_1(K,N_0,\ldots,N_K)\frac{p[K]}{p[0]}. \quad (89)$$

As an example, let us consider the channel where the noise powers behave according to a power law, i.e., $N_k = N_0(k+1)^\gamma$. The case when the noise powers are constant is a special case of this channel when $\gamma = 0$. In this case, we note that

$$J_1(K,N_0,\ldots,N_K) = \frac{\sum_{k=0}^{K}(k+1)^\gamma}{\sum_{k=0}^{K}(k+1)^{\gamma-1}} \quad (90)$$

$$J_2(K,N_0,\ldots,N_K) = \frac{\sum_{k=0}^{K}(k+1)^\gamma}{\sum_{k=0}^{K}(k+1)^{\gamma-2}} \quad (91)$$

and

$$\frac{P_T}{P_0} = \frac{\sum_{k=0}^{K}(k+1)^{\gamma-1}}{\sum_{k=0}^{K}(k+1)^{\gamma-2}}. \quad (92)$$

We can also make the following observations about the power-noise-law channel.

- If the total power is fixed, J_1 increases without bound as long as $\gamma \geq -1$. Moreover, if $\gamma \geq 0$, the transmitter's power goes to 0 as K goes to infinity.
- For the special case when all the noise powers are equal, $J_1 = O(\log K)$ while $J_2 = O(K)$. In fact, since [4]

$$\lim_{K\to\infty}\sum_{k=1}^{K}\frac{1}{k^2} = \frac{\pi^2}{6}$$

  the asymptotic slope of J_2 is equal to $\frac{6}{\pi^2}$.

Before we leave this topic, we illustrate the SNR improvement obtained from the optimal power distribution when all the noise stages have equal powers. Fig. 2 shows J_1 as a function of the number of relays and Fig. 3 shows J_2. From Fig. 3 it is clear

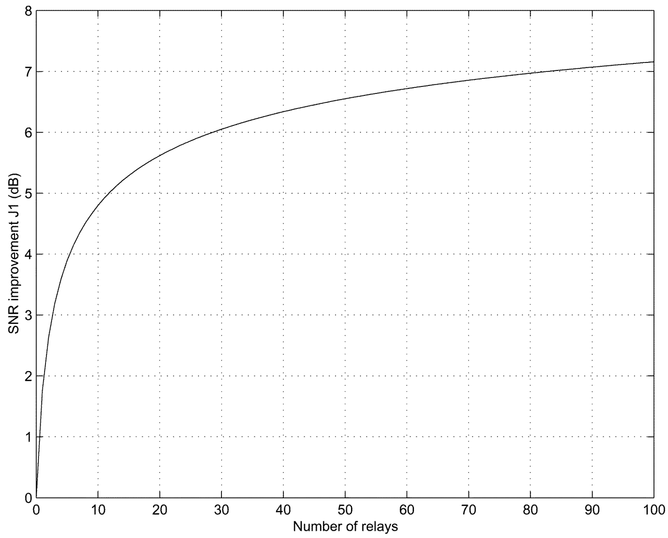

Fig. 2. SNR improvement J_1 in decibels with equal noise powers.

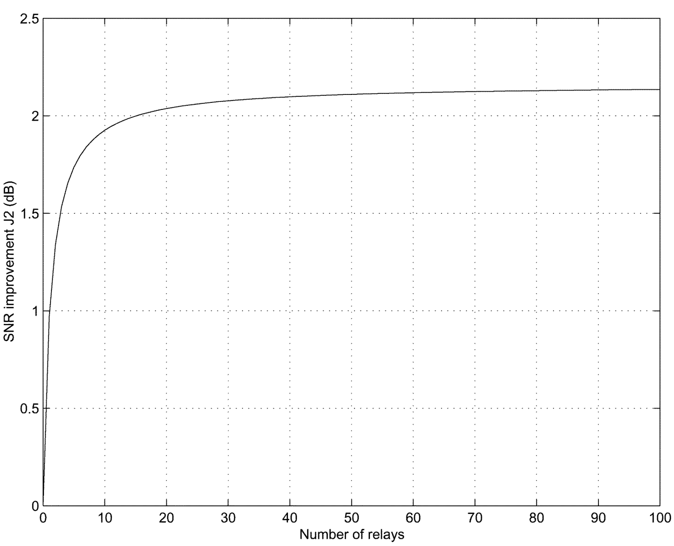

Fig. 3. SNR improvement J_2 in decibels with equal noise powers.

that the improvement levels off at $\frac{\pi^2}{6} = 2.16$ dB as the number of relays increases.

V. Wideband Performance in the Gaussian Degraded Relay Channel

In a recent paper [12], Verdú demonstrated that when bandwidth is not a free resource, second-order effects must be taken into account when wideband performance of communication systems is considered. Using these results it was shown [1] that in the wideband regime time-division multiple access (TDMA) is strictly suboptimal for both the multiple-access Gaussian channel and the broadcast Gaussian channel. This is so despite that fact that it does achieve the optimal $\frac{E_b}{N_0}_{\min}$ defined as

$$\frac{E_b}{N_0}_{\min} = \lim_{\mathsf{SNR}\to 0} \frac{\mathsf{SNR}}{\mathcal{C}(\mathsf{SNR})} = \frac{\log_e 2}{\dot{\mathcal{C}}(0)} \tag{93}$$

where $\mathcal{C}$ is the capacity of the channel as a function of SNR.

However, in the case of the degraded Gaussian relay channel, TDM does not even achieve the same $\frac{E_b}{N_0}_{\min}$ as the information theoretically optimal communication scheme. We illustrate this using the single-relay channel. Define $Q \stackrel{\text{def}}{=} \frac{P_0}{N_0}$ to be the SNR between the transmitter and the first relay stage. While Q may not always be the best measure of the SNR, it is sufficient for our purposes since any reasonable measure of SNR should go to 0 as Q goes to 0.

Then for the single-relay channel, the capacity as a function of Q is given by (6) as

$$\mathcal{C}_1 = C\left(\alpha_1^* Q\right). \tag{94}$$

Thus, $\lim_{Q\to 0} \mathcal{C}(Q) = \alpha_1^*$, resulting in

$$\frac{E_b}{N_0}_{\min} = \frac{\log_e 2}{\alpha_1^*}. \tag{95}$$

Turning now to TDM, we note that we can alternate between two modes of communication: the transmitter can transmit directly to the receiver or the transmitter can transmit to the relay which then relays the message to the receiver. In the second instance, the transmitter and relay transmissions have to be time-multiplexed as well. Let $\mathcal{C}_d(Q)$ be the capacity function for direct transmitter-to-receiver communication and $\mathcal{C}_r(Q)$ be the capacity function for communication using the relay. Then the TDM capacity function is given by

$$\mathcal{C}_{\mathrm{TDM}}(Q) = \max_{0\le\lambda\le 1} \left[\lambda \mathcal{C}_d(Q) + (1-\lambda)\,\mathcal{C}_r(Q)\right] \tag{96}$$

$$= \max\left[\mathcal{C}_d(Q), \mathcal{C}_r(Q)\right]. \tag{97}$$

Thus, the optimal strategy is to either always use a relay or to never use it. Which of the two options is to be chosen in a specific scenario depends on the values of β_1 and ν_1. In order to specify this more precisely, we need to examine the functions $\mathcal{C}_d$ and $\mathcal{C}_r$ themselves.

$\mathcal{C}_d$ is the capacity function for point-to-point communication without a relay, thus, it is given by

$$\mathcal{C}_d(Q) = C\left(Q \frac{1}{1+\nu_1}\right). \tag{98}$$

If we are using the relay to communicate, then we must time-share the channel between the transmitter and the relay. Thus, the capacity function in this case is given by

$$\mathcal{C}_r(Q) = \max_{0\le\lambda\le 1} \min\left[\lambda C(Q), (1-\lambda)\, C(\gamma Q)\right] \tag{99}$$

where we use the shorthand notation $\gamma \stackrel{\text{def}}{=} \frac{\beta_1}{\nu_1}$.

The minimax point of (99) is achieved by an λ^* such that

$$\lambda^* C(Q) = (1-\lambda^*) C(\gamma Q). \tag{100}$$

Solving this we get

$$\lambda^* = \frac{C(\gamma Q)}{C(Q) + C(\gamma Q)} \tag{101}$$

which gives us

$$\mathcal{C}_r(Q) = \frac{C(Q)C(\gamma Q)}{C(Q) + C(\gamma Q)} = \frac{1}{\frac{1}{C(Q)} + \frac{1}{C(\gamma Q)}}. \tag{102}$$

Since both $\mathcal{C}_d$ and $\mathcal{C}_r$ are monotonically increasing and concave, (93) and (96) give us

$$\frac{E_b}{N_0}_{\min}^{\mathrm{TDM}} = \frac{\log_e 2}{\max\left[\dot{\mathcal{C}}_d(0), \dot{\mathcal{C}}_r(0)\right]}. \tag{103}$$

Finding the derivatives in (103) we have

$$\dot{\mathcal{C}}_d(0) = \frac{1}{1+\nu_1} \dot{C}(0) = \frac{1}{1+\nu_1} \tag{104}$$

$$\dot{\mathcal{C}}_r(Q) = \frac{\dot{C}(Q)C^2(\gamma Q) + \gamma\dot{C}(\gamma Q)C^2(Q)}{(C(Q)+C(\gamma Q))^2} \tag{105}$$

$$= \frac{\dot{C}(Q)\frac{C(\gamma Q)}{C(Q)} + \gamma\dot{C}(\gamma Q)\frac{C(Q)}{C(\gamma Q)}}{2+\frac{C(\gamma Q)}{C(Q)}+\frac{C(Q)}{C(\gamma Q)}}. \tag{106}$$

To find $\dot{\mathcal{C}}_r(0)$ we note that $\lim_{Q\to 0}\frac{C(\gamma Q)}{C(Q)} = \gamma$. Using this fact we get

$$\dot{\mathcal{C}}_r(0) = \frac{1+\gamma}{2+\gamma+\frac{1}{\gamma}} = \frac{\gamma}{1+\gamma} = \frac{1}{1+\frac{\nu_1}{\beta_1}}. \tag{107}$$

Thus,

$$\max\left[\dot{\mathcal{C}}_d(0), \dot{\mathcal{C}}_r(0)\right] = \frac{1}{1+\nu_1\min\left(1, \frac{1}{\beta_1}\right)}. \tag{108}$$

We note that α_1^* is strictly greater than $\frac{1}{1+\nu_1\min(1,\frac{1}{\beta_1})}$ as long as $0 < \nu < \infty$ and $0 < \gamma < \infty$. To see this, consider first $\beta_1 \geq \nu_1$. Then, as we showed in Section II, $\alpha_1^* = 1$. Since $1+\nu_1\min(1,\frac{1}{\beta_1}) > 1$ the statement follows. Additionally note that $\alpha_1^* > \frac{1}{1+\nu_1}$ as long as $\beta_1 > 0$. Thus, it remains to consider the case when

$$\max\left[\dot{\mathcal{C}}_d(0), \dot{\mathcal{C}}_r(0)\right] = \frac{1}{1+\frac{\nu_1}{\beta_1}}$$

that is, when $1 < \beta_1 < \nu_1$. Rewriting (17) for ν_1 we have

$$\nu_1 = \frac{\left(\sqrt{\beta_1}+\sqrt{1-\alpha_1^*}\right)^2}{\alpha_1^*} > \frac{\beta_1}{\alpha_1^*}. \tag{109}$$

Then

$$\frac{1}{1+\frac{\nu_1}{\beta_1}} < \frac{1}{1+\frac{1}{\alpha_1^*}} = \frac{\alpha_1^*}{\alpha_1^*+1} < \alpha_1^*. \tag{110}$$

In fact, since the optimal communication strategy for the relay channel relies on coherent combining at the receiver, it is intuitive that any communication strategy that cannot provide this would not be able to attain the optimal $\frac{E_b}{N_0}_{\min}$. Moreover, the more relays we can use, the more signals we have to coherently combine at the receivers and thus the advantage of the optimal scheme over TDM grows with the number of available relays.

VI. Conclusion

In this paper, we extended the capacity result for a single-relay degraded Gaussian channel to a channel with multiple relays. In doing so, we demonstrated that a previously obtained achievable rate [13] for this channel is the capacity. We used our inductive proof of the achievability of the capacity region to determine the power allocation scheme that maximizes the capacity that results and showed that this is given by a particularly simple expression in terms of the noise power.

We note, however, that the degraded channel model is not generally the best model for real-life channels. A more accurate model of real-world channels would have independent noise sources at the input to the receiver and to each relay. Unfortunately, the capacity of such a nondegraded channel has not been solved even for a single-relay case. This is one of the several fundamental problems in network information theory that remains unsolved. Our results provide only an achievable rate region for a general relay channel.

Because of the degraded nature of the model considered, it is also challenging to use it to account for several important real-world phenomena, such as signal propagation properties in nonlinearly arranged communication networks and fading. Some attempts at this have been made in [13], however, much more work remains to be done.

The results obtained in this paper, as well as the results of the several other investigations (see, e.g., [6], [8]), do demonstrate that there are potentially significant advantages to be gained from using relays, provided that this is done is a way that makes sense from an information-theoretic considerations. In particular, it appears to be critical to be able to combine several simultaneous transmissions. In this sense, any scheme utilizing an orthogonal partition of the channel into subchannels is likely to yield suboptimal results. Our results on TDM schemes in Section IV support this assertion.

References

[1] G. Caire, D. Tuninetti, and S. Verdú, "Suboptimality of TDMA in the low power regime," *IEEE Trans. Inform. Theory*, to be published.

[2] A. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 572–584, Sept. 1979.

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

[4] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, Eds., *Tables of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic, 1994.

[5] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1877–1894, Aug. 2003.

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

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

[8] B. E. Schein, "Distributed coordination in network information theory," Ph.D. dissertation, MIT, Cambridge, MA, 2001.

[9] E. C. van der Muelen, "Transmission of information in a T-terminal discrete memoryless channel," Ph.D. dissertation, Univ. Calif., Berkeley, 1968.

[10] ——, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 12–154, 1971.

[11] ——, "A survey of multi-way channels in information theory: 1961–1976," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 1–37, Jan. 1977.

[12] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.

[13] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communciation: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, to be published.

[14] C. Zeng, F. Kuhlmann, and A. Buzo, "Achievability proof of some multiuser channel coding theorems using backward decoding," *IEEE Trans. Inform. Theory*, pp. 1160–1165, Nov. 1989.