

On the Capacity of Large Gaussian Relay Networks

Michael Gastpar, *Member, IEEE*, and Martin Vetterli, *Fellow, IEEE*

Abstract—The capacity of a particular large Gaussian relay network is determined in the limit as the number of relays tends to infinity. Upper bounds are derived from cut-set arguments, and lower bounds follow from an argument involving uncoded transmission. It is shown that in cases of interest, upper and lower bounds coincide in the limit as the number of relays tends to infinity. Hence, this paper provides a new example where a simple cut-set upper bound is achievable, and one more example where uncoded transmission achieves optimal performance. The findings are illustrated by geometric interpretations.

The techniques developed in this paper are then applied to a sensor network situation. This is a network joint source–channel coding problem, and it is well known that the source–channel separation theorem does not extend to this case. The present paper extends this insight by providing an example where separating source from channel coding does not only lead to suboptimal performance—it leads to an *exponential* penalty in performance scaling behavior (as a function of the number of nodes). Finally, the techniques developed in this paper are extended to include certain models of *ad hoc* wireless networks, where a capacity scaling law can be established: When all nodes act purely as relays for a single source–destination pair, capacity grows with the logarithm of the number of nodes.

Index Terms—Capacity, CEO problem, joint source–channel coding, network, relay, sensor network, separation theorem, uncoded transmission.

I. INTRODUCTION

THE relay channel (both with a single relay and with multiple relays) has been introduced by van der Meulen in his Ph.D. dissertation [3] and in [4]. Key results for the single-relay channel have been found by Cover and El Gamal [5]. Their capacity results are restricted to the so-called *degraded* relay channel. This restriction is considerably stronger than the common notion of degradedness in the case of broadcast channels introduced in [6]. For example, it is considered to be a weak model for the wireless relay channel. In extension of the single-relay channel, various relay network models

Manuscript received September 17, 2002; revised October 11, 2004. This work was supported in part by the National Science Foundation under Award CCF-0347298 (CAREER) and Grant CNS-0326503, and by the Swiss National Science Foundation in the framework of the National Competence Center in Research for Mobile Information and Communication Systems (<http://www.mics.org>).

The material in this paper was presented in part at the IEEE INFOCOM, New York, June 2002 and in part at the IEEE International Symposium on Information Theory, Lausanne, Switzerland, June/July 2002.

M. Gastpar is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720-1770 USA (e-mail: gastpar@eecs.berkeley.edu).

M. Vetterli is with the Institute of Communication Systems, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720 USA (e-mail: martin.vetterli@epfl.ch).

Communicated by L. Tassiulas, Associate Editor for Communication Networks.

Digital Object Identifier 10.1109/TIT.2004.842566

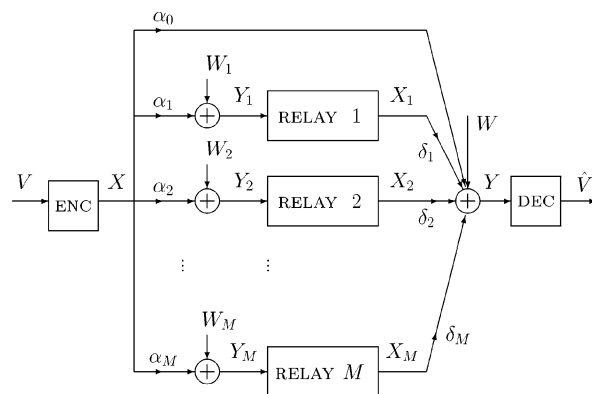


Fig. 1. The considered Gaussian relay network.

have been studied in the literature, some of the most recent examples being [7]–[14]. However, capacity results are rare, a few examples appearing in [15], [16], [12].

The simple Gaussian relay network studied in this paper is illustrated in Fig. 1. The corresponding single-relay channel, i.e., the system of Fig. 1 with $M = 1$, is not a degraded relay channel according to [5]. Its capacity is unknown to date. In this paper, we derive upper and lower bounds to the capacity of the system depicted in Fig. 1. While capacity is not known for any finite M , we show in this paper that for cases of interest, our upper and lower bounds coincide in the limit as the number of relays M tends to infinity, yielding an asymptotic capacity result.

After precisely defining the considered relay network in Section II, we derive two upper bounds to the capacity in Section III. These bounds can be seen as special cases of the cut-set bound presented in [17, Theorem 14.10.1]. They can also be understood as the capacities corresponding to two idealizations of the considered system: the first upper bound is the capacity of a multiple-antenna system with one transmit and $M + 1$ receive antennas, and the second upper bound is the capacity of a multiple-antenna system with $M + 1$ transmit and one receive antenna.

In Section IV, we determine a lower bound to the capacity of the relay network of Fig. 1. More precisely, we analyze a particular communication strategy for the relay channel in which the relays use uncoded forwarding as a strategy. This should be expected to be suboptimal in general.

Then, in Section V, we show that upper and lower bounds coincide as the number of relays M tends to infinity, at least under certain conditions. We illustrate these conditions by a number of examples.

In Section VI, the results of this paper are applied to a simple sensor network scenario. Here, the relevant tradeoff is between the power used by the sensors and the fidelity (or distortion) at which the interested party can reconstruct the object of interest.

While, in general, the optimal such tradeoff is not known to date, we show that for one particular sensor network situation, the arguments developed in this paper lead to a definite result.

Section VII provides a new example of the fact that the source–channel separation paradigm does *not* extend to networks. In fact, the considered example sharpens this fundamental insight: separate source and channel coding is shown to lead to an *exponentially suboptimal* performance scaling behavior, as a function of the number M of nodes. This example follows from the comparison of the result of Section VI with the CEO problem [18], [19].

In Section VIII, the results of this paper are extended to wireless *ad hoc* networks, operated in relay mode. To this end, the simple relay network model considered in this paper is slightly extended, and it is shown that our results permit to bound capacity to within a factor of two. A slightly different interpretation of this result was discussed in [1].

II. DEFINITIONS AND NOTATIONS

The $(M + 2)$ -terminal network depicted in Fig. 1 has a source terminal, M relay terminals, and a destination terminal. A message V at the source terminal is encoded into a codeword $\{X[i]\}_{i=1}^n$ of length n whose components are complex numbers that satisfy a power constraint

$$\frac{1}{n} \sum_{i=1}^n E [|X[i]|^2] \leq P. \quad (1)$$

The M relays are the square boxes in Fig. 1. At time i , relay m observes a noisy version of the input $X[i]$ at time i

$$Y_m[i] = \alpha_m X[i] + W_m[i] \quad (2)$$

where $\{W_m[i]\}$ is a sequence (in i , for $i = 1, 2, 3, \dots$) of independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables of mean zero and variance N . Moreover, we also assume that W_m and W_l are independent for all $m \neq l$. The coefficients α_m , $m = 1, 2, \dots, M$, are arbitrary fixed, complex-valued constants, assumed to be known throughout the network. Hence, the assumption that all noise processes W_m are of the same variance is without loss of generality: different noise variances can be taken care of by appropriately adjusting the coefficients α_m . Using the sequence of observations $\{Y_m[i]\}_{i=1}^n$, the relay m produces suitable outputs $\{X_m[i]\}_{i=1}^n$ that must satisfy two constraints. First, they can only depend in a *causal* fashion upon the relay's observations, that is,

$$X_m[i] = f_{m,i}(Y_m[i-1], Y_m[i-2], \dots, Y_m[1]). \quad (3)$$

Second, they must satisfy a power constraint. We consider power constraints of the form

$$\sum_{m=1}^M \frac{1}{n} \sum_{i=1}^n E [|X_m[i]|^2] \leq c(M). \quad (4)$$

This means that our model allows power allocation between the relays. The destination observes the sum of the signals transmitted by the source and the relays, and additive white noise

$$Y[i] = \alpha_0 X[i] + \sum_{m=1}^M \delta_m X_m[i] + W[i] \quad (5)$$

where $W[i]$ is a sequence of i.i.d. circularly symmetric complex Gaussian random variables of mean zero and variance N .¹ The coefficients δ_m , $m = 1, 2, \dots, M$, are arbitrary fixed, complex-valued constants, assumed to be known throughout the network. Without loss of generality, we assume that α_0 is real valued.

Based on the observed sequence $\{Y[i]\}_{i=1}^n$, the destination terminal forms an estimate \hat{V} of the original message V . Suppose that V has B_V bits. The *capacity* C is the supremum of rates $R = B_V/n$ at which the estimate \hat{V} can be made to satisfy $\Pr(\hat{V} \neq V) < \epsilon$ for any positive ϵ , for a suitably large n .

The significance and values of the coefficients α_m and δ_m is left open at present. As stated above, we assume these coefficients to be fixed² and known throughout the network for the scope of the present paper. The coefficients α_m and δ_m may represent the path loss of the signal and hence be related to the geometry of the network as

$$\alpha_m = \frac{1}{d_{0,m}^r} \quad \text{and} \quad \delta_m = \frac{1}{d_{m,d}^r} \quad (6)$$

for $m = 1, \dots, M$, where $d_{0,m}$ is the distance from the source to relay m and $d_{m,d}$ is the distance from relay m to the destination.

For notational convenience, we define the following functions:

$$a(M) = \sum_{m=0}^M |\alpha_m|^2 \quad (7)$$

$$d(M) = \alpha_0^2 + \sum_{m=1}^M |\delta_m|^2 \quad (8)$$

$$b(M) = \sum_{m=1}^M |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2}. \quad (9)$$

Recall that the function $c(M)$ denotes the total available relay power. All of our results can be stated in terms of these auxiliary functions.

III. UPPER BOUNDS TO CAPACITY

An upper bound to the capacity of the network of Fig. 1 can be found from the cut-set theorem that appears in [17, Theorem 14.10.1]. For the purpose of this paper, we consider a weaker version of this theorem in which a single cut through the network is chosen, and the mutual information is maximized across that cut. To formalize this, consider a network with $M + 2$ nodes, and let the nodes be numbered from 0 to $M + 1$, as illustrated in Fig. 2. Also, denote the rate at which node i transmits to node j

¹The assumption that the variance of W is the same as the variances of W_m , $m = 1, 2, \dots, M$, is without loss of generality: the coefficients α_m and δ_m can always be selected accordingly.

²Another interesting case is when α_m and δ_m represent fading effects and hence, may be modeled as random variables. A version of this problem is studied in [20].

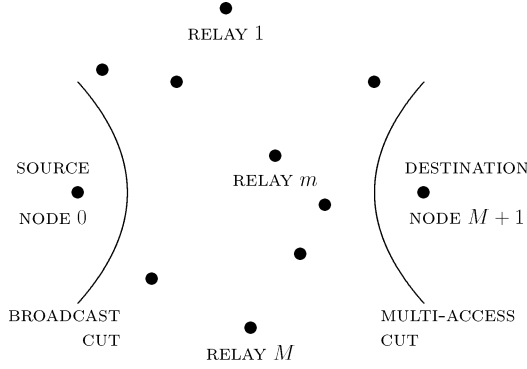


Fig. 2. Illustration of two natural cuts through the considered Gaussian relay network.

by R_{ij} . To obtain an upper bound on achievable rate allocations R_{ij} , $i, j \in \{0, 1, \dots, M+1\}$, the network is “cut,” i.e., the nodes are partitioned into two subsets, S and S^c . The following corollary to [17, Theorem 14.10.1] must be satisfied.

Corollary 1 (Weak Cut-Set Bound): The achievable rates $\{R_{ij}\}$, $i, j \in \{0, 1, \dots, M+1\}$, must satisfy, for any partitioning of the network into two sets S and S^c

$$\sum_{i \in S, j \in S^c} R_{ij} \leq \max_{p_{X_S, X_{S^c}}} I(X_S; Y_{S^c} | X_{S^c}) \quad (10)$$

where the maximization is over all $p_{X_S, X_{S^c}}$ that satisfy the power constraints.

Proof: This corollary follows straightforwardly from [17, Theorem 14.10.1], but there is also a direct way of proving it, namely by a “genie argument.” To see this, suppose that all the terminals in S can cooperate arbitrarily, and all the terminals in S^c can also cooperate arbitrarily. The resulting system cannot have smaller capacity than the original one simply because the original one is *one* way to implement the arbitrary cooperation. Hence, if we maximize the rate in the resulting system, this must lead to an upper bound to the rates in the original system. However, the resulting system is simply a point-to-point (vector) channel whose capacity is given by

$$\max_{p_{X_S, X_{S^c}}} I(X_S; Y_{S^c} | X_{S^c}) \quad (11)$$

which concludes the proof. \square

Remark 1 (Feedback): Since in the present paper, the coefficients α_m , $m = 0, 1, \dots, M$, and δ_m , $m = 1, 2, \dots, M$, are assumed to be fixed (rather than time varying or random) and known throughout the network, the point-to-point channels resulting from cutting as in Corollary 1 are memoryless channels. Therefore, the bound of Corollary 1 applies also when feedback is available. This follows directly from the proof. The capacity of the relay network is upper-bounded by the capacity of a memoryless point-to-point channel. Feedback cannot increase the capacity of such a channel (see, e.g., [17, pp. 217 and 256]).

Note that under different assumptions, feedback *may* alter capacity. For example, suppose that the coefficients α_m and δ_m are time varying and *not* known at the transmitters. Then, feedback may permit to learn the coefficients and enable higher rates

through “water-pouring” in time, see [21]. Similarly, in the case of nonergodic (quasi-static) fading, where α_m and δ_m are drawn once and for all, feedback can help, making it possible for the transmitter to prevent outage.

In order to evaluate the bound of Corollary 1, we need the capacity formula for Gaussian vector channels, i.e., channels defined by (see [22])

$$Y = GX + W \quad (12)$$

where $G \in \mathbb{C}^{L \times K}$ is a fixed and known matrix, $W \in \mathbb{C}^L$ is a vector of i.i.d. circularly symmetric complex Gaussian random variables of mean zero and unit variance, and the transmitted signal $X \in \mathbb{C}^K$ is constrained in its total power to

$$\text{trace}(E[XX^H]) \leq P. \quad (13)$$

The capacity of this channel can be characterized as follows.

Lemma 2 (MIMO Capacity [22]): The capacity of the vector channel specified by (12)–(13) is given by

$$C = \sum_{\ell=1}^{\min(L, K)} \max(0, \log_2(\lambda_\ell \nu)) \quad (14)$$

where λ_ℓ , $\ell = 1, 2, \dots, \min(L, K)$, are the eigenvalues of the matrix $G^H G$, and ν is chosen such that

$$\sum_{\ell=1}^{\min(L, K)} \max\left(0, \nu - \frac{1}{\lambda_\ell}\right) = P. \quad (15)$$

Corollary 1 and Lemma 2 permit to derive and evaluate simple upper bounds to the capacity of the relay network. From Fig. 1, it is clear that our relay network model consists of a broadcast section and a multiple-access section. Hence, there are two natural cuts to consider first, the “broadcast cut” and the “multiple-access cut.” One way to visualize this is suggested in Fig. 2: the black dots represent the terminals, i.e., the source and the destination, plus the M relay nodes.

For the broadcast cut, the mutual information expression to be maximized is

$$I(X; Y, Y_1, \dots, Y_M | X_1, \dots, X_M) \quad (16)$$

subject to the constraints

$$E[|X|^2] \leq P \quad \text{and} \quad \sum_{m=1}^M E[|X_m|^2] \leq c(M). \quad (17)$$

Under the assumption that the receiver knows the coefficients δ_m , we find

$$I(X; Y, Y_1, \dots, Y_M | X_1, \dots, X_M) = I(X; \tilde{Y}, Y_1, \dots, Y_M) \quad (18)$$

where

$$\tilde{Y} = Y - \sum_{m=1}^M \delta_m X_m. \quad (19)$$

Hence, whether or not the receiver knows the coefficients δ_m , the capacity C of the relay network can be bounded by

$$C \leq C_{\text{BC}} \stackrel{\text{def}}{=} \max_{p_X: E[|X|^2] \leq P} I(X; \tilde{Y}, Y_1, \dots, Y_M). \quad (20)$$

The right-hand side can be evaluated using Lemma 2. In particular, the channel at hand has a scalar input and a vector output,

hence, the matrix G in Lemma 2 consists of only one row, namely, $(\alpha_0, \alpha_1, \dots, \alpha_M)$, and thus, has only one singular value, given by $a(M) = \sum_{m=0}^M |\alpha_m|^2$. Hence, C_{BC} is found to be

$$C_{BC} = \log_2 \left(1 + \frac{P}{N} a(M) \right). \quad (21)$$

Let us also comment on the tightness of this bound. It was essentially found under the additional hypothesis that the relays do not send anything, but that they are the multiple antennae of the destination. However, under the hypothesis that the relays do not send anything, it is easy to find the *true* capacity from the source to the destination: $\log_2(1 + \alpha_0^2 P/N)$. Hence, the bound should not be expected to be tight.

The second cut for which we evaluate Corollary 1 is the multiple-access cut, as drawn in Fig. 2. This gives the bound

$$C \leq \max I(X, X_1, \dots, X_M; Y) \quad (22)$$

where the maximization is over all $p(x, x_1, \dots, x_M)$ that satisfy

$$E[|X|^2] \leq P \quad \text{and} \quad \sum_{m=1}^M E[|X_m|^2] \leq c(M). \quad (23)$$

To obtain a simple expression for the bound, we relax the power constraint to

$$E[|X|^2] + \sum_{m=1}^M E[|X_m|^2] \leq P + c(M). \quad (24)$$

Clearly, the maximum in the problem specified by (22) and (23) cannot be larger than in the problem specified by (22) and (24). To evaluate the latter, we can again use Lemma 2. In particular, the channel at hand has a vector input and a scalar output, hence, the matrix G in Lemma 2 consists of only one column, namely, $(\alpha_0, \delta_1, \dots, \delta_M)^T$, and thus, has only one singular value, given by $d(M) = \alpha_0^2 + \sum_{m=1}^M |\delta_m|^2$. With this, we find

$$C \leq C_{MAC} \stackrel{\text{def}}{=} \log_2 \left(1 + \frac{P + c(M)}{N} d(M) \right). \quad (25)$$

The capacity must be smaller than either one of these two bounds. Hence, we have proved the following proposition.

Proposition 3: The capacity of the Gaussian relay network of Fig. 1 is upper-bounded by

$$C \leq \log_2 \left(1 + \frac{\min \{Pa(M), (P + c(M))d(M)\}}{N} \right). \quad (26)$$

Tighter bounds can be obtained by evaluating [17, Theorem 14.10.1]. For the single-relay channel, the corresponding upper bound is [17, Theorem 14.7.1]. For the scope of this paper, however, we will only use the upper bound given by Proposition 3.

IV. LOWER BOUND TO CAPACITY

Lower bounds to capacity are found by analyzing transmission strategies. Such transmission strategies are usually set up in the following way. A sequence (in n) of codes containing $\lceil 2^{nR} \rceil$ codewords of length n is proposed, and it is shown that the error probability goes to zero as $n \rightarrow \infty$. This means that the capacity of the considered channel is at least $C \geq R$. The art consists in cleverly choosing the $\lceil 2^{nR} \rceil$ codewords.

However, whenever the separation theorem applies (e.g., for stationary ergodic point-to-point channels), there is an alternative approach that also leads to lower bounds on capacity. We

select any source (with any distortion measure). Then, we propose a transmission strategy for that source across the channel at hand (i.e., a joint source–channel code). Finally, we evaluate the performance of that communication strategy; that is, we compute the achieved distortion Δ . By the separation theorem [23, Theorem 21]

$$R(\Delta) \leq C \quad (27)$$

where $R(\cdot)$ denotes the rate-distortion function of the source with respect to the selected distortion measure (see, e.g., [17, Theorem 13.2.1]). In other words, the capacity of the channel cannot be smaller than the rate necessary to encode the source at distortion Δ . Clearly, the art consists in selecting the right source with the right distortion measure to get the best lower bound.³ The drawback of this approach is that we have to know the rate-distortion function of that source with respect to the selected distortion measure, or at least a lower bound to this function.

In this paper, we use this approach to determine a lower bound to the capacity of the relay network. We start by fixing the functions according to which the relays operate. These functions must be chosen to satisfy the power and causality constraints. Once they are fixed, the relay network is turned into a point-to-point channel. Clearly, the capacity C' of this point-to-point channel cannot be larger than the capacity C of the relay network. The goal is to determine the capacity C' of this point-to-point channel.

In particular, for the Gaussian relay network of Fig. 1, we propose the following joint source–channel coding problem. Suppose that an i.i.d. Gaussian source of variance P is transmitted without coding on the broadcast section of the relay channel. The relays simply delay the input by one time unit (to satisfy the causality condition), and scale it (up or down) to their power level. This coding technique is certainly suboptimal, but its complexity is the absolute minimum. Moreover, we will see later that this coding technique is sufficient to achieve the right scaling behavior in the number of relays M for a large class of Gaussian relay networks of the type depicted in Fig. 1.

The goal is, therefore, to determine the distortion D_1 achieved by the suggested coding scheme when the source is zero-mean i.i.d. Gaussian of variance P . The input of relay m at time n is

$$Y_m[n] = \alpha_m X[n] + W_m[n]. \quad (28)$$

The strategy of the relay is simply to scale this received value to meet its own power constraint P_m , and to transmit this result onwards. Hence, the output of relay m at time $n + 1$ is

$$X_m[n + 1] = e^{i\theta_m} \sqrt{\frac{P_m}{|\alpha_m|^2 P + N}} Y_m[n] \quad (29)$$

where θ_m is an appropriately chosen phase. There is a power allocation P_m that permits to incur the following distortion.

Proposition 4: The achievable distortion for the transmission of an i.i.d. Gaussian source (i.e., a sequence of i.i.d.

³Note that with this insight, it becomes a simple matter to prove the lower bound $R \geq \log_2(1 + P/\sigma_W^2)$ to the capacity of the standard additive white Gaussian noise channel, i.e., Fig. 1 with $M = 0$: Simply analyze the performance of uncoded transmission of an i.i.d. Gaussian source with respect to mean-squared error. See also [24].

Gaussian random variables of mean zero and variance P) across the Gaussian relay network of Fig. 1 with M relays, and where the source node has power no larger than P (see (1)) and the sum power of the relays does not exceed $c(M)$ (see (4)) is no larger than D_1 , where D_1 is given by

$$D_1 = \frac{P(N + \tau(M)(\alpha_0^2 P + N))}{\tilde{a}(M)P + N + 2\sqrt{\frac{b(M)}{c(M)}}\alpha_0 P + \tau(M)(\alpha_0^2 P + N)} \quad (30)$$

where $\tilde{a}(M) = a(M) - \alpha_0^2$, and $\tau(M) = b(M)/(\tilde{a}(M)c(M))$.

Proof: The proof is given in the Appendix.

As discussed earlier, this implies a lower bound to capacity through the separation theorem. For the case at hand, the following statement can be made.

Corollary 5: The capacity of the Gaussian relay network of Fig. 1 with M relays and where the source node has power no larger than P (see (1)) and the sum power of the relay nodes does not exceed $c(M)$ (see (4)) is at least

$$C \geq R_1 \stackrel{\text{def}}{=} \log_2 \frac{P}{D_1} \quad (31)$$

with D_1 given by (30).

Proof: By the separation theorem [23, Theorem 21], any stationary ergodic point-to-point communication system satisfies $R(\Delta) \leq C$, where Δ is the distortion achieved by that communication system. \square

Remark 2: The lower bound of Corollary 5 can be improved upon easily in at least two ways. First, fixing the relay operation as in (29), the resulting scenario is an intersymbol interference channel whose capacity could be determined via known tools. Second, the relay operation could be generalized to any (not necessarily linear) function of all the past received symbols. For the purpose of this paper, the lower bound of Corollary 5 turns out to be sufficient. Moreover, this bound has applications beyond the capacity problem, as illustrated in Sections VI and VII.

The transmission strategy leading to Corollary 5 is rather simple, yet it is a genuine “network strategy” in a sense that we now clarify. In the first step (the broadcasting from the source node to the relays), a “code” is used that permits every relay to understand the message at its particular level of fidelity. An actual code, to be decoded without error at the relays, could not possibly achieve this (unless the channels from the source to the relays all are identical). The reason for this is that a relay with a better channel would have to receive a code of higher rate than the rest. However, this higher rate would appear as noise to the relays with worse channels. This effect is also well known in the *single-source* broadcast problem, and is reflected by the fact that the separation theorem does not extend to such a topology. In fact, when a single Gaussian source is transmitted across a Gaussian broadcast channel to multiple destinations, each of which desires to reconstruct that source to within the smallest mean-squared error, then it is easy to see that uncoded transmission is optimal, while any approach based on capacity-approaching codes is not. (See also [24], [25].) In the second step

(the multiple accessing from the relays to the destination), cooperative transmission is used to boost transmit power: the signals transmitted by the relays are all correlated.

V. SCALING BEHAVIOR AND ASYMPTOTIC CAPACITY

In this section, we compare the upper bounds derived in Section III to the lower bound found in Section IV. We split the main result of this paper into two parts. In Theorem 6 below, we characterize the asymptotic difference between the lower bound R_1 and the upper bound that follows from the broadcast cut C_{BC} . Thereafter, in Theorem 8, we characterize the asymptotic difference between R_1 and the upper bound that follows from the multiple-access cut C_{MAC} .

Theorem 6 (Broadcast Cut): The capacity C of the Gaussian relay network of Fig. 1, where the source node has power no larger than P (see (1)) and the sum power of the relay nodes does not exceed $c(M)$ (see (4)), is bounded between $C_{\text{BC}} \geq C \geq R_1$. Since $a(M)$ is a nondecreasing function of M , $\lim_{M \rightarrow \infty} 1/a(M) = \theta_a < \infty$. If moreover

$$\lim_{M \rightarrow \infty} \frac{b(M)}{a(M)c(M)} = \tau < \infty \quad (32)$$

then

$$\lim_{M \rightarrow \infty} (C_{\text{BC}} - R_1) = \gamma_{\text{BC}}. \quad (33)$$

The constant γ_{BC} takes the following values.

i) If $\theta_a = 0$, then

$$\gamma_{\text{BC}} = \log_2 \left(1 + \frac{\alpha_0^2 P + N}{N} \tau \right) \quad (34)$$

which means that when $\tau = 0$, then $\gamma_{\text{BC}} = 0$.

ii) In general

$$\gamma_{\text{BC}} = \log_2 \left(\frac{1 + \frac{N}{P}\theta_a + \frac{\alpha_0^2 P + N}{N}\tau}{1 + \frac{N - 2\alpha_0^2 P}{P}\theta_a + 2\alpha_0\sqrt{\theta_a\tau} + \frac{\alpha_0^2 P + N}{P}\theta_a\tau} \frac{-\alpha_0^2 \frac{N}{P}\theta_a^2 + \frac{\alpha_0^2 P + N}{P}\theta_a\tau}{+(\alpha_0^3 - \alpha_0^2 \frac{N}{P})\theta_a^2 - 2\alpha_0^4\theta_a\sqrt{\theta_a\tau}} \right).$$

Proof: The proof is given in the Appendix.

Theorem 6 gives general conditions for the convergence of C_{BC} and R_1 . The following is a simple concrete illustration.

Example 1 (No Attenuation, Increasing Total Relay Power):

To illustrate the conditions of Theorem 6, we now study the concrete example where $|\alpha_m| = |\delta_m| = 1$ for all m , and the power constraint on the relays is $c(M) = M^u Q$ for some constant Q and some $u > 0$. Hence, $a(M) = M + 1$ and

$$b(M) = \sum_{m=1}^M |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2} = M(P + N). \quad (35)$$

This implies (using $a(M) \geq M$)

$$\begin{aligned} \lim_{M \rightarrow \infty} \frac{b(M)}{a(M)c(M)} &\leq \lim_{M \rightarrow \infty} \frac{M(P + N)}{M^{u+1}Q} \\ &= \lim_{M \rightarrow \infty} \frac{P + N}{M^u Q} = 0 \end{aligned} \quad (36)$$

since $u > 0$ by assumption. Hence, in Theorem 6, we have $\theta = \tau = 0$, which yields

$$\lim_{M \rightarrow \infty} (C_{\text{BC}} - R_1) = 0 \quad (37)$$

which means that the capacity of this network behaves asymptotically like

$$C_{\text{BC}} = \log_2 \left(1 + \frac{(M+1)P}{N} \right). \quad (38)$$

In this example, Theorem 6 is asymptotically tight, i.e., it leads to a capacity result.

Example 1 is the simplest possible (nontrivial) case of a network according to Fig. 1. In spite of this fact, the example generalizes in a straightforward manner to cover a large class of interesting cases: whenever the total available power increases and the attenuation coefficients are lower and upper bounded (strictly larger than zero, strictly smaller than infinity), then Theorem 6 yields a capacity result. We formulate this in the shape of the following corollary.

Corollary 7 (Bounded Attenuation Coefficients): In the setup of Theorem 6, suppose that the attenuation coefficients are strictly bounded, $0 < |\alpha_m| < \infty$ and $0 < |\delta_m| < \infty$, for all m , and that the power constraint is $c(M) = M^u Q$, with $u > 0$ and Q some constant. Then

$$\lim_{M \rightarrow \infty} (C_{\text{BC}} - R_1) = 0 \quad (39)$$

i.e., in this case, the capacity behaves asymptotically like

$$C_{\text{BC}} = \log_2 \left(1 + \frac{a(M)P}{N} \right). \quad (40)$$

Proof: Since the attenuation coefficients are strictly bounded, we can upper-bound $b(M) \leq Mb_{\text{max}}^2$ and lower-bound $a(M) \geq Ma_{\text{min}}^2$, and hence,

$$\frac{b(M)}{a(M)c(M)} \leq \frac{Mb_{\text{max}}^2}{Ma_{\text{min}}^2 M^u Q} = \frac{b_{\text{max}}^2}{a_{\text{min}}^2 Q} \frac{1}{M^u}. \quad (41)$$

Since $u > 0$ by assumption, this implies that in Theorem 6, $\tau = 0$. By the same token

$$\frac{1}{a(M)} \leq \frac{1}{Ma_{\text{min}}^2}, \quad (42)$$

and hence, $\theta_a = 0$. But then, Theorem 6 asserts the claim. \square

The main point of Corollary 7 is to illustrate that there is a meaningful and interesting parameter range for which our arguments establish capacity (in the limit as $M \rightarrow \infty$). Moreover, it also shows that in a scaling sense, as M becomes large, an optimal relay ‘‘recoding’’ strategy is simply uncoded forwarding. To make a stronger point, it is unclear if there is a different strategy than the one considered here (and its ramifications along Remark 2) that *also* achieves the optimum capacity scaling behavior. In particular, consider the simple ‘‘point-to-point coding’’ scheme used, e.g., in [26]: all transmissions are point-to-point and receivers consider any interfering signal purely as noise. For the *ad hoc* wireless network scenario considered there, it has been shown in [27] that such a simple strategy does achieve the optimum capacity scaling behavior. For the feedback scenario considered here,

it is immediately clear that this coding strategy will perform suboptimally; in fact, its performance will *not* improve with M : the bottleneck is the source node. Under the point-to-point coding restriction, the source node can only transmit to one relay in any given time slot, and hence, the resulting rate is bounded by $\log(1 + P/N)$, where P is the power of the source node. Hence, for the Gaussian relay network as considered here, going beyond point-to-point coding significantly changes the asymptotic behavior. This conclusion is certainly of interest in the interpretation of the result of [26]: it suggests the possibility that the asymptotic behavior of capacity *does change* when the point-to-point coding restriction is removed.

We now proceed to the examination of the asymptotic difference between the upper bound stemming from the multiple-access cut ((25)) and the achievable rate R_1 . The main result is the following theorem, which is the analog of Theorem 6.

Theorem 8 (Multiple-Access Cut): The capacity C of the Gaussian relay network of Fig. 1 where the source node has power no larger than P (see (1)) and the sum power of the relay nodes does not exceed $c(M)$ (see (4)), is bounded between $C_{\text{MAC}} \geq C \geq R_1$. Since $a(M)$, $c(M)$, and $d(M)$ are nondecreasing functions of M

$$\lim_{M \rightarrow \infty} 1/a(M) = \theta_a < \infty$$

$$\lim_{M \rightarrow \infty} 1/c(M) = \theta_c < \infty$$

and

$$\lim_{M \rightarrow \infty} 1/d(M) = \theta_d < \infty.$$

If moreover

$$\lim_{M \rightarrow \infty} \frac{c(M)d(M)}{a(M)} = \tau_1 < \infty$$

and

$$\lim_{M \rightarrow \infty} \frac{b(M)d(M)}{a^2(M)} = \tau_2 < \infty \quad (43)$$

then

$$\lim_{M \rightarrow \infty} (C_{\text{MAC}} - R_1) = \gamma_{\text{MAC}}. \quad (44)$$

The constant γ_{MAC} takes the following values.

i) If $\theta_a = \theta_d = 0$, then

$$\gamma_{\text{MAC}} = \log_2 \left(\tau_1 \theta_c + \frac{\tau_1}{P} + \frac{(\alpha_0^2 P + N)(1 + P\theta_c)}{PN} \tau_2 \right). \quad (45)$$

Note that the bound is tight if the argument of the logarithm is 1.

ii) In general

$$\gamma_{\text{MAC}} = \log_2 \left(\frac{N\tau_1 - N\alpha_0^2\tau_1\theta_a + PN\tau_1\theta_c + N^2\theta_a}{PN - 2PN\alpha_0^2\theta_a + 2PN\alpha_0^4\theta_a^2 + N^2\theta_a} - \frac{PN\alpha_0^2\tau_1\theta_c\theta_a - N^2\alpha_0^2\theta_a^2}{+2\alpha_0 PN\sqrt{\tau_2\theta_c\theta_d} - 2\alpha_0^3 N\sqrt{\tau_2\theta_c\theta_d}\theta_a} + \frac{(\alpha_0^2 P + N)(\tau_2 + P\tau_2\theta_c + N\tau_2\theta_c\theta_d)}{-\alpha_0^2 N^2\theta_a^2 + (\alpha_0^2 PN + N^2)\tau_2\theta_c\theta_d} \right).$$

Proof: The proof is given in the Appendix.

Example 1, continued 1: Consider again the setup of Example 1. Theorem 8 is of no value here: bounding $a(M) \leq 2M$, we find

$$\frac{c(M)d(M)}{a(M)} \geq \frac{M^{u+1}Q}{2M} = M^u \frac{Q}{2} \quad (46)$$

which does not converge as M tends to infinity. In other words, the multiple-access cut leads to a very loose upper bound to capacity in this example.

Example 2 (No Attenuation, Constant Total Relay Power):

Like in Example 1, suppose that $|\alpha_m| = |\delta_m| = 1$, for all m . However, the power constraint for the present example is $c(M) = Q$, where Q is some constant. As in Example 1, $a(M) = M + 1$ and

$$b(M) = \sum_{m=1}^M |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2} = M(P + N). \quad (47)$$

First, consider Theorem 6. To evaluate the conditions, note that

$$\lim_{M \rightarrow \infty} \frac{b(M)}{a(M)c(M)} = \lim_{M \rightarrow \infty} \frac{M(P + N)}{(M + 1)Q} = \frac{P + N}{Q} \quad (48)$$

Hence, in Theorem 6, we have $\theta = 0$, but $\tau = (P + N)/Q$, which yields

$$\lim_{M \rightarrow \infty} (C_{\text{BC}} - R_1) = \log_2 \left(1 + \frac{(P + N)^2}{NQ} \right). \quad (49)$$

Hence, R_1 is asymptotically only a constant additive term away from C_{BC} ; and thus, it also grows like $\log M$. Since the capacity lies between $C_{\text{BC}} \geq C \geq R_1$, we conclude that C grows like $\log M$ as well. Let us also briefly discuss the difference $C_{\text{BC}} - R_1$. It is seen that this difference decreases with Q , but increases with P . This is due to the fact that for our decoding scheme, the original signal of power P is an *interferer* at the destination.

In order to apply Theorem 8 to this example, we first have to determine τ_1 and τ_2 in (43), as follows:

$$\tau_1 = \lim_{M \rightarrow \infty} \frac{c(M)d(M)}{a(M)} = \lim_{M \rightarrow \infty} \frac{QM}{M + 1} = Q \quad (50)$$

and

$$\tau_2 = \lim_{M \rightarrow \infty} \frac{b(M)d(M)}{a^2(M)} = \lim_{M \rightarrow \infty} \frac{(P + N)M^2}{(M + 1)^2} = P + N.$$

Since they are both finite, Theorem 8 does apply to this example. We evaluate moreover $\theta_a = 0$ and $\theta_d = 0$, and hence, the value of the bound supplied by Theorem 8 is determined by (45). With $\theta_c = 1/Q$, we find

$$\begin{aligned} \lim_{M \rightarrow \infty} (C_{\text{MAC}} - R_1) \\ = \log_2 \left(1 + \frac{Q}{P} + \frac{(\alpha_0^2 P + N)(P + N)(1 + P/Q)}{PN} \right). \end{aligned}$$

This bound is always weaker than (49). To verify this, note that in the present example, the broadcast bound C_{BC} is asymptotically always smaller than the multiple-access bound C_{MAC} , which follows immediately from a comparison of (20) and (25)

$$\lim_{M \rightarrow \infty} 2^{C_{\text{MAC}} - C_{\text{BC}}} = 1 + \frac{Q}{P}. \quad (51)$$

Example 3: In this example, we study the scenario where $|\alpha_m|^2 = m$ and $|\delta_m| = 1$. We can bound $a_0 M^2 \leq a(M) \leq a_1 M^2$, $b_0 M^3 \leq b(M) \leq b_1 M^3$, and $d(M) = M$. Moreover, suppose a constant total power $c(M) = Q$. Then

$$\frac{b(M)}{a(M)c(M)} \geq \frac{b_0 M^3}{a_1 M^2 Q} \quad (52)$$

which diverges, and hence, Theorem 6 does not apply.

As for Theorem 8, we find

$$\frac{c(M)d(M)}{a(M)} \leq \frac{QM}{a_0 M^2} \quad (53)$$

which tends to zero (and hence, $\tau_1 = 0$), and

$$\frac{b(M)d(M)}{a^2(M)} \leq \frac{b_1 M^3 M}{a_0^2 M^4} = \frac{b_1}{a_0} \quad (54)$$

which converges. This means that in this example, Theorem 8 yields a tighter bound than Theorem 6. In fact, evaluating (45), we find

$$\lim_{M \rightarrow \infty} (C_{\text{MAC}} - R_1) \leq \log_2 \left(\frac{(\alpha_0^2 P + N)(1 + P/Q)}{PN} \tau_2 \right). \quad (55)$$

It is also clear from this expression that a capacity result is obtained if it is possible to slightly alter α_m and δ_m in such a way as to make

$$\tau_2 = \frac{PN}{(\alpha_0^2 P + N)(1 + P/Q)}$$

(while keeping all the limits fixed as in this example).

Another example is given in Section VIII-C for a simple scenario inspired by wireless networks. There, the values of α_m and δ_m are randomly chosen according to an appropriate distribution, and Fig. 6 compares the presented inner bound R_1 to the outer bounds C_{BC} and C_{MAC} .

VI. APPLICATION: GAUSSIAN SENSOR NETWORK

The topology of our network model, Fig. 1, also resembles a particular sensor network situation: in that case, the sequence $\{X[i]\}$ is the physical phenomenon to be measured, $\{W_m[i]\}$ are due to the fact that the phenomenon cannot be measured directly as well as due to measurement noise, and the relays are the sensors themselves. For the sensor network situation, we take $\alpha_0 = 0$. This is illustrated in Fig. 3. The goal is for a central unit to learn about the physical phenomenon $\{X[i]\}$. More precisely, its goal is to provide a reconstruction $\{\hat{X}[i]\}$ such that

$$D_M \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E \left[|X[i] - \hat{X}[i]|^2 \right] \quad (56)$$

is minimized. The sensors communicate to the central unit over a common wireless channel, and their transmitted powers have to satisfy a sum power constraint. This is clearly a source-channel communication problem; the relevant tradeoff is between the total power of the sensors and the fidelity at which the central unit can reconstruct the sequence $\{X[i]\}$.

The optimal such tradeoff is not known to date. Note that the separation paradigm does not extend to this case: compressing

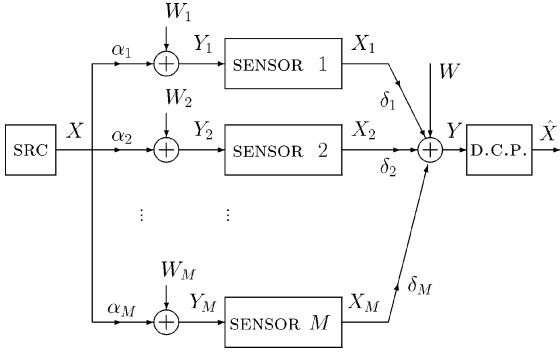


Fig. 3. The considered sensor network: The data collection point (DCP) needs to reconstruct X , i.e., the underlying source (SRC).

the sources (using the concepts of Slepian and Wolf [28], and their extension to the case of lossy compression [29]) and transmitting the source codewords using a capacity-achieving code on the multiple-access channel is a suboptimal strategy. A simple illustration of this can be found in [17, p. 448]. We address this problem in Section VII.

The results of this paper permit to determine the optimal tradeoff for the following scenario. Suppose that $\{X[i]\}$ is a sequence of i.i.d. circular complex Gaussian random variables of variance P . Sensor m measures

$$Y_m[i] = \alpha_m X[i] + W_m[i]$$

where $W_m[i]$ is i.i.d. circular complex Gaussian noise of variance N . Sensor m is allowed to get to know the entire sequence $\{Y_m[i]\}$ before transmitting a sequence

$$\{X_m[i]\} = f_m(\{Y_m[i]\})$$

at average power

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[|X_m[i]|^2] \leq P_m.$$

There are M sensors, and their total power is constrained to be

$$\sum_{m=1}^M P_m \leq MQ. \quad (57)$$

The final destination receives

$$Y[i] = \sum_{m=1}^M \delta_m X_m[i] + W[i] \quad (58)$$

where $W[i]$ is i.i.d. circular complex Gaussian noise of variance N . The final destination is allowed to get to know the entire sequence $\{Y[i]\}$ before producing the sequence of estimates $\{\hat{X}[i]\} = g(\{Y[i]\})$.

Corollary 9: For the sensor network model described in this section, if the conditions of Theorem 6 are satisfied with $\theta_a = 0$ and $\tau = 0$, then the optimum distortion $D_{M,\text{opt}}$ satisfies

$$\lim_{M \rightarrow \infty} \min_{f_1, \dots, f_M, g} \frac{D_{M,\text{opt}}}{\frac{PN}{a(M)P+N}} = 1 \quad (59)$$

and the minimum is achieved when the sensors use a simple scaling, $X_m[i] = \gamma_m Y_m[i]$, and the final destination uses $\hat{X}[i] = \gamma Y[i]$.

Remark 3: The encoding functions f_m are much more general here than in the capacity consideration for the relay channel: there is no causality constraint like (3) in the considered sensor network model. This means that there are more degrees of freedom in choosing the encoding functions f_m in the present scenario. Nevertheless, our result says that the optimum (asymptotically as $M \rightarrow \infty$) can be achieved without exploiting these additional degrees of freedom, by simple *causal* encoding functions respecting the constraint (3). More explicitly, in this example, the global optimum can be achieved by *real-time processing*.

Remark 4: Extensions of this result, involving multiple underlying sources, multiple data collection points and fading effects, and addressing the tradeoff between the total sensor power and the achievable distortion more explicitly, have been presented in [30]–[32], [20]. In particular, it is shown that (57) is not a necessary condition for the result to hold.

Proof: Suppose that the physical phenomenon $X[i]$ itself uses optimal coding. This is clearly an idealization, and hence leads to a lower bound to the distortion. For a given source, the achievable end-to-end distortion certainly cannot be smaller than the rate-distortion function of the source, evaluated at the capacity upper bound C_{BC} . This is immediate since the multiple-antenna idealization of the multiple-relay channel is a simple stationary ergodic point-to-point channel, hence the separation theorem applies. The distortion for this idealized system can be calculated as

$$D \geq D_{\text{BC}} \stackrel{\text{def}}{=} D_{\mathcal{N}}(C_{\text{BC}}) = \frac{PN}{a(M)P+N} \quad (60)$$

where $D_{\mathcal{N}}(\cdot)$ denotes the distortion-rate function of the i.i.d. circularly complex Gaussian source. An achievable distortion D_1 has been found in Proposition 4. Now consider the quotient D_1/D_{BC} , which, using the distortion-rate function $D_{\mathcal{N}}(R) = \sigma_S^2 2^{-R}$, can be expressed as

$$\frac{D_1}{D_{\text{BC}}} = 2^{C_{\text{BC}} - R_1}. \quad (61)$$

The convergence of the latter is established in the proof of Theorem 6. Hence, as M tends to infinity, the smallest achievable distortion behaves like the D_{BC} . \square

VII. APPLICATION: THE CEO PROBLEM

The problem studied in Section VI resembles the CEO problem, proposed and solved in [18]. This problem is illustrated in Fig. 4: M agents (or sensors) observe each a different noisy copy of one and the same underlying source sequence $\{X[i]\}_{i=1}^n$, just like in our model. However, the agents' task is different: in the CEO problem, agent m is allotted R_m bits per source sample, and must use them so as to enable the CEO (or data collection point) to provide the best possible estimate $\{\hat{X}[i]\}_{i=1}^n$ of the underlying source sequence, in the sense of (56). The problem is to determine, for a given rate budget $R = \sum_{m=1}^M R_m$, the smallest achievable distortion. The compression must be performed in a distributed fashion: agent m must encode its observation sequence $\{Y_m[i]\}_{i=1}^n$ without

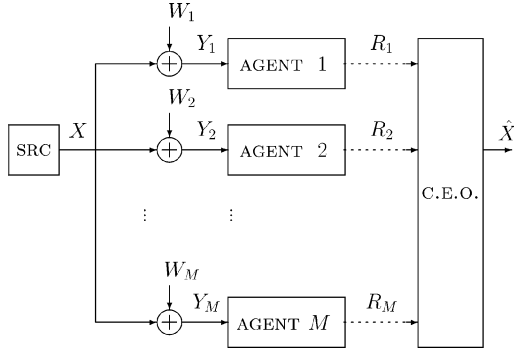


Fig. 4. Illustration of the CEO source coding problem, quoted from [18], [19], [33]: M agents have to keep their C.E.O. up to date about the underlying source X .

knowing what the other agents observed. A central decoder receives all M codewords, and produces an estimate sequence $\{\hat{X}[i]\}_{i=1}^n$, incurring a distortion of

$$D_n(R) = \frac{1}{n} \sum_{i=1}^n E \left[\left| X[i] - \hat{X}[i] \right|^2 \right]. \quad (62)$$

In [19], [33], it is established that

$$D_{\text{CEO}}(R) = \lim_{n \rightarrow \infty} D_n(R)$$

is characterized, in the limit as $M \rightarrow \infty$, by

$$D_{\text{CEO}}(R) = \frac{PN}{RP + N} \quad (63)$$

which is [33, eq. (6)], adapted to the case of a circularly complex Gaussian source.

The goal of this section is to reconsider the CEO problem as in Fig. 4, but with the abstract communication channels (“bit pipes” of rates R_1, R_2, \dots, R_M) replaced by a “physical” channel model, hence extending the source coding problem into a joint source–channel coding problem. Let us consider the standard additive white Gaussian noise multiple-access channel, as defined in [17, p. 378], under a sum power constraint on the participating nodes, as in (4), and let the total power function be $c(M) = MQ$. This is, of course, the channel model used in the system of Fig. 3, with $\delta_m = 1$, for $m = 1, 2, \dots, M$.

Clearly, one tempting strategy to tackle the resulting joint source–channel communication problem consists in letting the agents compress their observation using the optimal code for the CEO problem, and then using the multiple-access channel at the highest possible rates (i.e., capacity) to communicate the source codewords down to the CEO. To evaluate this performance, we can simply use (63), and it remains to evaluate how large R can be made on the channel model at hand. The capacity region of the additive white Gaussian noise multiple-access channel is well known (see, e.g., [17, p. 378] and the references therein), but the additional twist here is that the source codewords may be *dependent* since they are drawn from a common source sequence $\{X[i]\}_{i=1}^n$. The capacity of multiple-access channels when the messages are allowed to be “*somewhat*” dependent is not generally known [34], but if they can be *arbitrarily* dependent, then the multiple-access channel becomes a point-to-point channel with an input vector of length M , and the capacity of

such a “multiple-antenna” channel is well known (see [22] and the references therein)

$$R \leq \log_2 \left(1 + \frac{M^2 Q}{N} \right), \quad (64)$$

see also (25). Clearly, the total rate available to the agents in our example to communicate their source codewords cannot exceed (64), and hence, for any such strategy, the resulting distortion cannot be smaller than

$$D_{\text{sep}} \geq \frac{PN}{\log_2 \left(1 + \frac{M^2 Q}{N} \right) P + N}. \quad (65)$$

The question is whether this is the optimum performance, or whether a smaller distortion can be achieved under the same constraints.

To answer this question, we can use Corollary 9, determining optimum performance. Evaluating the corollary for the case at hand, the coefficients are $\alpha_0 = 0$ and $\alpha_m = \delta_m = 1$, for $m = 1, \dots, M$. Hence, $a(M) = M$ and $b(M) = M(P + N)$. In the current consideration, $c(M) = MQ$, and hence, $\theta_a = 0$ and $\tau = 0$. Therefore, the asymptotic behavior (as $M \rightarrow \infty$) of the distortion achieved by the uncoded strategy is found from Corollary 9 as

$$D_{\text{opt}} = \frac{PN}{MP + N}. \quad (66)$$

Recall that this scheme satisfies an additional property of causality and real-time processing, as described in Corollary 9.

The comparison of (65) with (66) confirms the well-known fact that separate source and channel coding is generally sub-optimal in networks, see, e.g., [17, p. 448] and [25]. But the comparison of (65) with (66) also provides a new insight: separate source and channel codes are not only suboptimal, they may scale *exponentially worse* (as a function of the number of nodes M) than the optimum scheme. More precisely, for a fixed target distortion D , denote the number of nodes required by the optimum scheme by M_{opt} , and the number of nodes required by the separation-based scheme by M_{sep} . Then, comparing (65) with (66), we find (using the concavity of the square root)

$$M_{\text{sep}} \geq \sqrt{\frac{N}{Q}} \left(e^{M_{\text{opt}}/2} - 1 \right) \quad (67)$$

i.e., M_{sep} is exponentially larger than M_{opt} .

VIII. EXTENSION: WIRELESS NETWORKS

It is clear that the network of Fig. 1 does not well model a wireless situation. Rather, we now consider the following extended network model. Relay k does not only receive the transmission from the source node, but also the transmissions from all other relay nodes. We replace (2) by

$$Y_m[i] = \alpha_m X[i] + \sum_{j=1, j \neq m}^M \alpha_{jm} X_m[j] + W_m[i]. \quad (68)$$

The network model considered by Gupta and Kumar in [26] is of that kind. Their network is discussed in more detail in Section VIII-B. The upper bound formulated in Proposition 3 applies unchanged: the cut-set idealization still yields upper bounds to capacity. The lower bound, however, is changed

since there are many more interfering terms at the input of the relays, as a comparison of (2) with (68) reveals. One way to obtain a simple lower bound is to apply the same strategy, but to split it into two time slots: in the first time slot, the source sends to the relays, in the second time slot, the relays send to the destination. That is, a factor of two in rate is lost due to this time-multiplexing. Apart from that, the same rate is achievable as for the network specified by (2).⁴ That is, our arguments permit to determine capacity up to a factor of two. One such result is summarized in the following corollary.

Corollary 10: For the wireless relay network model defined by (68) and (5), where the source node has power no larger than P (see (1)) and the sum power of the relay nodes does not exceed $c(M)$ (see (4)), suppose that the conditions of Theorem 6 are satisfied with $\theta_a = 0$ and $\tau = 0$, i.e.,

$$\lim_{M \rightarrow \infty} \frac{1}{a(M)} = 0 \quad \text{and} \quad \lim_{M \rightarrow \infty} \frac{b(M)}{a(M)c(M)} = 0.$$

Then, the achievable rate R'_1 and the upper bound C_{BC} satisfy

$$\lim_{M \rightarrow \infty} C_{BC} - 2R'_1 = 0 \quad (69)$$

i.e., the true capacity C of the considered relay network is bounded by $C_{BC}/2 \leq C \leq C_{BC}$ as $M \rightarrow \infty$.

Proof: The scheme used to prove Proposition 4 is operated in every other time slot: in odd time slots, the source outputs a new source value; in even time slots, the relays forward an appropriately scaled version of the signal they received in the preceding odd time slot. Hence, the links that are added in (68) over (2) are of no relevance to this scheme. In other words, the distortion D_1 of Proposition 4 is still achieved (but only in every other time slot). By Corollary 5, this implies that the rate $R'_1 = R_1/2$ is achievable. Moreover, the outer bound C_{BC} remains unaffected when (2) is replaced by (68) since none of the additional links in (68) cross the broadcast cut illustrated in Fig. 2. Hence, Theorem 6 can be applied to the difference between $2R'_1$ and C_{BC} , establishing the corollary. \square

Clearly, Corollary 10 can be extended by taking into account the remaining cases covered by Theorem 6 as well as the ones covered by Theorem 8. We do not discuss this explicitly here.

Instead, the result of Corollary 10 is illustrated for two special classes of networks. For networks similar to the ones studied in [26], where the nodes are arbitrarily placed *except* for a certain dead zone, and for networks that are forced to employ frequency duplex: the source transmits on a different frequency band than the relays.⁵

A. Arbitrary Node Placement With a Dead Zone

Consider the following network, illustrated in Fig. 5: $M + 2$ nodes are placed arbitrarily in a disk of unit area, and the coefficients α_{jm} characterize the path losses, i.e.,

$$\alpha_{jm} = \frac{1}{d_{j,m}^r} \quad (70)$$

⁴Again, in line with Remark 2, note that the considered scheme can be finetuned in many ways. For the scaling-law perspective taken in this paper, this seems to lead to only marginal improvements and is therefore omitted.

⁵A simplified version of these capacity results was presented at the 2001 IMA “Hot Topics” Workshop on Wireless Networks, August 8–10, 2001, Minneapolis, MN, and appeared in [1].

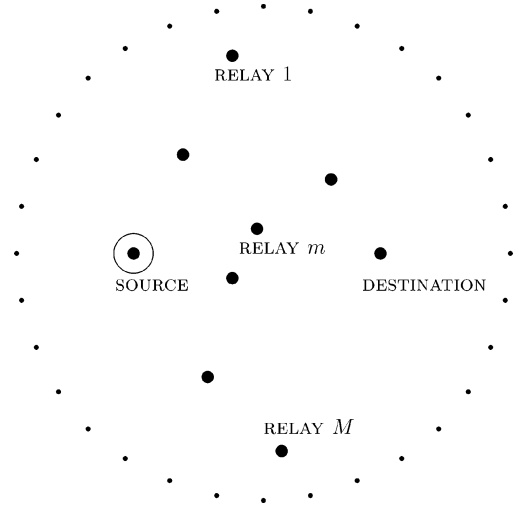


Fig. 5. A wireless relay network model, based on [26].

where $d_{j,m}$ is the Euclidean distance between nodes j and m , and r is the path loss exponent.⁶ These networks are sometimes called *dense* networks since, as the number of nodes M tends to infinity, the distance between adjacent nodes vanishes.

In [26], the nodes are partitioned into pairs (uniformly at random), each pair consisting of a source and a destination, and the goal of the analysis is to characterize the maximum throughput per pair. The analysis presented in [26] is limited to the case where all transmissions are carried out in a point-to-point fashion, considering simultaneous transmissions purely as noise.⁷ In contrast to this, our argument considers the traffic scenario where two special nodes are selected, namely, a source and a destination, and all other nodes serve purely as relays (see Fig. 5), and *all possible* coding schemes are allowed, not only those satisfying the point-to-point coding hypothesis mentioned above.

To apply Corollary 10 to the network of Fig. 5, we have to compute the values of θ_a and τ . To make a simple, but precise statement, suppose that there is a “dead zone” of radius $\epsilon(M)$ around the source node: inside this zone, no other node can be placed.⁸ This is suggested in Fig. 5 by the solid circle around the source node. The dead-zone radius $\epsilon(M)$ may depend on the number of nodes M in the network. In particular, in interesting cases, it should *decrease* as the number of nodes M is increased. Notice that the remaining area of the network is unconstrained—the relay nodes can be placed as close to each other as desired. For this setup, we can prove the following capacity bound.

Corollary 11: For the wireless relay network model defined by (68) and (5), suppose that the nodes are placed arbitrarily inside a disk of unit area, but outside of a dead zone of radius $\epsilon(M)$ around the source node. Suppose that the coefficients α_m

⁶In [26], this exponent is denoted by α .

⁷More general results beyond this restriction on the coding scheme were recently presented in [27]. However, those results do not seem to apply to dense networks; rather, they impose a lower bound on the distance between any two nodes.

⁸A more restrictive dead zone assumption, namely, a *fixed* zone around *every* node in the network (and hence not leading to dense networks), has recently been studied in [27].

and δ_m depend on the distance between nodes according to (70), that the source node has power no larger than P (see (1)), and the sum power of the relay nodes does not exceed $c(M)$ (see (4)). If

$$\lim_{M \rightarrow \infty} \frac{1}{\epsilon^{4r}(M)c(M)} = 0 \quad (71)$$

then, the following asymptotic capacity result holds:

$$\lim_{M \rightarrow \infty} C_{\text{BC}} - 2R_1' = 0 \quad (72)$$

implying that the capacity C of the wireless relay network is bounded by $C_{\text{BC}}/2 \leq C \leq C_{\text{BC}}$ as $M \rightarrow \infty$, where

$$C_{\text{BC}} = \log_2 \left(1 + \frac{a(M)P}{N} \right). \quad (73)$$

Remark 5: At fixed sum power $c(M)$, (71) characterizes the fastest rate at which the radius of the dead zone around the source node may decrease such that the arguments developed in this paper suffice to establish a point-wise convergence in the sense that (72) holds for *arbitrary* node placements. If the dead zone radius is to decrease faster than (71), then the arguments developed in this paper may still be used to establish weaker convergence results. For example, for many *probability distributions* for the node locations, (72) can be used to determine the behavior of the *expected value* of capacity even in cases where (71) is violated.

Proof: With reference to Corollary 10, we merely have to establish that the coefficients α_m and δ_m satisfy the conditions of Theorem 6. Since the largest distance in the network is $2/\sqrt{\pi}$, we have that $|\alpha_m|^2 \geq (\pi/4)^r$. This implies that $a(M) \geq (M+1)(\pi/4)^r$, and hence, $\theta_a = 0$. Moreover, since by the same token, $|\delta_m|^2 \geq (\pi/4)^r$, we find

$$b(M) = \sum_{m=1}^M |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2} \quad (74)$$

$$\leq (\pi/4)^r \sum_{m=1}^M |\alpha_m|^2 (|\alpha_m|^2 P + N). \quad (75)$$

By the dead zone assumption, $|\alpha_m|^2 \leq 1/\epsilon^{2r}(M)$, and hence,

$$b(M) \leq (\pi/4)^r M \frac{1}{\epsilon^{2r}(M)} \left(\frac{1}{\epsilon^{2r}(M)} P + N \right). \quad (76)$$

This permits to bound as follows:

$$\frac{b(M)}{a(M)c(M)} \leq (\pi/4)^{2r} \frac{M}{M+1} \frac{P + N\epsilon^{2r}(M)}{\epsilon^{4r}(M)c(M)} \quad (77)$$

showing that (71) implies that $\tau = 0$, as required in Corollary 10. \square

Corollary 11 is true for all node placements that respect the dead zone assumption. Suppose, for example, that the total power available to the nodes is $c(M) = MQ$ for some constant Q . Then, for our arguments to establish capacity, the dead-zone radius cannot decrease faster than

$$\epsilon(M) \geq \frac{1}{M^{1/(4r)}}. \quad (78)$$

Let us assume this, and that the node placement is random⁹ according to some law that respects the dead zone. For any fixed

⁹This is understood in a quasi-static sense. At the beginning of time, the node placements are drawn from a distribution, but they then stay fixed for the rest of the experiment.

node placement, Corollary 11 provides upper and lower bounds on capacity. Hence, in expectation (taken over all node placements), the expected capacity $E[C]$ of the wireless relay network is bounded as $E[C_{\text{BC}}/2] \leq E[C] \leq E[C_{\text{BC}}]$, where

$$E[C_{\text{BC}}] = \int_{\alpha_1} d\alpha_1 \cdots \int_{\alpha_M} d\alpha_M \left(\prod_{m=1}^M p(\alpha_m) \right) \log_2 \left(1 + \frac{P}{N} \sum_{m=1}^M \alpha_m^2 \right).$$

Without evaluating this integral explicitly, we can use the fact that $(\pi/4)^r \leq |\alpha_m|^2 \leq 1/\epsilon^{2r}(M)$ to bound the expected capacity as

$$\log_2 \left(1 + \frac{P}{N} \left(\frac{\pi}{4} \right)^r M \right) \leq E[C_{\text{BC}}] \leq \log_2 \left(1 + \frac{P}{N} \frac{1}{\epsilon^{2r}(M)} M \right) \quad (79)$$

revealing the “scaling law” of the (expected) capacity of the considered wireless network under the “relay” traffic pattern, in Knuth’s notation (cf. the “scaling laws” in [27])

$$E[C] = \Theta(\log_2 M) \quad (80)$$

or, more explicitly, $\lim_{M \rightarrow \infty} E[C]/(\log_2 M) = c$, for some constant $0 < c < \infty$.

B. Frequency Duplex

As a second application to wireless networks, suppose now that the source transmits on a frequency band different from the relay transmitters’ frequency bands. This implies that the relay receiver does not incur the interference terms present in (68). Rather, the latter is replaced by (2). Let us further assume that the destination can only receive on the relay transmitters’ frequency band.¹⁰ This frequency duplex situation is represented by Fig. 1, with $\alpha_0 = 0$. Hence, the highest rate that can be achieved by any such frequency-duplex scheme, denoted as C_{duplex} , can be characterized by Theorems 6 and 8, whenever they apply.

Let us reconsider the geometry described in Section VIII-B. Corollary 11 still applies, the only change being that the achievable rate is now given by $R_1 = 2R_1'$, since the source node transmits in every time slot. In particular, Corollary 11 implies that whenever the dead-zone radius satisfies (71), $R_1 \leq C_{\text{duplex}} \leq C_{\text{BC}}$, where

$$\lim_{M \rightarrow \infty} C_{\text{BC}} - R_1 = 0. \quad (81)$$

To gain insight, let us now investigate how fast these two bounds converge.

Example 4: Consider the geometry of Section VIII-B, i.e., consider a network inside a disk of unit area centered at the origin of a Cartesian coordinate system, where the coefficients α_m and δ_m depend on the Euclidean distances between nodes according to (70). Suppose the source node is located at coordinates $(-0.25, 0)$ and the destination node at $(0.25, 0)$. Let the source node have power $P = 10$. The locations of the M relays are chosen uniformly (outside the dead zone of fixed radius

¹⁰If the destination can also listen on the source transmitter’s frequency bands, this can at most change the capacity by an additive constant, and is thus not scaling-law relevant. To keep considerations simple, we therefore omit this possibility.

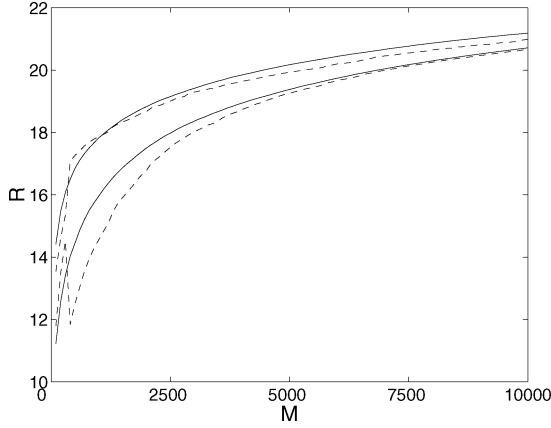


Fig. 6. Numerical evaluation of upper (see (21)) and lower (see (31)) bounds versus the number of nodes in the network generated according to Example 4. For a single experiment (dashed lines), and averaged over 100 experiments (solid lines).

$\epsilon = 0.01$), and their total power is $c(M) = 10M$. Let the noise variances be $N = 1$. Fig. 6 shows the corresponding bounds R_1 and C_{BC} against the number of nodes M for one realization of the node locations, and averaged over 100 trials. The figure clearly illustrates the typical $\log_2 M$ behavior of capacity, and it shows the convergence of the upper and lower bounds established in this paper.

IX. CONCLUSION

For the Gaussian relay network of Fig. 1, capacity is not known to date, not even for $M = 1$. In this paper, we determined the asymptotic capacity in the limit as the number of relays M tends to infinity. For many interesting relay networks, our paper gives an *exact* asymptotic capacity result, most notably for all cases where the fading coefficients are strictly larger than zero and strictly smaller than infinity. For the lower bound that leads to our capacity result, the relays simply use uncoded forwarding. This can be seen as an extension of [24].

Beyond the exact capacity results, we show for a larger class of Gaussian relay networks that the typical *scaling behavior* of capacity is $\log M$, where M is the number of relay nodes. This is demonstrated even for network models beyond Fig. 1, including certain wireless scenarios. In contrast to this, the point-to-point coding hypothesis of [26] only leads to a constant rate, independent of the number of relays. This shows that at least in certain situations, going beyond point-to-point coding (and into genuine “network strategies”) can alter the scaling behavior of the capacity of wireless networks.

Finally, we also demonstrate how our results can be applied to sensor networks. There, the tradeoff is between sensor power and reconstruction fidelity, and is generally unknown to date. For a particular sensor network situation, we determine the optimal tradeoff using the arguments developed in this paper. We also demonstrate that this optimal tradeoff cannot be achieved by separate source and channel code design, illustrating the known fact that the source–channel separation principle does not extend to such sensor networks, and, in fact, furnishing new insight to the effect that source–channel separation can lead

to a substantially (exponentially, that is) suboptimal scaling behavior of performance as the network becomes large.

The bounding techniques developed in this paper can also be applied to more complex situations, such as a general wireless network with collaborating users. A simple extension of the work presented here is for the individual data sources to take turns. In its time (and/or frequency) slot, each source uses the network as a relay network. The respective performance can be lower-bounded using the techniques developed in this paper. Another scenario involves channel fading effects, and unknown parameters in the network. Here, the key insight is that the strategy developed to prove lower bounds has some degree of natural robustness to such effects. In particular, such robustness can be easily and conveniently exploited in the sensor network context described in Section VI.

APPENDIX

Proof of Proposition 4: At time n , the relay receives

$$Y_m[n] = \alpha_m X[n] + W_m[n] \quad (82)$$

and transmits in the next time slot

$$X_m[n+1] = e^{i\theta_m} \sqrt{\frac{P_m}{|\alpha_m|^2 P + N}} Y_m[n] \quad (83)$$

where i is the imaginary unit (the square root of -1), $\theta_m \in [0, 2\pi)$ an appropriately chosen phase, and P_m an appropriately chosen nonnegative real constant. Note that this recoding coefficient makes the expected power of relay m equal to P_m . The received random variable at time $n+1$ is, therefore,

$$Y[n+1] = \alpha_0 X[n+1] + \sum_{m=1}^M \beta_m (\alpha_m X[n] + W_m[n]) + W[n+1] \quad (84)$$

where β_m is defined as

$$\beta_m = \delta_m e^{i\theta_m} \sqrt{\frac{P_m}{|\alpha_m|^2 P + N}}. \quad (85)$$

Suppose that we use $\gamma Y[n+1]$ as the estimate of $X[n]$. Then

$$\begin{aligned} & E|X[n] - \gamma Y[n+1]|^2 \\ &= E \left| X[n] \left(1 - \gamma \sum_{m=1}^M \beta_m \alpha_m \right) - \gamma \alpha_0 X[n+1] \right. \\ &\quad \left. - \gamma \sum_{m=1}^M \beta_m W_m[n] - \gamma W[n+1] \right|^2 \\ &= P \left| 1 - \gamma \sum_{m=1}^M \beta_m \alpha_m \right|^2 + |\gamma|^2 |\alpha_0|^2 P \\ &\quad + |\gamma|^2 N \left(1 + \sum_{m=1}^M |\beta_m|^2 \right). \end{aligned} \quad (86)$$

Recall that α_0 is real by assumption, hence, $|\alpha_0|^2 = \alpha_0^2$. The optimal single-letter decoding function γ can be found by taking the derivative of (86) with respect to γ and setting this derivative equal to zero. This yields

$$\gamma_{\text{opt}} = \frac{P \sum_{m=1}^M \alpha_m \beta_m}{P \sum_{m=1}^M |\alpha_m \beta_m|^2 + \alpha_0^2 P + N(1 + \sum_{m=1}^M |\beta_m|^2)}$$

where \bar{x} denotes the complex conjugate of x . The minimum achievable distortion becomes

$$\begin{aligned} D_{\text{opt}} &= E|X[n] - \gamma_{\text{opt}} Y[n+1]|^2 \\ &= P \left| 1 - \gamma_{\text{opt}} \sum_{m=1}^M \beta_m \alpha_m \right|^2 \\ &\quad + |\gamma_{\text{opt}}|^2 \left(\alpha_0^2 P + N \left(1 + \sum_{m=1}^M |\beta_m|^2 \right) \right) \\ &= \frac{P \left(\alpha_0^2 P + N \left(1 + \sum_{m=1}^M |\beta_m|^2 \right) \right)}{P \left| \sum_{m=1}^M \alpha_m \beta_m \right|^2 + \alpha_0^2 P + N \left(1 + \sum_{m=1}^M |\beta_m|^2 \right)}. \end{aligned}$$

The next goal is to find suitable β_m 's, i.e., to find a good power allocation between the relays. To this end, we may rewrite D_{opt} using vector notation. Define the two vectors $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_M)$ and $\beta = (1, \beta_1, \dots, \beta_M)$. Then

$$D_{\text{opt}} = P \frac{\alpha_0^2 P + \langle \beta, \beta \rangle N}{|\langle \alpha, \bar{\beta} \rangle|^2 P + \langle \beta, \beta \rangle N} = P \frac{\frac{\alpha_0^2}{\|\beta\|^2} P + N}{\frac{|\langle \alpha, \bar{\beta} \rangle|^2}{\|\beta\|^2} P + N}. \quad (87)$$

Minimizing this expression over all vectors β under the given power constraint does not seem to have a simple solution. However, a good (but generally suboptimal) solution is found by recalling that

$$\max_{\beta: \|\beta\|^2 = \text{const}} \frac{\langle \alpha, \bar{\beta} \rangle^2}{\|\beta\|^2} = \langle \alpha, \alpha \rangle = \|\alpha\|^2 \quad (88)$$

and that the maximum is achieved when $\bar{\beta} = B\alpha$, for a scalar B .

For the purposes of this paper, this relay coding strategy is sufficient. That is, we pick the recoding function given by (83) such that

$$\beta_m = B\bar{\alpha}_m, \quad \text{for } k = 1, \dots, M \quad (89)$$

which makes the phase θ_m in (83)

$$\theta_m = -\arg \alpha_m - \arg \delta_m \quad (90)$$

and the power P_m of relay m

$$P_m = B^2 |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2} \quad (91)$$

where B must be chosen to satisfy the power constraint $\sum_{m=1}^M P_m \leq c(M)$, i.e., B^2 can be determined as

$$B^2 = \frac{c(M)}{\sum_{m=1}^M |\alpha_m|^2 \frac{|\alpha_m|^2 P + N}{|\delta_m|^2}}. \quad (92)$$

To simplify the expression for the distortion, we first point out that

$$\begin{aligned} \langle \alpha, \beta \rangle &= Ba(M) - B\alpha_0^2 + \alpha_0 \\ \langle \beta, \beta \rangle &= B^2 a(M) - B^2 \alpha_0^2 + 1. \end{aligned}$$

Using this in (87), the distortion can be expressed as

$$D_1 \stackrel{\text{def}}{=} P \frac{B^2(a(M) - \alpha_0^2)N}{B^2(a(M) - \alpha_0^2)^2 P + 2B(a(M) - \alpha_0^2)\alpha_0 P + \alpha_0^2 P + N} \frac{1}{+B^2(a(M) - \alpha_0^2)N + \alpha_0^2 P + N}. \quad (93)$$

Introducing in B from above yields the claimed result. \square

Remark 6 (Optimal (Multiple-Letter) Decoding): In Remark 2, we outlined how the considered coding scheme could be improved. Similarly, the decoding considered in the proof of Proposition 4 can also be improved. In fact, the two transmission steps of our coding scheme take place simultaneously. This acts like a convolutional code. Hence, the optimum decoder must consider all outputs $Y[n]$ simultaneously; it cannot operate on a single-letter basis. This does not seem to lead to simple expressions for the achieved distortion, and asymptotically, the improvement of the optimal decoder over the single-letter decoder should not be expected to be large: for many cases, the interfering term $\alpha_0 X[n+1]$ in (84) does not influence the asymptotic behavior.

Proof of Theorem 6: Recall

$$C_{\text{BC}} = \log_2 \left(1 + \frac{a(M)P}{N} \right). \quad (94)$$

We compare this to R_1 . Let us write out as follows:

$$\begin{aligned} 2^{C_{\text{BC}} - R_1} &= \left(\frac{B^2(a(M) - \alpha_0^2)N}{B^2(a(M) - \alpha_0^2)^2 P + 2B(a(M) - \alpha_0^2)\alpha_0 P + \alpha_0^2 P + N} \right) \frac{a(M)P + N}{N} \\ &= \frac{PNB^2 a^2(M) + N^2 B^2 a(M)}{PNB^2 a^2(M) + (N - 2\alpha_0^2 P)NB^2 a(M) + (\alpha_0^2 P + N)Pa(M) - \alpha_0^2 N^2 B^2 + 2\alpha_0 PNB a(M) + (\alpha_0^4 PN - \alpha_0^2 N^2)B^2 + N(\alpha_0^2 P + N)} \\ &\quad \frac{1}{-2\alpha_0^3 PNB + N(\alpha_0^2 P + N)}. \end{aligned} \quad (95)$$

Using (92), we can replace $B^2 = c(M)/b(M)$. Let us then multiply both numerator and denominator by $b(M)$ to obtain

$$\begin{aligned} 2^{C_{\text{BC}} - R_1} &= \frac{P^2 N c(M) a^2(M) + P N^2 c(M) a(M)}{P^2 N c(M) a^2(M) + (N - 2\alpha_0^2 P) P N c(M) a(M) + (\alpha_0^2 P + N) P^2 a(M) b(M) + 2\alpha_0 P^2 N \sqrt{c(M)} a(M) \sqrt{b(M)} + \alpha_0^3 P^2 N c(M) - \alpha_0^2 P N^2 c(M)} \\ &\quad \frac{1}{- \alpha_0^2 P N^2 c(M) - 2\alpha_0^4 P^2 N \sqrt{c(M)} b(M) + P N (\alpha_0^2 P + N) b(M) + P N (\alpha_0^2 P + N) b(M)} \\ &= \frac{1 + \frac{N}{P} \frac{1}{a(M)} + \frac{\alpha_0^2 P + N}{N} \frac{b(M)}{a(M)c(M)}}{1 + \frac{N - 2\alpha_0^2 P}{P} \frac{1}{a(M)} + 2\alpha_0 \frac{\sqrt{b(M)}}{a(M)\sqrt{c(M)}} - \alpha_0^2 \frac{N}{P} \frac{1}{a^2(M)}} \\ &\quad \frac{1}{+ (\alpha_0^3 - \alpha_0^2 \frac{N}{P}) \frac{1}{a^2(M)} - 2\alpha_0^4 \frac{\sqrt{b(M)}}{a^2(M)\sqrt{c(M)}} + \frac{\alpha_0^2 P + N}{P} \frac{b(M)}{a^2(M)c(M)} + \frac{\alpha_0^2 P + N}{P} \frac{b(M)}{a^2(M)c(M)}}. \end{aligned} \quad (96)$$

Next, we argue that under the stated assumptions, each summand both in the numerator as well as in the denominator of M

in (96) converges. By assumption, $1/a(M) \rightarrow \theta_a$, and hence, $1/a^2(M) \rightarrow \theta_a^2$. This implies the following:

$$\begin{aligned} & \lim_{M \rightarrow \infty} \frac{b(M)}{a^2(M)c(M)} \\ &= \left(\lim_{M \rightarrow \infty} \frac{1}{a(M)} \right) \left(\lim_{M \rightarrow \infty} \frac{b(M)}{a(M)c(M)} \right) \\ &= \theta_a \tau. \end{aligned}$$

Similarly

$$\begin{aligned} & \lim_{M \rightarrow \infty} \frac{\sqrt{b(M)}}{a(M)\sqrt{c(M)}} \\ &= \left(\lim_{M \rightarrow \infty} \sqrt{\frac{1}{a(M)}} \right) \left(\lim_{M \rightarrow \infty} \sqrt{\frac{b(M)}{a(M)c(M)}} \right) \\ &= \sqrt{\theta_a \tau}, \end{aligned} \quad (97)$$

which also implies that

$$\sqrt{b(M)/(a^2(M)c(M))} \xrightarrow{M \rightarrow \infty} \theta_a \sqrt{\theta_a \tau}.$$

Finally

$$\lim_{M \rightarrow \infty} \frac{\alpha_0^2 P + N}{N} \frac{b(M)}{a(M)c(M)} = \tau \frac{\alpha_0^2 P + N}{N} \quad (99)$$

leading to the following relationship:

$$\begin{aligned} & \lim_{M \rightarrow \infty} 2^{C_{\text{BC}} - R_1} \\ &= \frac{1 + \frac{N}{P} \theta_a + \frac{\alpha_0^2 P + N}{N} \tau}{1 + \frac{N - 2\alpha_0^2 P}{P} \theta_a + 2\alpha_0 \sqrt{\theta_a \tau} + (\alpha_0^3 - \alpha_0^2 \frac{N}{P}) \theta_a^2} \\ & \quad \frac{-\alpha_0^2 \frac{N}{P} \theta_a^2 + \frac{\alpha_0^2 P + N}{P} \theta_a \tau}{-2\alpha_0^4 \theta_a \sqrt{\theta_a \tau} + \frac{\alpha_0^2 P + N}{P} \theta_a \tau} \end{aligned} \quad (100)$$

which concludes the proof. \square

Proof of Theorem 8: Recall that

$$C_{\text{MAC}} = \log \left(1 + \frac{(P + c(M))d(M)}{N} \right). \quad (101)$$

We now compare this to R_1 . Let us write out as follows:

$$\begin{aligned} 2^{C_{\text{MAC}} - R_1} &= \left(\frac{B^2 N (a(M) - \alpha_0^2)}{B^2 (a(M) - \alpha_0^2)^2 P + 2B (a(M) - \alpha_0^2) \alpha_0 P} \right. \\ & \quad \left. + \frac{\alpha_0^2 P + N}{B^2 (a(M) - \alpha_0^2) N + \alpha_0^2 P + N} \right) \\ & \quad \cdot \frac{(P + c(M))d(M) + N}{N}. \end{aligned}$$

Using (92), we can replace $B^2 = c(M)/b(M)$. Let us then multiply both numerator and denominator by $b(M)/c(M)$ to obtain

$$\begin{aligned} & 2^{C_{\text{MAC}} - R_1} \\ &= \frac{Na(M)c(M)d(M) - N\alpha_0^2 c(M)d(M)}{PNa^2(M) - 2PN\alpha_0^2 a(M) + 2PN\alpha_0^4} \\ & \quad \frac{+PNa(M)d(M) - PN\alpha_0^2 d(M) + N^2 a(M) - N^2 \alpha_0^2}{+2\alpha_0 PN a(M) \sqrt{\frac{b(M)}{c(M)}} - 2\alpha_0^3 N \sqrt{\frac{b(M)}{c(M)}}} \\ & \quad \frac{+(\alpha_0^2 P + N) \left(b(M)d(M) + P \frac{b(M)d(M)}{c(M)} + N \frac{b(M)}{c(M)} \right)}{+N^2 a(M) - \alpha_0^2 N^2 + (\alpha_0^2 PN + N^2) \frac{b(M)}{c(M)}}. \end{aligned}$$

Dividing both the numerator and the denominator by $a^2(M)$ yields the desired form. By assumption

$$\lim_{M \rightarrow \infty} \frac{c(M)d(M)}{a(M)} = \tau_1 < \infty$$

and

$$\lim_{M \rightarrow \infty} \frac{b(M)d(M)}{a^2(M)} = \tau_2 < \infty. \quad (102)$$

Under these assumptions, we find that

$$\lim_{M \rightarrow \infty} \sqrt{\frac{b(M)}{c(M)}} \frac{1}{a(M)} = \sqrt{\tau_2 \theta_c \theta_d}. \quad (103)$$

Finally

$$\begin{aligned} & 2^{C_{\text{MAC}} - R_1} \\ &= \frac{N\tau_1 - N\alpha_0^2 \tau_1 \theta_a + PN\tau_1 \theta_c - PN\alpha_0^2 \tau_1 \theta_c \theta_a +}{PN - 2PN\alpha_0^2 \theta_a + 2PN\alpha_0^4 \theta_a^2 + 2\alpha_0 PN \sqrt{\tau_2 \theta_c \theta_d} -} \\ & \quad \frac{+N^2 \theta_a - N^2 \alpha_0^2 \theta_a^2}{-2\alpha_0^3 N \sqrt{\tau_2 \theta_c \theta_d} \theta_a + N^2 \theta_a - \alpha_0^2 N^2 \theta_a^2} \\ & \quad \frac{+(\alpha_0^2 P + N) (\tau_2 + P\tau_2 \theta_c + N\tau_2 \theta_c \theta_d)}{+(\alpha_0^2 PN + N^2) \tau_2 \theta_c \theta_d} \end{aligned} \quad (104)$$

which concludes the proof. \square

ACKNOWLEDGMENT

Fruitful discussions with Dr. G. Kramer (Bell Labs) are gratefully acknowledged. The insightful comments of the anonymous reviewers improved the quality of this manuscript.

REFERENCES

- [1] M. Gastpar and M. Vetterli, "On the capacity of wireless networks: The relay case," in *Proc. IEEE INFOCOM 2002*, New York, Jun. 2002.
- [2] —, "On the asymptotic capacity of Gaussian relay networks," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 195.
- [3] E. C. van der Meulen, "Transmission of information in a T -terminal discrete memoryless channel," Ph.D. dissertation, Univ. Calif., Berkeley, CA, 1968.
- [4] —, "Three-terminal communication channels," *Adv. Appl. Probab.*, vol. 3, pp. 120–154, 1971.
- [5] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
- [6] P. P. Bergmans, "Random coding theorems for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 197–207, Mar. 1973.
- [7] B. Schein and R. G. Gallager, "The Gaussian parallel relay network," in *Proc IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 22.
- [8] P. Gupta and P. R. Kumar, "Toward an information theory of large networks: An achievable rate region," in *Proc IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 159.
- [9] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 136.
- [10] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc IEEE Int. Symp. Information Theory*, Sorrento, Italy, Jun. 2000, p. 40.
- [11] 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. 2–4, 2002.
- [12] G. Kramer, M. Gastpar, and P. Gupta, "Capacity theorems for wireless relay channels," in *Proc 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1074–1083.

- [13] M. A. Khojastepour, A. Sabharwal, and B. Aazhang, "On the capacity of 'cheap' relay networks," in *Proc 37th Annu. Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 12–14, 2003.
- [14] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," *IEEE Trans. Inf. Theory*, submitted for publication.
- [15] A. A. El Gamal and M. Aref, "The capacity of the semideterministic relay channel," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 3, p. 536, May 1982.
- [16] Z. Zhang, "Partial converse for a relay channel," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 1106–1110, Sep. 1988.
- [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [18] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
- [19] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.
- [20] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Sel. Areas Commun.* (Special Issue on Self-Organizing Distributive Collaborative Sensor Networks), 2005, to be published.
- [21] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.
- [22] İ. E. Telatar, "Capacity of multi-antenna gaussian channels," Bell Labs, Tech. Memo., 1995. Also published in *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov./Dec. 1999.
- [23] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
- [24] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [25] M. Gastpar, "To code or not to code," Ph.D. dissertation, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.
- [26] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.
- [27] L.-L. Xie and P. Kumar, "A network information theory for wireless communications: Scaling laws and optimal operation," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 748–767, May 2004.
- [28] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
- [29] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977.
- [30] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *Proc. 2nd Int. Workshop Information Processing in Sensor Networks (IPSN'03) (Lecture Notes in Computer Science)*, L. J. Guibas and F. Zhao, Eds. Berlin, Germany: Springer-Verlag, 2003, pp. 162–177.
- [31] —, "Scaling laws for homogeneous sensor networks," in *Proc 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 1084–1091.
- [32] —, "Power-bandwidth-distortion scaling laws for sensor networks," in *Proc. 3rd Int. Symp. Information Processing in Sensor Networks (IPSN'04)*, Berkeley, CA, Apr. 2004, pp. 320–329.
- [33] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.
- [34] T. M. Cover, A. A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.