

Carbon Copying Onto Dirty Paper

Ashish Khisti, *Student Member, IEEE*, Uri Erez, *Member, IEEE*,
Amos Lapidoth, *Fellow, IEEE*, Gregory W. Wornell *Fellow, IEEE*

Abstract—A generalization of the problem of writing on dirty paper is considered in which one transmitter sends a common message to multiple receivers. Each receiver experiences on its link an additive interference, which is known noncausally to the transmitter but not to any of the receivers. Applications range from wireless multi-antenna multicasting to robust dirty paper coding.

We develop results for memoryless channels in Gaussian and noiseless binary special cases. In general, we show that the availability of side information at the transmitter increases capacity relative to systems without such side information, and that the lack of side information at the receivers decreases capacity relative to systems with such side information.

For the noiseless binary case, we derive the capacity when there are two receivers. When there are many receivers, we show that the side information provides a vanishingly small benefit. When the interference is large, we show that time sharing is optimal.

For the Gaussian case with two users and independent interferences, we provide upper and lower bounds on capacity. At high interference-to-noise ratios, we show that time-sharing is (asymptotically) optimal. This settles the conjecture by Steinberg and Shamai [22]. At high signal-to-noise ratios, we propose a superposition dirty paper code that achieves within $1/4$ bit/symbol of capacity. Extensions and generalizations are also discussed.

I. INTRODUCTION

The study of communication over channels controlled by a random state parameter known only to the transmitter was initiated by Shannon [20]. Shannon considered the case where the state sequence is known causally at the encoder. Subsequently, Gel'fand and Pinsker [10] analyzed the case where the state sequence is available noncausally. The noncausal model has found application in diverse areas, ranging from coding for memory with defects [12], [17], to digital watermarking [3], [4], [19], and to coding for the multiple-input/multiple-output (MIMO) broadcast channel [1], [23].

Costa [6] considered a version of the Gel'fand-Pinsker model in which there is an additive white Gaussian interference (“dirt”), which constitutes the state, in addition to independent additive white Gaussian noise. The key result in this “dirty paper coding” scenario is that there is no loss in capacity if the interference is known only to the transmitter.

By contrast, there has been very limited work to date on *multiuser* channels with state parameters known to the

transmitter(s). In an early work in this area, Gel'fand and Pinsker [11] show that the Gaussian broadcast channel with *independent messages* incurs no loss in capacity if the interference sequences are known noncausally to the transmitter. More recently, the degraded broadcast channel with independent messages and state sequence known to the transmitter is examined in [21], while the multiple-access channel with state sequence known to one or more of the transmitters is examined in, e.g., [16], [2], [15].

This paper examines the *common-message* broadcast channel, which we refer to as the *multicast* channel. Specifically, we consider a scenario in which one transmitter broadcasts a common message to multiple receivers. In the case of interest, associated with the link to each receiver is a corresponding additive interference. The collection of such interferences is thus the (random) state of the multiuser channel. In our model, the transmitter has perfect noncausal knowledge of all these interference sequences, but none of the receivers have knowledge of any of them. This model and its generalizations arise rather naturally not only in a variety of multi-antenna wireless multicasting problems, but also in applications of dirty paper coding where only imperfect knowledge of the state is available to the transmitter.

The capacity of some binary versions of such multicast channels is reported in [14] [13]. For more general channels, [22] reports achievable rates for broadcasting common and independent messages over a discrete memoryless channel with noncausal state knowledge at the transmitter. The case of two-user Gaussian channels with jointly and individually independent identically distributed (i.i.d.) Gaussian interferences on each link is also considered in [22], for which it is conjectured that in the limit of large interference, time-sharing between the two receivers is optimum even when both are only interested in a common message. Among other results, in this paper we establish that this conjecture is true. We upper bound the capacity of the Gaussian channel and show that it approaches the time-sharing rate in this limit. In addition, we also present a coding scheme that achieves within $1/4$ bit/symbol of the capacity¹ in the high signal to noise ratio (SNR) limit for any interference power.

An outline of the paper is as follows. Section II presents the general multicast channel model of interest. The binary special cases of interest are analyzed in Section III, and the Gaussian special cases of interest are analyzed in Section IV. Finally, Section V contains some conclusions and directions for future work.

¹Throughout this work, symbol refers to a *real* symbol.

This work has been supported in part by the National Science Foundation under Grant No. CCF-0515109, and by Hewlett-Packard through the MIT/HP Alliance. This work was presented in part at the International Symposium on Information Theory, Chicago, IL, June 2004.

A. Khisti and G. W. Wornell are with the Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA (E-mail: {khisti,gww}@mit.edu). U. Erez is with the Department of Electrical Engineering-Systems, Tel Aviv University, Ramat Aviv, 69978, Israel (E-mail: uri@eng.tau.ac.il). A. Lapidoth is with the Institute for Information and Signal Processing, Swiss Federal Institute of Technology (ETH) – Zurich, CH-8092, Switzerland (E-mail: lapidoth@isi.ee.ethz.ch).

II. MULTICAST CHANNEL MODEL

The K -user multicast channel of interest is defined as follows.

Definition 1: A K -user discrete memoryless multicast channel with random parameters consists of an input alphabet \mathcal{X} , output alphabets $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_K$ for receivers $1, 2, \dots, K$, respectively, and a state alphabet \mathcal{S} . For a given state sequence² $s^n = (s_1, s_2, \dots, s_n)$ such that $s_i \in \mathcal{S}$ and input $x^n = (x_1, x_2, \dots, x_n)$ such that $x_i \in \mathcal{X}$, the channel outputs are distributed according to

$$p(y_1^n, y_2^n, \dots, y_K^n | x^n, s^n) = \prod_{i=1}^n p(y_{1i}, y_{2i}, \dots, y_{Ki} | x_i, s_i) \quad (1)$$

where $y_k^n = (y_{k1}, y_{k2}, \dots, y_{kn})$, for all $y_{ki} \in \mathcal{Y}_k$, $k = 1, 2, \dots, K$. Moreover, $p(s^n) = \prod_i p(s_i)$. The particular realization s^n is known noncausally to the transmitter before using the channel, but not to any of the K receivers.

It is worth emphasizing that the above definition includes the case where the channel of user k is controlled by its own state s_k^n . In such cases, the joint state is, with slight abuse of notation, $s^n = (s_1^n, s_2^n, \dots, s_K^n)$, so that $p(s_i) = p(s_{1i}, s_{2i}, \dots, s_{Ki})$.

The capacity of the channel of Definition 1 is defined as follows.

Definition 2: A $(2^{nR}, n)$ code consists of a message set $\mathcal{W} = \{1, 2, \dots, 2^{nR}\}$, an encoder $f_n : \mathcal{W} \times \mathcal{S}^n \rightarrow \mathcal{X}^n$, and K decoders $g_{k,n} : \mathcal{Y}_k^n \rightarrow \mathcal{W}$. The rate R is *achievable* if there exists a sequence of codes such that for W uniformly distributed over \mathcal{W} we have

$$\lim_{n \rightarrow \infty} P_e^n = \lim_{n \rightarrow \infty} \Pr \left\{ \bigcup_{k=1}^K \{g_{k,n}(Y_k^n) \neq W\} \right\} = 0. \quad (2)$$

Note that the error probability in (2) is averaged over all state sequences and messages. The capacity C is the supremum of achievable rates.

In the remainder of the paper, we focus on special cases of the memoryless channel in Definition 1. In particular, we focus on binary and Gaussian cases in which the state is an additive interference; for results on the memory with defects multicast channel, see, e.g., [13].

III. NOISELESS BINARY CASE

We first consider the noiseless binary special case of Definition 1. Specifically, the channel outputs $Y_1^n, Y_2^n, \dots, Y_K^n$ depend on the input X^n and the states $S_1^n, S_2^n, \dots, S_K^n$ according to

$$Y_k^n = X^n \oplus S_k^n \quad (3)$$

where $X_i, S_{ki} \in \{0, 1\}$, and where \oplus denotes symbol-by-symbol modulo-two addition (i.e., exclusive-or). In (3), the memoryless case of interest corresponds to the requirement that the $(S_{1i}, S_{2i}, \dots, S_{Ki})$ for $i = 1, 2, \dots, n$ form an i.i.d. sequence of K -tuples. In particular, for each i the variables $\{S_{1i}, S_{2i}, \dots, S_{Ki}\}$ may in general be statistically dependent,

²More generally throughout the paper the superscript notation n indicates an n -vector.

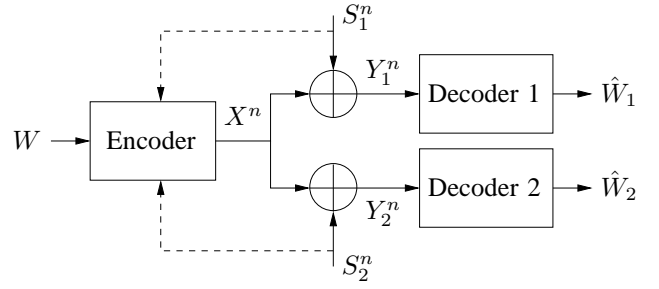


Fig. 1. Two-user memoryless, noiseless binary multicast channel with additive interference. The encoder maps message W into codeword X^n . The state takes the form of interference sequences S_1^n and S_2^n . Each channel output $Y_k^n = X^n \oplus S_k^n$, where \oplus denotes symbol-by-symbol modulo-two addition, is decoded to produce message estimate \hat{W}_k .

and do not need to be identically distributed. As a result, we express our results in terms of the properties of a generic K -tuple in this sequence, which we denote by (S_1, S_2, \dots, S_K) .

Note that with only a single receiver ($K = 1$), the capacity is trivially 1 bit/symbol, which is achieved by interference precancellation, i.e., by choosing $X^n = S^n \oplus B^n$, so that $Y^n = B^n$, where B^n is the bit representation for the message W . As we will now develop, when there are multiple receivers, capacity is generally less than this ideal single-user rate.

A. The Case of $K = 2$ Receivers

The case of two receivers, which is depicted in Fig. 1, is the simplest nontrivial scenario since perfect interference precancellation is not possible simultaneously for both users.

One lower bound on the two-user capacity corresponds to a time-sharing approach that precancels the interference of one of the receivers at a time, yielding a rate of $R_{\text{TS}} = 1/2$ bit/symbol. Another lower bound corresponds to ignoring the interference at the transmitter, i.e., treating each of the channels as a binary symmetric channel. This strategy yields a rate of $R_{\text{IS}} = 1 - \max\{H(S_1), H(S_2)\}$ bit/symbol. It turns out that the former bound is only tight in the high interference regime, and the latter bound is only tight in the low interference regime.

A coding theorem for the channel is as follows.

Theorem 1: The capacity of two-user noiseless, memoryless binary channel with additive interference is given by

$$C = 1 - \frac{1}{2}H(S_1 \oplus S_2). \quad (4)$$

Proof: A converse is provided in Appendix I. The achievability follows from an application of the following lemma.

Lemma 1: The capacity of the K -user multicast channel with random parameters is lower bounded by

$$C \geq \max_{p(U|S), p(X|U,S)} \min_k \{I(U; Y_k) - I(U; S)\}, \quad (5)$$

where U is an auxiliary random variable (over some alphabet \mathcal{U}) that satisfies the Markov constraint $U \leftrightarrow (X, S) \leftrightarrow Y_k$ for $k = 1, 2, \dots, K$.

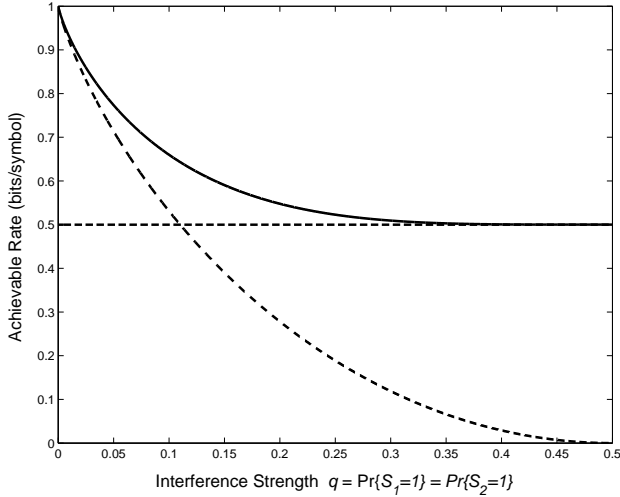


Fig. 2. Achievable rates for the two-user noiseless binary multicast channel with independent interferences, as a function of the strength of the interference. Capacity is indicated by the solid curve, time-sharing performance is indicated by the horizontal dashed line, and the performance of a system that ignores the side information is indicated by the downward sloping dashed curve.

To prove Lemma 1 it suffices to recognize the lower bound (5) as the rate obtained by a straightforward extension of the achievability scheme used in [10].

Our achievability result, in turn, follows by a suitable choice of the auxiliary random variable in (5) with $K = 2$. Specifically, we choose $\mathcal{U} = \{U_1, U_2, U_3, U_4\}$ and let

$$U = A \{U_1 (X \oplus S_1) + U_2 (\overline{X \oplus S_1})\} + \bar{A} \{U_3 (X \oplus S_2) + U_4 (\overline{X \oplus S_2})\}, \quad (6)$$

where A is a random variable taking values in $\{0, 1\}$ with equal probability and that is independent of X , S_1 and S_2 , and where $\bar{\cdot}$ denotes the complement of a (binary-valued) variable. ■

Fig. 2 shows the performance gains of optimal coding relative to time-sharing and disregarding the side-information. In particular, the achievable rate in the case of independent interferences is plotted as a function of the strength of the interference as measured by $q = \Pr\{S_1 = 1\} = \Pr\{S_2 = 1\}$.

Three immediate conclusions can be drawn from Theorem 1. First, transmitter-only side information incurs a penalty relative to system-wide side information unless S_1 and S_2 are completely dependent random variables, i.e., unless $S_2 = S_1$ or $S_2 = \overline{S_1}$. This follows from the fact that the capacity of the latter is obviously 1 bit/symbol since each receiver can postcancel its own interference. Second, time-sharing is strictly sub-optimal except when S_1 and S_2 are independent $\mathcal{B}(1/2)$ random variables.³ We emphasize that, by contrast, when there are *independent* messages for each of the receivers in Fig. 1, time-sharing between the receivers is optimal and there is no loss in the capacity region with side information only at the transmitter. Finally ignoring the side information at the transmitter is strictly suboptimal except when $H(S_1) = H(S_2) = 0$.

³We use $\mathcal{B}(q)$ to denote a Bernoulli random variable with parameter q .

The auxiliary random variable in this proof implies the following intuitively rather natural code construction:

- 1) Select 2^{nR} codewords randomly according to an i.i.d. $\mathcal{B}(1/2)$ distribution in a codebook \mathcal{C} of rate R strictly less than the capacity (4). Denote these codewords as $B^n(1), B^n(2), \dots, B^n(2^{nR})$, so a message w is represented by codeword $B^n(w)$.
- 2) Select a sequence A^n by flipping a fair coin for each symbol index (the realization of which is also known at the decoders [24]). Select the set \mathcal{A}_1 of symbol indices where $A_i = 1$, and precancel the interference at those indices for user 1, and precancel the interference at the remaining indices \mathcal{A}_2 for user 2. Specifically, the transmitted codeword is of the form⁴

$$X_i(w) = \begin{cases} B_i(w) \oplus S_{1i} & i \in \mathcal{A}_1 \\ B_i(w) \oplus S_{2i} & i \in \mathcal{A}_2. \end{cases} \quad (7)$$

With this encoding, receiver 1 then observes a version of $B^n(w)$ where $|\mathcal{A}_1|$ symbols are correct, and the remaining $|\mathcal{A}_2|$ symbols are corrupted by interference $S_{1i} \oplus S_{2i}$, $i \in \mathcal{A}_2$, corresponding to a binary symmetric channel with crossover probability $q' = \Pr\{S_1 \oplus S_2 = 1\}$. Receiver 2 experiences the opposite effect. Thus for large n we have, since $|\mathcal{A}_1|/n \rightarrow 1/2$,

$$\frac{1}{n} I(B^n; Y_k^n | A^n) \rightarrow \frac{1}{2} + \frac{1}{2} (1 - H(S_1 \oplus S_2)), \quad k = 1, 2, \quad (8)$$

which is C in (4). The decoding of Y_k^n to the message \hat{W}_k is done by exploiting the knowledge of \mathcal{A}_1 and \mathcal{A}_2 (i.e., A^n) at the decoders. As long as $R \leq C$, \hat{W}_k equals W with high probability.

We conclude this section with a few additional observations. *Some Further Remarks:*

- 1) Note that the code construction outlined above suggests that the transmitter does not require noncausal knowledge of the interference. In fact, it can be verified that $I(U; S_1) = I(U; S_2) = 0$, and thus *causal* knowledge of S_1^n and S_2^n is sufficient to achieve the capacity. We emphasize, however, this result is specific to the noiseless binary channel model.
- 2) It is straightforward to verify that random linear codes are sufficient to achieve the capacity of Theorem 1. It suffices to use an argument analogous to that used by Gallager for the binary symmetric channel [9, Sec. 6.2].
- 3) Theorem 1 can be readily generalized to the case of state sequences that are not in general i.i.d. In this case the term $H(S_1 \oplus S_2)$ in (4) is simply replaced with the entropy rate of $S_1^n \oplus S_2^n$.
- 4) Theorem 1 cannot be readily generalized to the case in which there is additional noise present on each link. In particular, for a discrete memoryless channel model of the form

$$\begin{aligned} Y_1 &= X \oplus S_1 \oplus Z_1 \\ Y_2 &= X \oplus S_2 \oplus Z_2, \end{aligned}$$

⁴In practice, it suffices to follow a deterministic strategy, such as precanceling the interference in the first half of the codeword for user 1, and in the second half for user 2.

where Z_1 and Z_2 are mutually independent and identically distributed Bernoulli random variables and independent of all other variables, the converse does not generalize. However, it is straightforward to generalize the direct part of the theorem to show that a rate

$$R = 1 - \frac{1}{2}H(S_1 \oplus S_2 \oplus Z_1) - \frac{1}{2}H(Z_1)$$

is achievable. It is also possible to show, in the special case where S_1 and S_2 are independent $\mathcal{B}(1/2)$ random variables, that time-sharing is optimal⁵.

B. The Case of $K > 2$ Receivers

When there are more than two receivers further losses in capacity ensue, as we now develop. Specifically, we have the following bounds on capacity.

Theorem 2: The capacity of the K -user noiseless binary channel in which the generic S_1, S_2, \dots, S_K are mutually independent and identically distributed⁶ is bounded according to:

$$R_- \leq C \leq R_+, \quad (9a)$$

where

$$R_+ = 1 - \frac{1}{K}H(S_1 \oplus S_2, S_1 \oplus S_3, \dots, S_1 \oplus S_K), \quad (9b)$$

$$R_- = \max \left\{ 1 - H(S_1), 1 - \left(1 - \frac{1}{K} \right) H(S_1 \oplus S_2) \right\}. \quad (9c)$$

Proof: The upper bound (9b) is established in Appendix II. The lower bound (9c) is obtained via a direct generalization of the code construction (7) in the case of two users. Specifically, it suffices to consider a code construction that divides each codeword into K equally sized blocks and precancels the interference for a different user in each of the blocks. Each user then experiences one clean block and $K - 1$ noisy blocks governed by a binary symmetric channel with crossover probability $q' = \Pr\{S_1 \oplus S_2 = 1\}$ as before. ■

In general, the lower and upper bounds in (9) do not coincide.⁷ However, the associated rate gap decreases monotonically with the number of receivers K . Moreover, even for $K = 3$, it is small, as Fig. 3 illustrates.

The rate gap also decays to zero in the limit of large K , which follows readily from Theorem 2. In particular, $C \rightarrow 1 - H(S)$ as $K \rightarrow \infty$, where S denotes a generic random variable with the distribution of the S_k . To see this, it suffices to recognize that when S_1, S_2, \dots, S_K are i.i.d.,

$$\begin{aligned} \left(1 - \frac{1}{K} \right) H(S) &\leq \frac{1}{K} H(S_1 \oplus S_2, S_1 \oplus S_3, \dots, S_1 \oplus S_K) \\ &\leq H(S). \end{aligned} \quad (10)$$

⁵The converse argument mirrors the one used in the Gaussian case in the next section and will not be presented here.

⁶Our results actually hold more generally provided the distribution across the interference sequences is symmetric, i.e., if for all m , $p(s_{k_1}, s_{k_2}, \dots, s_{k_m})$ is independent of the specific choice of $k_1, k_2, \dots, k_m \in \{1, 2, \dots, K\}$.

⁷A slightly improved lower bound appears in [13], but it, too, does not match the upper bound.

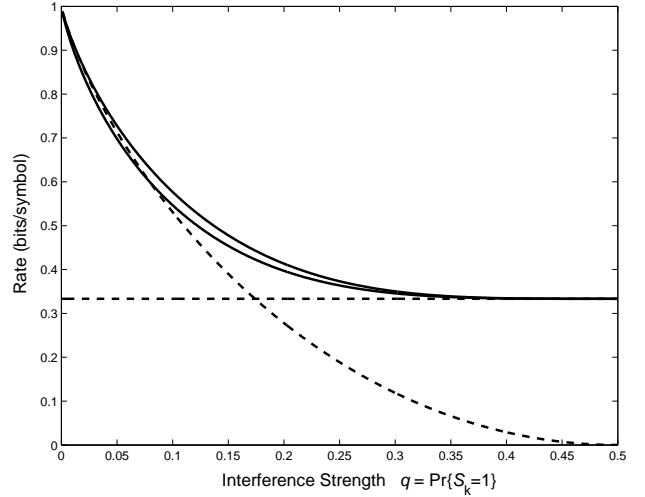


Fig. 3. Upper bound and lower bounds on the capacity of the three-user noiseless binary multicast channel, as a function of the strength of the interference. The solid curves depict the two bounds of (9). The horizontal dashed line indicates the performance of time-sharing, while the other dashed curve indicates the performance of a strategy in which the side information is ignored by the transmitter.

As $K \rightarrow \infty$, the lower and upper bounds in (10) converge, so that the upper bound on capacity (9b) converges to $R_+ = 1 - H(S)$. However, this rate is achievable by simply treating the interference as noise at the receivers, so it is the limiting capacity. It should be emphasized that this implies that when the number of receivers is large, the side-information available to the transmitter is essentially useless.

We can also use (10) to bound the rate penalty associated with ignoring side information as a function of the number of receivers K . In particular, the gap is at most $H(S)/K$.

Finally, we can use Theorem 2 to establish that in the limit of large interference, time-sharing is optimal for every K . Specifically, when $S_k \sim \mathcal{B}(1/2)$, the capacity is $C = 1/K$ and is achieved through time-sharing. To see this, it suffices to specialize the upper bound in (9b). Specifically, $S_1 \oplus S_k$ for $k = 2, 3, \dots, K$ are independent $\mathcal{B}(1/2)$ random variables, so the joint entropy is $K - 1$.

IV. GAUSSIAN CASE

In this section we consider a memoryless Gaussian special case of the general model of Definition 1 and incorporate an average power constraint on the input. We restrict our attention to the two-user ($K = 2$) case. In the scenario of interest, depicted in Fig. 4, the state is additive, and the associated interferences S_k^n are zero-mean white Gaussian sequences of power Q . Unless otherwise specified we focus on the case of independent interferences. In addition, each receiver's link also has a zero-mean additive white Gaussian noise Z_k^n of power N . Thus, the observation at receiver k takes the form

$$Y_k^n = X^n + S_k^n + Z_k^n, \quad k = 1, 2. \quad (11)$$

Our power constraint takes the form

$$\frac{1}{n} \sum_{i=1}^n x_i^2(w) \leq P, \quad \forall w \in \mathcal{W}. \quad (12)$$

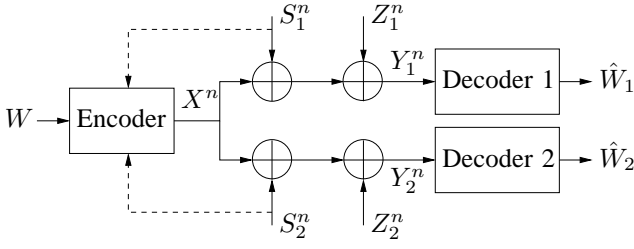


Fig. 4. Two-user Gaussian multicast channel model with additive interference. The encoder maps message W into codeword X^n . The state takes the form of interference sequences S_1^n and S_2^n . Each channel output $Y_k^n = X^n + S_k^n + Z_k^n$ is decoded to produce message estimate \hat{W}_k .

Finally, note that without loss of generality, we may set $N = 1$, and interpret P as the signal-to-noise ratio (SNR), and Q as the interference-to-noise ratio (INR).

For this channel, we have the following bounds on capacity.

Theorem 3: The Gaussian multicast channel capacity is bounded according to:

$$R_- \leq C \leq R_+, \quad (13)$$

where^{8,9}

$$R_+ = \begin{cases} \frac{1}{4} \log(1+P) + \frac{1}{4} \log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q}\right) & Q \geq 4 \\ \frac{1}{4} \log\left(\frac{1+P}{Q/4+1}\right) + \frac{1}{4} \log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/4+1}\right) & Q < 4 \end{cases} \quad (14)$$

and

$$R_- = \begin{cases} \frac{1}{2} \log\left(1 + \frac{P}{Q/2+1}\right) & Q/2 < 1 \\ \frac{1}{2} \log\left(\frac{P+Q/2+1}{Q}\right) + \frac{1}{4} \log\left(\frac{Q}{2}\right) & 1 \leq Q/2 < P+1 \\ \frac{1}{4} \log(1+P) & Q/2 \geq P+1. \end{cases} \quad (15)$$

Proof: The upper bound is established in Appendix III. It follows by considering a single-interference channel instead of the two-interference channel in Fig. 4 and establishing an upper bound for this channel.

The lower bound¹⁰ (15) is an explicit expression of the following maximization:

$$R_- = \max_{\{(P_S, P_V): P_S \geq 0, P_V \geq 0, P_S + P_V \leq P\}} R(P_S, P_V) \quad (16a)$$

with

$$R(P_S, P_V) \triangleq \frac{1}{2} \log\left(1 + \frac{P_S}{P_V + Q/2 + 1}\right) + \frac{1}{4} \log(1 + P_V). \quad (16b)$$

Accordingly, we show the achievability of (16b). The proposed scheme, which combines superposition coding, dirty

⁸All logarithms are to the base 2 in this work.

⁹We can also tighten the upper bound further by considering $\min(R_+, \frac{1}{2} \log(1+P))$, where the second expression corresponds to the multicasting rate when the interference is absent.

¹⁰Our lower bound for $Q/2 < 1$ was also independently reported by Costa [5].

paper coding, and time-sharing, and exploits a representation of the interferences in the form

$$\begin{aligned} S_1^n &= S^n + V^n \\ S_2^n &= S^n - V^n, \end{aligned} \quad (17)$$

where

$$\begin{aligned} S^n &= (S_1^n + S_2^n)/2 \\ V^n &= (S_1^n - S_2^n)/2. \end{aligned} \quad (18)$$

Specifically, the encoding is as follows:

- 1) Decompose the message W into two submessages W_S and W_V and divide the power P into two powers P_S and P_V so that $P = P_S + P_V$.
- 2) Use dirty paper coding [6] to (independently) encode message W_S into a codeword U_S^n in a (random i.i.d. Gaussian) codebook \mathcal{C}_S using power P_S for a channel with interference S^n and Gaussian noise of power $P_V + Q/2 + 1$. Transmit $X_S^n = U_S^n - \alpha_S S^n$, where $\alpha_S = P_S/(P + Q/2 + 1)$. The corresponding rate is

$$R_S = \frac{1}{2} \log\left(1 + \frac{P_S}{P_V + Q/2 + 1}\right). \quad (19)$$

- 3) Use dirty paper coding to encode message W_V into a codeword U_V^n in a codebook \mathcal{C}_V using power P_V for a channel with interference either $(1 - \alpha_S)S^n + V^n$ or $(1 - \alpha_S)S^n - V^n$, and Gaussian noise of power 1. Transmit the corresponding $X_V^n = U_V^n - \alpha_V\{(1 - \alpha_S)S^n + V^n\}$ or $X_V^n = U_V^n - \alpha_V\{(1 - \alpha_S)S^n - V^n\}$, with $\alpha_V = P_V/(P_V + 1)$. Time-share evenly between these two possibilities. In each case the corresponding rate is

$$R_V = \frac{1}{2} \log(1 + P_V). \quad (20)$$

- 4) Send the superposition $X^n = X_S^n + X_V^n$, which has power P over the channel.

The decoding exploits successive cancellation (stripping) and proceeds as follows:

- 1) Decode U_S^n from Y_1^n or Y_2^n treating X_V^n as part of the noise. The received signals are of the form

$$\begin{aligned} Y_1^n &= X_S^n + S^n + (V^n + Z_1^n + X_V^n) \\ &= U_S^n + (1 - \alpha_S)S^n + (V^n + Z_1^n + X_V^n), \\ Y_2^n &= X_S^n + S^n + (-V^n + Z_2^n + X_V^n) \\ &= U_S^n + (1 - \alpha_S)S^n + (-V^n + Z_2^n + X_V^n). \end{aligned}$$

The rate R_S in (19) ensures that the resulting \hat{W}_S equals W_S with high probability.

- 2) Subtract the decoded U_S^n from each of Y_1^n and Y_2^n , so that the residual signals $\tilde{Y}_i^n = Y_i^n - U_S^n$ are of the form

$$\tilde{Y}_1^n = X_V^n + ((1 - \alpha_S)S^n + V^n) + Z_1^n, \quad (21)$$

$$\tilde{Y}_2^n = X_V^n + ((1 - \alpha_S)S^n - V^n) + Z_2^n. \quad (22)$$

The rate R_V in (20) ensures that X_V^n can be decoded from either \tilde{Y}_1^n or \tilde{Y}_2^n so that the resulting \hat{W}_V equals W_V with high probability at the corresponding receiver. Specifically, for the fraction of time that the transmitter encodes W_V for interference $(1 - \alpha_S)S^n + V^n$, user 1 can recover W_V , while for the fraction of

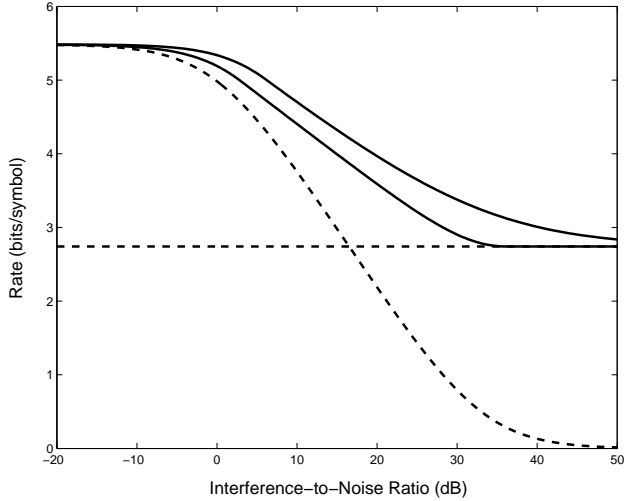


Fig. 5. Upper and lower bounds on the capacity of the two-user Gaussian multicast channel, as a function of INR Q for an SNR $P = 33$ dB. The solid curves depict the two bounds of (13). The horizontal dashed line indicates the performance of time-sharing, while the other dashed curve indicates the performance of a strategy in which the side information is treated by the transmitter as additional noise on each link.

time that the transmitter encodes W_V for interference $(1 - \alpha_S)S^n - V^n$, user 2 can recover W_V .

From this coding strategy, we see that the average rate delivered to each receiver is identical, i.e., $R_S + (1/2)R_V$. Maximizing this rate over the choices of P_S and P_V subject to the constraint $P = P_S + P_V$ optimizes the lower bound, whence (16a).

From (15), we obtain several useful insights. First, note that in the high INR regime ($Q/2 \geq P + 1$), our lower bound reduces to time-sharing, while in the low INR regime ($Q/2 \leq 1$) it reduces to dirty paper coding with respect to S^n . In the moderate interference regime, our bound shows that one can generally achieve a gain over these two strategies by a superposition coding approach that combines them.

The behavior of the bounds as a function of INR is depicted in Fig. 5 for a fixed SNR of $P = 33$ dB. When the INR is very small, Fig. 5 reflects that the side information can be ignored by the transmitter without sacrificing rate. Similarly, when the INR is large, Fig. 5 reflects that time-sharing is capacity-achieving. Both these observations can be quantified.

For the low INR behavior, it suffices to note from (14) that

$$\lim_{Q \rightarrow 0} C \leq \lim_{Q \rightarrow 0} R_+ = \frac{1}{2} \log(1 + P), \quad (23)$$

which is the same as $\lim_{Q \rightarrow 0} R_{\text{IS}}$, where R_{IS} is the rate achieved by ignoring the side-information, i.e.,

$$R_{\text{IS}} = \frac{1}{2} \log \left(1 + \frac{P}{Q + 1} \right). \quad (24)$$

For the high INR behavior, it suffices to note from (14) that

$$\lim_{Q \rightarrow \infty} C \leq \lim_{Q \rightarrow \infty} R_+ = \frac{1}{4} \log(1 + P), \quad (25)$$

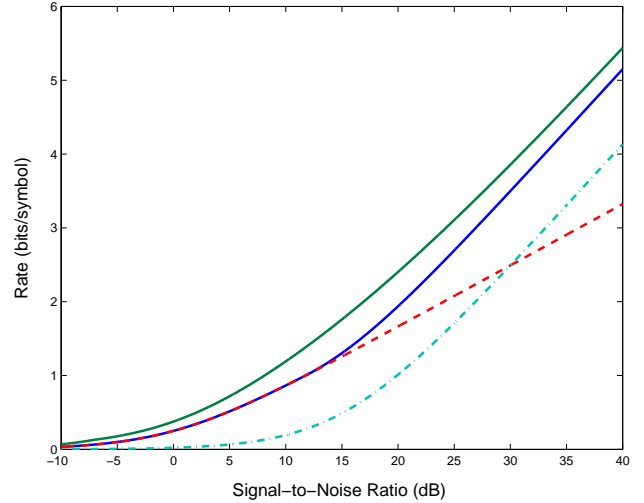


Fig. 6. Upper and lower bounds on the capacity of the two-user Gaussian multicast channel, as a function of SNR P for an INR $Q = 15$ dB. The solid curves depict the two bounds of (13). The dashed curve indicates the performance of time-sharing, while the dash-dotted curve indicates the performance of a strategy in which the side information is treated by the transmitter as additional noise on each link.

which can be achieved by time-sharing between the two users and doing Costa dirty paper coding for each user being served. We note that this result settles the conjecture made in [22].

The behavior of the bounds as a function of SNR is depicted in Fig. 6 for a fixed INR of $Q = 15$ dB. When the SNR is large, Fig. 5 reflects that having the transmitter ignore the side information does not achieve a rate particularly close to capacity, but the superposition dirty paper coding strategy corresponding to our lower bound does. More generally, the difference between the upper bound and the lower bound can be bounded as follows:

$$\begin{aligned} \lim_{P \rightarrow \infty} (C - R_-) &\leq \lim_{P \rightarrow \infty} (R_+ - R_-) \\ &= \begin{cases} \frac{1}{4} & Q \geq 4 \\ \frac{1}{4} \log \left(\frac{2Q}{(1+Q/4)^2} \right) & 2 \leq Q < 4 \\ \frac{1}{2} \log \left(1 + \frac{Q}{4+Q} \right) & Q < 2 \end{cases} \quad (26) \end{aligned}$$

Hence, for INR above 3 dB, our superposition dirty paper code achieves within 1/4 bit/symbol of capacity, while for INR below 3 dB, the superposition part of the code does not help, but the dirty paper code alone achieves within $\frac{1}{2} \log(4/3) = 0.2075$ bit/symbol of capacity.

To verify (26) for $Q \geq 4$, first note that since $P \rightarrow \infty$, the middle case of the lower bound (15) applies which we can alternately express in the form

$$R_- = \frac{1}{4} \log \left(\frac{P + Q/2 + 1}{Q} \right) + \frac{1}{4} \log \left(\frac{P + Q/2 + 1}{2} \right). \quad (27)$$

Combining (27) with the first case of the upper bound (14)

then yields

$$R_+ - R_- = \frac{1}{4} \log \frac{2(1+P)}{P+Q/2+1} + \frac{1}{4} \log \frac{Q(P+Q+1+2\sqrt{PQ})}{Q(P+Q/2+1)}, \quad (28)$$

which in the limit gives (26).

For the case $2 < Q \leq 4$, we combine the middle case in (15) and the second case of (14).

$$R_+ - R_- = \frac{1}{4} \log \frac{2(1+P)}{(1+Q/4)(P+Q/2+1)} + \frac{1}{4} \log \frac{Q(P+Q+1+2\sqrt{PQ})}{(1+Q/4)(P+Q/2+1)}, \quad (29)$$

Taking the limit $P \rightarrow \infty$, we recover the second case of (26). To verify (26) for $Q < 2$, first note that the top case of the lower bound (15) applies, from which we obtain

$$R_+ - R_- = \frac{1}{4} \log \left[\frac{(1+P) \left(\frac{P+Q+1+2\sqrt{PQ}}{(1+Q/4)^2} \right)}{\left(1 + \frac{P}{Q/2+1} \right)^2} \right]. \quad (30)$$

Taking the limit $P \rightarrow \infty$ in (30) yields the relevant case in (26).

We conclude this section with a few additional observations.

Some Further Remarks:

- 1) Feedback does not help much. We show in Appendix III-A that an upper bound when the transmitter has perfect causal feedback from the receiver is given by

$$R_+^F = \frac{1}{4} \log(1+P) + \frac{1}{4} \log \left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1} \right). \quad (31)$$

Thus in the limit $Q \rightarrow \infty$, feedback can gain by no more than 1/4 bit/symbol over time-sharing.

- 2) Correlation among noise sequences does not matter, i.e., the upper bound in Theorem 3 is valid even when the noises Z_1^n and Z_2^n are not independent. The argument is analogous to that for the standard broadcast channel (e.g. [7, Ch. 14]). For the sake of completeness, we prove this fact in Lemma 2 in Appendix III. We further exploit this observation to derive the upper bound (14) in Appendix III.
- 3) The capacity-achieving strategy for the binary channel does not extend immediately to the Gaussian channel. While one might speculate that an adaptation of the achievability approach in Theorem 1 for the Gaussian channel would improve on the lower bound (16a) in Theorem 3, the obvious generalizations do not. In particular, strategies which precancel the interference in part of the codeword for each user achieved strictly lower rates than our superposition dirty paper coding; for a detailed analysis, see [13].

A. Extensions for Robust Dirty Paper Coding

Consider the a memoryless Gaussian point-to-point channel model with output

$$Y^n = X^n + S^n + Z^n, \quad (32)$$

where X^n is the channel input subject to power constraint P , S^n is a white Gaussian interference sequence of power Q not known to decoder, and Z^n is a white Gaussian noise sequence of unit power. When the interference S^n is perfectly known to the encoder, Costa's dirty paper coding is capacity achieving. However, in many applications, only imperfect knowledge of S^n is available to the encoder. One special case is the case of *causal* knowledge considered by Shannon. Another is the case of *noisy* noncausal knowledge. For these kinds of generalizations, there is interest in understanding the capacity of such channels and the structure of the associated capacity-achieving codes, which we refer to as *robust* dirty paper codes.

It is often natural to analyze such problems via their equivalent Gaussian multicast model. As an illustration, suppose that the interference in (32) is of the form $S^n = \beta S_0^n$ where $S_0^n \sim \mathcal{N}(0, \mathbf{QI})$ is known to the encoder but β is not. Then if β is from a finite alphabet (or can be approximated as being so), i.e., $\beta \in \{\beta_1, \beta_2, \dots, \beta_K\}$, the problem is equivalent to a Gaussian multicast problem with K users where the interference for the k th user is $\beta_k S_0^n$.

From this example it is apparent that for at least some applications, there is a need to accommodate *correlated* interferences in the Gaussian multicast model. While a thorough treatment of such generalizations is beyond the scope of the present paper, we note that our results can be used to establish potentially useful lower bounds.

We illustrate such a bound for the case $K = 2$ of the example above. Specifically, consider a superposition dirty paper coding strategy analogous to that in the proof of the lower bound in Theorem 3, whereby we decompose the interferences according to (17). In this case, we have that (18) specializes to

$$\begin{aligned} S^n &= \beta_S S_0^n \\ V^n &= \beta_V S_0^n, \end{aligned} \quad (33)$$

where

$$\begin{aligned} \beta_S &= (\beta_1 + \beta_2)/2 \\ \beta_V &= (\beta_1 - \beta_2)/2. \end{aligned} \quad (34)$$

When we turn to implement step 2) of the encoding in the proof of the lower bound of Theorem 3, in which S^n is treated as interference and V^n as noise, the results of [6] cannot be directly applied since the interferences S^n and V^n in (33) are correlated. However, the stronger version of Costa's results in [8] establishes that provided there is common randomness at the transmitter and receivers, any correlation between interference and noise can be ignored in encoding and does not affect the (worst case) capacity.¹¹

Thus, we obtain the following lower bound on capacity of our example multicast channel with correlated interferences:

$$C_\beta \geq \max_{\{(P_S, P_V): P_S \geq 0, P_V \geq 0, P_S + P_V \leq P\}} R_\beta(P_S, P_V), \quad (35a)$$

¹¹The common randomness enables the channel input to be statistically independent of the effective noise.

where

$$R_\beta(P_S, P_V) = \frac{1}{2} \log \left(1 + \frac{P_S}{1 + (\beta_1 - \beta_2)^2 Q/4 + P_V} \right) + \frac{1}{4} \log(1 + P_V). \quad (35b)$$

While the capacity of this channel remains unknown, we note that the rate (35) is nontrivial. Indeed, it can be quite large even when the INRs $\beta_1 Q$ and $\beta_2 Q$ are each large, provided that $|\beta_1 - \beta_2|$ is small.

V. CONCLUDING REMARKS

We introduced the multicast channel model and analyzed the special cases of binary and Gaussian channels with additive interference. One of the main conclusions of this work is that the lack of side information at the receiver strongly limits capacity. For example, we show that in either case if the interfering sequences are independent, time-sharing is optimal in the limit of large interference.

In terms of future work, it would be interesting to extend the upper bound in Theorem 3 in two directions: (i) when the interference sequences are correlated and (ii) when there are more than two receivers. Both these extensions will lead to further insights into the robustness of dirty paper coding schemes. Another interesting direction would be to develop new schemes that exploit the correlation between interfering sequences in the Gaussian case, extending the work of Section IV-A. In particular, it would be worthwhile to investigate other correlation models and coding schemes that are well-matched to them. Finally, the impact of our results on practical applications remains to be investigated, but is likely to yield important insights for system designers. Examples include the MIMO broadcast channel with imperfect knowledge of fading coefficients [18], as well as the MIMO broadcast channel with common message [13].

APPENDIX I

PROOF OF THE CONVERSE IN THEOREM 1

We have to show that for any sequence of $(2^{nR}, n)$ codes with $P_e^n \rightarrow 0$, we must have $R \leq C$, where C is defined in (4).

Since each receiver is able to decode the message we have from Fano's inequality

$$H(W|Y_k^N) \leq n\epsilon_n, \quad \text{for } k = 1, 2. \quad (36)$$

We can use Fano's inequality to bound the rate as

$$\begin{aligned} nR &= H(W) \\ &= H(W|Y_1^n) + I(W; Y_1^n) \\ &\leq n\epsilon_n + H(Y_1^n) - H(Y_1^n|W) \end{aligned} \quad (37)$$

$$\leq n\epsilon_n + \sum_{j=1}^n H(Y_{1j}) - H(Y_1^n|W) \quad (38)$$

$$\leq n\epsilon_n + n - H(Y_1^n|W), \quad (39)$$

where (37) follows by using the Fano inequality (36), (38) follows from the chain rule and the fact that conditioning

reduces the entropy, and (39) follows from the fact that each Y_{1j} is binary valued. We can similarly bound the rate on the second user's channel as

$$nR \leq n\epsilon_n + n - H(Y_2^n|W). \quad (40)$$

Combining (39) and (40), we obtain

$$\begin{aligned} nR &\leq n - \max\{H(Y_1^n|W), H(Y_2^n|W)\} + n\epsilon_n \\ &\leq n - \frac{1}{2}\{H(Y_1^n|W) + H(Y_2^n|W)\} + n\epsilon_n \\ &\leq n - \frac{1}{2}H(Y_1^n, Y_2^n|W) + n\epsilon_n \end{aligned} \quad (41)$$

$$\leq n - \frac{1}{2}H(Y_1^n \oplus Y_2^n|W) + n\epsilon_n \quad (42)$$

$$= n - \frac{1}{2}H(S_1^n \oplus S_2^n) + n\epsilon_n \quad (43)$$

$$= n \left(1 - \frac{1}{2}H(S_1 \oplus S_2) + \epsilon_n \right), \quad (44)$$

where (41) follows from the fact that conditioning reduces entropy, (42) follows from the fact that $Y_1^n \oplus Y_2^n$ is a deterministic function of (Y_1^n, Y_2^n) , (43) follows from the fact that $Y_1 \oplus Y_2 = S_1 \oplus S_2$, and (44) follows from the fact that both S_1 and S_2 are i.i.d. so the joint entropy of the sequence $S_1^n \oplus S_2^n$ is the sum of the individual terms.

APPENDIX II

PROOF OF UPPER BOUND (9b) IN THEOREM 2

The upper bound mirrors the converse for two-user case. In particular, following the same steps as in the two-user case to derive (41), we have that any achievable rate satisfies

$$nR \leq n - \frac{1}{K}H(Y_1^n, Y_2^n, \dots, Y_K^n|W) + n\epsilon_n. \quad (45)$$

Proceeding from (45) we obtain

$$\begin{aligned} nR - n\epsilon_n &\leq n - \frac{1}{K}H(Y_1^n, Y_2^n, \dots, Y_K^n|W) \\ &= n - \frac{1}{K}H(Y_1^n, Y_1^n \oplus Y_2^n, \dots, Y_1^n \oplus Y_2^n|W) \end{aligned} \quad (46)$$

$$= n - \frac{1}{K}H(X^n \oplus S_1^n, S_1^n \oplus S_2^n, \dots, S_1^n \oplus S_K^n|W)$$

$$= n - \frac{1}{K}H(S_1^n \oplus S_2^n, \dots, S_1^n \oplus S_K^n|W)$$

$$- \frac{1}{K}H(X^n \oplus S_1^n|S_1^n \oplus S_2^n, \dots, S_1^n \oplus S_K^n, W)$$

$$= n - \frac{n}{K}H(S_1 \oplus S_2, \dots, S_1 \oplus S_K)$$

$$- \frac{1}{K}H(X^n \oplus S_1^n|S_1^n \oplus S_2^n, \dots, S_1^n \oplus S_K^n, W) \quad (47)$$

$$\leq n - \frac{n}{K}H(S_1 \oplus S_2, \dots, S_1 \oplus S_K),$$

where (46) follows from the fact that the mapping $(Y_1^n, Y_2^n, \dots, Y_K^n) \rightarrow (Y_1^n, Y_1^n \oplus Y_2^n, \dots, Y_1^n \oplus Y_2^n)$ is invertible, and (47) follows from the fact that $S_1^n, S_2^n, \dots, S_K^n$ are all i.i.d. and independent of W .

APPENDIX III

PROOF OF UPPER BOUND (14) IN THEOREM 3

Our proof is structured as follows. We derive an upper bound for a particular single-interference Gaussian channel, and reason that the achievable rate for the two-interference channel of interest in Theorem 3 cannot be higher.

The single-interference channel is one in which $S_1^n = 0$ and $S_2^n = S^n$ in (11) and Fig. 4, i.e., only the second receiver experiences interference. We now argue that provided the encoder and decoder 1 share common randomness in this channel, they can simulate the two-interference channel. In particular, receiver 1 can generate an i.i.d. $\mathcal{N}(0, Q)$ sequence \hat{S}_1^n and add to Y_1^n . Because of the shared randomness, the encoder also knows this sequence. Now if we restrict user 1 to ignore the knowledge of \hat{S}_1^n while decoding, the scenario is exactly the same as in the two-interference case in Fig. 4. Thus any rate achievable for the two-interference channel is also achievable for the single-interference channel with common randomness, so we focus in the sequel on the latter.

In the remainder of this section, we will first develop an upper bound on the single-interference channel and then show that the upper bound continues to hold even when there is common randomness available.

In our single-interference channel, S^n is a white Gaussian interference of power Q . In addition, the noises Z_1 and Z_2 on the two channels are correlated, zero mean, unit variance, jointly Gaussian random variables with correlation coefficient $E[Z_1 Z_2] = \rho$. We note that the capacity of the channel only depends on the marginal distributions $p(Y_1^n | X^n)$ and $p(Y_2^n | X^n, S^n)$ and not on the joint distribution $p(Y_1^n, Y_2^n | X^n, S^n)$. Thus, allowing such correlation between the noise does not change capacity. Specifically, we have

Lemma 2: Let P_e^n be the probability of decoding error in (2). If P_e^n is bounded away from zero for a certain correlation between Z_1 and Z_2 , above then it is bounded away from zero for *any* other correlation between Z_1 and Z_2 .

Proof: The argument is essentially the same as given in [7, Ch 14, Page 454]. We repeat it here for completeness. Let $P_e^{1,n}$ and $P_e^{2,n}$ denote the error probabilities in decoding at receiver 1 and 2 respectively. We have

$$P_e^{1,n} = \Pr(g_1(Y_1^n) \neq W)$$

$$P_e^{2,n} = \Pr(g_2(Y_2^n) \neq W)$$

$$P_e^n = \Pr\left(\bigcup_{k=1,2} \{g_k(Y_k^n) \neq W\}\right)$$

Next, note that

$$\max\{P_e^{1,n}, P_e^{2,n}\} \leq P_e^n \leq P_e^{1,n} + P_e^{2,n}, \quad (48)$$

where the left inequality in (48) follows from the fact that by definition $P_e^n \geq P_e^{k,n}$ for $k = 1, 2$, and the right inequality follows from the union bound. In turn, note that both $P_e^{1,n}$ and $P_e^{2,n}$ do not depend on the correlation between Z_1 and Z_2 . Accordingly, both the left and right hand terms in (48) do not depend on the correlation between Z_1 and Z_2 . In particular if P_e^n is bounded away from 0 for some correlation between Z_1 and Z_2 , then necessarily one of $P_e^{1,n}$ and $P_e^{2,n}$ is bounded

away from zero. Thus the probability of error is bounded away from zero for all possible correlations. ■

In the rest of the section we will fix $E[Z_1 Z_2] = \rho$ and derive an upper bound. Thereafter, we will optimize over ρ , to tighten the upper bound. We will need the following additional properties of Z_1 and Z_2 , which are readily computed.

Lemma 3: Let Z_1 and Z_2 be standard normal, jointly Gaussian random variables with correlation ρ . Define $Z_- \triangleq (Z_1 - Z_2)/\sqrt{2}$ and $Z_+ \triangleq (Z_1 + Z_2)/\sqrt{2}$. Then Z_+ and Z_- are independent zero-mean Gaussian random variables with variances $1 + \rho$ and $1 - \rho$, respectively.

To obtain our upper bound we relax our power constraint¹² to $E[X_i^2] < P_i$ for $i = 1, 2, \dots, n$ with $\frac{1}{n} \sum_{i=1}^n P_i \leq P$ and show that a sequence of $(2^{nR}, n)$ codes that can be decoded by both the receivers with $P_e^n \rightarrow 0$ must satisfy $R \leq R_+$ in (14).

Suppose R_1 and R_2 denote the rates at which the two receivers can reliably decode the common message. The rate of the common message must satisfy $R \leq \min(R_1, R_2)$.

From Fano's inequality, we have

$$H(W|Y_k^N) \leq n\epsilon_n, \quad \text{for } k = 1, 2. \quad (49)$$

We first upper bound R_1 as

$$\begin{aligned} nR_1 &< I(W; Y_1^n) + n\epsilon_n \\ &= h(Y_1^n) - h(Y_1^n|W) + n\epsilon_n \\ &\leq \sum_{i=1}^n h(Y_i) - h(Y_1^n|W) + n\epsilon_n \end{aligned} \quad (50)$$

$$\leq \sum_{i=1}^n \log 2\pi e(P_i + 1) - h(Y_1^n|W) + n\epsilon_n. \quad (51)$$

$$\leq \frac{n}{2} \log 2\pi e(P + 1) - h(Y_1^n|W) + n\epsilon_n, \quad (52)$$

where (50) follows from the chain rule and the fact that conditioning reduces entropy, and (51) follows from the fact that each Y_i has a variance no larger than $P_i + 1$ and its differential entropy can be upper bounded by that of a Gaussian RV. Finally, (52) is a consequence of Jensen's inequality. Rearranging the terms in (52) we obtain

$$h(Y_1^n|W) < \frac{n}{2} \log 2\pi e(P + 1) - nR_1. \quad (53)$$

We will also need the following corresponding lower bound on $h(Y_2^n|W)$.

Lemma 4:

$$\begin{aligned} h(Y_2^n|W) &\geq \frac{n}{2} \log 2\pi e \left(\frac{Q}{2} + 1 - \rho \right) + \frac{n}{2} \log 2\pi e(1 + \rho) - h(Y_1^n|W). \end{aligned} \quad (54)$$

¹²Rather than requiring that *for each codeword* the time-averaged power not exceed P , this constraint only requires that this time-average power not exceed P *when averaged over the codewords*.

Proof: First note that:

$$\begin{aligned} & h(Y_2^n|W) \\ &= h(Y_2^n, Y_1^n|W) - h(Y_1^n|W, Y_2^n) \\ &= h\left(\frac{Y_1^n - Y_2^n}{\sqrt{2}}, \frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W\right) - h(Y_1^n|W, Y_2^n) \quad (55) \end{aligned}$$

$$\begin{aligned} &= h\left(\frac{Y_1^n - Y_2^n}{\sqrt{2}} \middle| W\right) \\ &+ h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\right) - h(Y_1^n|W, Y_2^n) \quad (56) \end{aligned}$$

$$\begin{aligned} &= h\left(\frac{Z_1^n - Z_2^n}{\sqrt{2}} - \frac{S^n}{\sqrt{2}}\right) \\ &+ h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\right) - h(Y_1^n|W, Y_2^n) \quad (57) \end{aligned}$$

$$\begin{aligned} &= \frac{n}{2} \log 2\pi e \left(\frac{Q}{2} + 1 - \rho\right) \\ &+ h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\right) - h(Y_1^n|W, Y_2^n) \quad (58) \end{aligned}$$

$$\begin{aligned} &\geq \frac{n}{2} \log 2\pi e \left(\frac{Q}{2} + 1 - \rho\right) \\ &+ h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\right) - h(Y_1^n|W), \quad (59) \end{aligned}$$

where (55) follows from the fact that the differential entropy is invariant under an orthogonal transformation, (57) follows from substituting Y_1^n and Y_2^n from (11), (58) follows from the independence of S^n and (Z_1^n, Z_2^n) and Lemma 3, and (59) follows from the fact removing Y_2^n from the conditioning in the last term increases its entropy.

We next lower bound the second term in (59) according to

$$\begin{aligned} & h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}} \middle| W, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\right) \\ &= h\left(\sqrt{2}X^n + \frac{1}{\sqrt{2}}S^n + Z_+^n \middle| W, Z_-^n - \frac{1}{\sqrt{2}}S^n\right) \quad (60) \end{aligned}$$

$$\begin{aligned} &\geq h\left(\sqrt{2}X^n + \frac{1}{\sqrt{2}}S^n + Z_+^n \middle| W, Z_-^n - \frac{1}{\sqrt{2}}S^n, X^n, S^n\right) \quad (61) \end{aligned}$$

$$= h(Z_+^n|Z_-^n) \quad (62)$$

$$= h(Z_+^n) = \frac{n}{2} \log 2\pi e(1 + \rho), \quad (63)$$

where (60) follows from the fact that substituting Y_1^n and Y_2^n from (11), (61) from the fact that additional conditioning on X^n and S^n only reduces the differential entropy, (62) from the fact that in our channel model Z_+^n is independent of (W, X^n, S^n) and (63) from the fact that using Lemma 3 we have that Z_+^n and Z_-^n are independent and $Z_+ \sim \mathcal{N}(0, 1 + \rho)$.

Substituting this lower bound into (59), we obtain (54). ■

Finally, using the standard Fano arguments we have the

following chain of inequalities

$$\begin{aligned} & nR_2 - n\epsilon'_n \\ &\leq I(W; Y_2^n) \\ &= h(Y_2^n) - h(Y_2^n|W) \\ &\leq \sum_{i=1}^n h(Y_{2i}) - h(Y_2^n|W) \quad (64) \end{aligned}$$

$$\leq \sum_{i=1}^n \log 2\pi e(P_i + Q + 1 + 2\sqrt{P_i Q}) - h(Y_2^n|W) \quad (65)$$

$$\leq \frac{n}{2} \log 2\pi e(P + Q + 1 + 2\sqrt{PQ}) - h(Y_2^n|W) \quad (66)$$

$$\begin{aligned} &\leq \frac{n}{2} \log 2\pi e(P + Q + 1 + 2\sqrt{PQ}) - \frac{n}{2} \log 2\pi e\left(\frac{Q}{2} + 1 - \rho\right) \\ &\quad - \frac{n}{2} \log 2\pi e(1 + \rho) + h(Y_1^n|W) \quad (67) \end{aligned}$$

$$\begin{aligned} &\leq \frac{n}{2} \log 2\pi e(P + Q + 1 + 2\sqrt{PQ}) - \frac{n}{2} \log 2\pi e\left(\frac{Q}{2} + 1 - \rho\right) \\ &\quad - \frac{n}{2} \log 2\pi e(1 + \rho) + \frac{n}{2} \log 2\pi e(P + 1) - nR_1 \quad (68) \end{aligned}$$

$$\begin{aligned} &= \frac{n}{2} \log\left(\frac{1+P}{1+\rho}\right) + \frac{n}{2} \log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1-\rho}\right) - nR_1, \quad (69) \end{aligned}$$

where (64) follows from the chain rule and the fact that conditioning reduces entropy. Since $Y_{2i} = X_i + S_i + Z_{2i}$, its variance can be at most $P_i + Q + 1 + 2\sqrt{P_i Q}$, which follows from the Cauchy-Schwarz inequality and the fact that Z_{2i} is independent of (X_i, S_i) . Accordingly, we can upper bound $h(Y_{2i})$ by the differential entropy of a Gaussian random variable as in (65) and use Jensen's inequality in (66). Eq. (67) follows by substituting the lower bound for the differential entropy term from Lemma 4, while (68) follows by substituting the upper bound for $h(Y_1^n|W)$ in (53).

Rearranging the terms in (69) we have

$$\begin{aligned} nR_1 + nR_2 - n\epsilon_n &\leq \frac{n}{2} \log\left(\frac{1+P}{1+\rho}\right) \\ &\quad + \frac{n}{2} \log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1-\rho}\right), \\ &\quad 0 \leq \rho \leq 1. \quad (70) \end{aligned}$$

To obtain the tightest bound we minimize the right hand side of the above over ρ . The tightest bounds is obtained with the choice

$$\rho^*(Q) = \begin{cases} Q/4 & \text{if } Q \leq 4 \\ 1 & \text{if } Q > 4. \end{cases} \quad (71)$$

Finally note that the common information rate R satisfies

$$R \leq \min(R_1, R_2) \leq \frac{R_1 + R_2}{2}. \quad (72)$$

Combining (70), (71) and (72) we obtain (14) upon letting $n \rightarrow \infty$.

A. Gains from Feedback

Suppose that the transmitter has access to perfect causal feedback from the receiver. Accordingly we can express the

$$h\left(\sqrt{2}X^n + \frac{1}{\sqrt{2}}S^n + Z_+^n \middle| W, Z_-^n - \frac{1}{\sqrt{2}}S^n\right) \quad (73)$$

$$= \sum_i h\left(\sqrt{2}X_i + \frac{1}{\sqrt{2}}S_i + Z_{+,i} \middle| W, Z_-^n - \frac{1}{\sqrt{2}}S^n, \sqrt{2}X_1^{i-1} + \frac{1}{\sqrt{2}}S_1^{i-1} + Z_{+,1}^{i-1}\right) \quad (74)$$

$$\geq \sum_i h\left(\sqrt{2}X_i + \frac{1}{\sqrt{2}}S_i + Z_{+,i} \middle| W, Z_-^n - \frac{1}{\sqrt{2}}S^n, \sqrt{2}X_1^{i-1} + \frac{1}{\sqrt{2}}S_1^{i-1} + Z_{+,1}^{i-1}, S_i, X_i\right) = \sum_i h(Z_{+,i}) \quad (75)$$

transmitted symbol as $x_i = f(w, y_1^{i-1}, y_2^{i-1}, s^n)$. We argue that if $E[Z_1 Z_2] = \rho$, then the upper bound (70) continues to hold even with feedback.

Indeed, a close examination of the proof above shows that the dependence between X^n and (Z_1^n, Z_2^n) can only be potentially problematic while deriving the lower bound in (60)-(63). In particular since Z_+^n is dependent on X^n , we cannot simplify (61) to (62). However we can decompose (61) via chain rule as shown in (73)-(75) on the top of the page. We then use the fact that $Z_{+,i}$ is independent of (W, Z_-^n, S^n, X_1^i) to get the lower bound in (63).

Note however that the optimization over ρ cannot be done when we allow feedback. In particular Lemma 2 does not continue to hold if we allow for feedback. This is because $P_e^{1,n}$ and $P_e^{2,n}$ in (48) are no longer only dependent on the marginals. Hence, when Z_1 and Z_2 are independent, we select $\rho = 0$. The corresponding upper bound is given by

$$R^F \leq \frac{1}{4} \log(1+P) + \frac{1}{4} \log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1}\right). \quad (76)$$

B. Common randomness does not help

To complete the argument, it remains to show that common randomness does not increase the capacity of this single-interference channel. We will provide an argument for the general multicast channel in Definition 1. The result for the Gaussian case of interest then follows as a special case.

Suppose that the transmitter and receiver operations depend on a common random variable D , whose realization is known at both the transmission ends. Given the message m and state sequence s^n , the encoding function (c.f. Definition 1) is given by $x^n = f(m, s^n, d)$. Similarly the decoding functions are given by $\hat{m}_k = g_k(y_k^n, d)$ for $k = 1, 2, \dots, K$. The average probability of error for the rate R randomized code is then defined by

$$\begin{aligned} & P_e^{n, \text{randomized}} \\ &= \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} E_D \left[\sum_{y^n: \exists k: g_k(y_k^n, d) \neq m} \sum_{s^n} p(s^n) p(y^n | f(m, s^n, d)) \right] \\ &= E_D \left[\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \sum_{y^n: \exists k: g_k(y_k^n, d) \neq m} \sum_{s^n} p(s^n) p(y^n | f(m, s^n, d)) \right] \\ &= E_D \left[\Pr \left\{ \bigcup_{k=1}^K \{g(Y_k^n, d) \neq W\} \middle| D = d \right\} \right], \end{aligned}$$

where the second equality follows by interchanging the expectation and summation over m , and the third equality follows by observing that given a realization of the random variable D , the encoding and decoding are both deterministic and we can use the definition of the average probability of error in (2). Finally note that there must be some value of d for which the term inside the expectation is minimized. We can design the encoding and decoding function for this deterministic value of d and our probability of error will be lower than the average. Thus having access to common randomness cannot decrease the probability of error for the K-user multicast channel. An analogous argument can be made for the continuous alphabet extension of the multicast channel. Since the single-interference Gaussian channel is a special case of this latter model, common randomness cannot lower the probability of error and the upper bound derived in (14) still holds.

REFERENCES

- [1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, 2003.
- [2] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoder," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3992–4003, Nov., 2005.
- [3] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. IT-47, pp. 1423–1443, May 2001.
- [4] A. S. Cohen and A. Lapidot, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1639–1667, June 2002.
- [5] M. H. M. Costa, "Private communication."
- [6] —, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.
- [8] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.
- [9] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: Wiley, 1968.
- [10] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Peredachi Inform. (Probl. Inform. Trans.)*, vol. 9, no. 1, pp. 19–31, 1980.
- [11] —, "On Gaussian channels with random parameters," in *Proc. Int. Symp. Inform. Theory*, Sep. 1984, pp. 247–250.
- [12] C. Heegard and A. E. Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inform. Theory*, vol. 29, pp. 731–739, Sep. 1983.
- [13] A. Khisti, "Coding techniques for multicasting," Master's thesis, M.I.T, Cambridge, MA, 2004, <http://web.mit.edu/khisti/www/SMThesis.pdf>.
- [14] A. Khisti, U. Erez, and G. Wornell, "Writing on two pieces of dirty paper at once," in *Proc. Int. Symp. Inform. Theory*, June 2004.
- [15] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proc. Int. Symp. Inform. Theory*, June 2004.

- [16] S. Kotagiri and J. N. Laneman, "Achievable rates for multiple access channels with state information known at one encoder," in *Allerton Conf. Commun., Contr., Computing*, Monticello, IL, Oct. 2004.
- [17] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Peredachi Inform. (Probl. Inform. Trans.)*, vol. 10, pp. 52–60, Apr.-June 1974.
- [18] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Annual Allerton Conference on Communication, Control, and Computing*, September, 2005.
- [19] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, pp. 563–593, Mar. 2003.
- [20] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Dev.*, vol. 2, pp. 289–293, Oct. 1958.
- [21] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.
- [22] Y. Steinberg and S. Shamai (Shitz), "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. Int. Symp. Inform. Theory*, 2005.
- [23] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," in *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, 2004.
- [24] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1964.