Correspondence

Carbon Copying Onto Dirty Paper

Ashish Khisti, *Student Member, IEEE*, Uri Erez, *Member, IEEE*, Amos Lapidoth, *Fellow, IEEE*, and 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 (in addition to the additive noise), which is known noncausally to the transmitter but not to any of the receivers. Applications range from wireless multiple-antenna multicasting to robust dirty paper coding.**

We develop results for memoryless channels in Gaussian and binary special cases. In most cases, we observe 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 establish the capacity when there are two receivers. When there are many receivers, we show that the transmitter side information provides a vanishingly small benefit. When the interference is large and independent across the users, we show that time sharing is optimal. For the Gaussian case, we present a coding scheme and establish its optimality in the high signal-to-interference-plus-noise limit when there are two receivers. When the interference power is large and independent across all the receivers, we show that time-sharing is again optimal. Connections to the problem of robust dirty paper coding are also discussed.

***Index Terms*—Common information, dirty paper coding, Gel'fand–Pinsker channels, multiple-input/multiple-output (MIMO) broadcast channel, writing on dirty paper.**

I. Introduction

The study of communication over channels controlled by a random state parameter known only to the transmitter was initiated by Shannon [21]. 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], [18], to digital watermarking [3], [4], [20], and to coding for the multiple-input/multiple-output (MIMO) broadcast channel [1], [25].

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. Some other multiuser settings are also discussed. The degraded broadcast channel with independent messages and state sequence known to the transmitter either causally or noncausally is examined in [23]. Other works on multiuser channels with state parameters include [17], [2], [16], [13], and [22].

This correspondence 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 addition to additive noise, 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 in a variety of multiple-antenna wireless multicasting problems as well as in applications of robust 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], [15]. For more general channels, [24] 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 and identically distributed (i.i.d.) Gaussian interferences on each link is also considered in [24], 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 correspondence 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 is asymptotically optimal in the limit of high signal-to-interference-plus-noise (SINR) ratio.[1]

An outline of the correspondence 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. The proofs of the converses are deferred to the Appendices.

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 sequence $s^n = (s_1, s_2, \dots, s_n)$ 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

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

Manuscript received November 28, 2005; revised September 29, 2006. This work was supported in part by the National Science Foundation under Grant CCF-0515109, and by Hewlett-Packard through the MIT/HP Alliance. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Chicago, IL, June/July 2004 and at the International Zurich Seminar, Zurich, Switzerland, February 2006.

A. Khisti and G. W. Wornell are with the Department Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: khisti@mit.edu; 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 Signal and Information Processing Laboratory, ETH Zurich, 8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

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

Digital Object Identifier 10.1109/TIT.2007.894693

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

where $y_k^n = (y_{k1}, y_{k2}, \ldots, y_{kn})$, for all $y_{ki} \in \mathcal{Y}_k$, $k = 1, 2, \ldots, K$. 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, $s^n = (s_1^n, s_2^n, \ldots, s_K^n)$, so that $p(s_i) = p(s_{1i}, s_{2i}, \ldots, s_{Ki})$.

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

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

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

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 correspondence, 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. While we believe that our techniques also apply to the general channel model in Definition 1, we will not consider it further.

III. Noiseless Binary Case

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

$$Y_k^n = X^n \oplus S_k^n \tag{3}$$

where $X_i, S_{ki} \in \{0, 1\}$, and $\oplus$ denotes symbol-by-symbol modulo-two addition (i.e., EXCLUSIVE-OR). In (3), the memoryless case of interest corresponds to the requirement that the $(S_{1i}, S_{2i}, \ldots, S_{Ki})$ for $i = 1, 2, \ldots, n$ form an i.i.d. sequence of K-tuples. In particular, for each i, the variables $\{S_{1i}, S_{2i}, \ldots, S_{Ki}\}$ may in general be statistically dependent, 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 $(S_1, S_2, \ldots, S_K)$.

Note that with only a single receiver ($K = 1$), the capacity is trivially 1 (bit per channel use),[2] 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 $R_{\mathrm{TS}} = 1/2$. 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 $R_{\mathrm{IS}} = 1 - \max\{H(S_1), H(S_2)\}$. It turns out that the former bound is only tight when S_1 and S_2 are independent and $\mathcal{B}(1/2)$, and the latter bound is only tight when both S_1 and S_2 are $\mathcal{B}(0)$.[3]

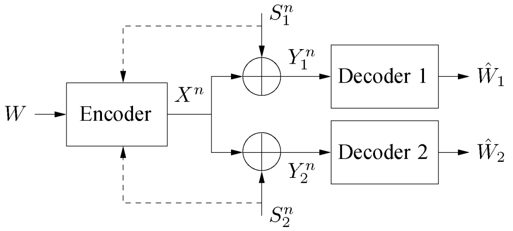

Fig. 1. Two-user memoryless, noiseless, binary multicast channel with additive interference. The encoder maps message W into codeword X^n. The state takes the form of interference sequences S_1^n and S_2^n. Each channel output $Y_k^n = X^n \oplus S_k^n$, where $\oplus$ denotes symbol-by-symbol modulo-two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

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

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

Proof: A converse is provided in Appendix I. The achievability argument is detailed as follows.

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 $B^n(1), B^n(2), \ldots, B^n(2^{nR})$, 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 [26]). 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 $\mathcal{A}_2$ (with $A_i = 0$) for user 2. Specifically, the transmitted sequence is of the form

$$X_i(w) = \begin{cases} B_i(w) \oplus S_{1i} & i \in \mathcal{A}_1 \\ B_i(w) \oplus S_{2i} & i \in \mathcal{A}_2. \end{cases} \tag{5}$$

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 $|\mathcal{A}_1|/n \to 1/2$

$$\frac{1}{n}I(B^n; Y_k^n | A^n) \to \frac{1}{2} + \frac{1}{2}(1 - H(S_1 \oplus S_2)), \qquad k = 1, 2 \tag{6}$$

which is C in (4). As the mutual information expression in (6) indicates, the decoding of Y_k^n to the message $\hat{W}_k$ is done by using the knowledge of $\mathcal{A}_1$ and $\mathcal{A}_2$ (i.e., A^n) at the decoders. In particular, receiver 1 selects a codeword which agrees with the received symbols in the set $\mathcal{A}_1$ and which is typical with noise $S_1 \oplus S_2$ with the symbols in the set $\mathcal{A}_2$. For decoder 2, the order of the sets is reversed. As long as $R \leq C$, $\hat{W}_k$ equals W with high probability. □

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

[2] From now on, except in the case of ambiguity, the units of "bits per channel use" will be omitted.

[3] We use $\mathcal{B}(q)$ to denote a Bernoulli random variable with parameter q i.e., $\Pr(S = 1) = q, \Pr(S = 0) = 1 - q$.

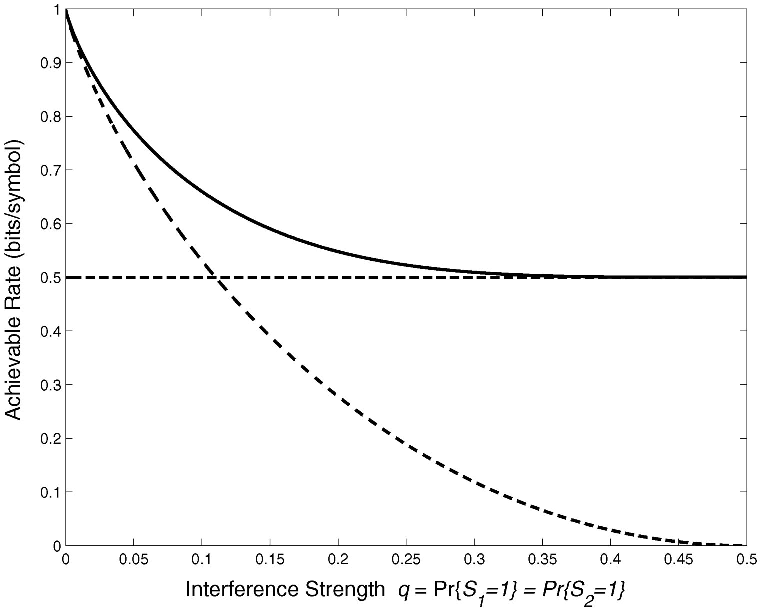

Fig. 2. Achievable rates for the two-user noiseless binary multicast channel with i.i.d. interferences, as a function of the strength of the interference. Capacity is indicated by the solid curve, time-sharing performance is indicated by the horizontal dashed line, and the performance of a system that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

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 $S_2 = \overline{S_1}$. Second, time sharing is strictly suboptimal except when S_1 and S_2 are independent $\mathcal{B}(1/2)$ random variables. 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 $H(S_1) = H(S_2) = 0$.

We make a few additional observations.

Some Further Remarks:

1) The achievability argument can also be obtained via a more direct, but perhaps less intuitive route as follows. First note that a straightforward extension of the random binning argument for the single user case [10] shows that the following rate is achievable for the K-user multicast channel with random parameters

$$R_K = \max_{p(U|S),p(X|U,S)} \{\min_k I(U;Y_k) - I(U;S)\}. \quad (7)$$

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

For the two-user binary channel, the following choice of U yields the achievability of (4). Let the alphabet of U be $\mathcal{U} = \{\Psi_1,\Psi_2,\Psi_3,\Psi_4\}$.

$$U = A\left\{\Psi_1\left(X \oplus S_1\right) + \Psi_2\left(\overline{X \oplus S_1}\right)\right\} + \bar{A}\left\{\Psi_3\left(X \oplus S_2\right) + \Psi_4\,\overline{\left(X \oplus S_2\right)}\right\} \quad (8)$$

where X is $\mathcal{B}(1/2)$ random variable, independent of S_1 and S_2, and A is also $\mathcal{B}(1/2)$ that is independent of X, S_1, and S_2, and where $\bar{\cdot}$ denotes the complement of a (binary-valued) variable.

2) For the code construction outlined above the transmitter does not require noncausal knowledge of the interference. We emphasize, however, this result is specific to the noiseless binary channel model.

3) 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].

4) 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 $S_1^n \oplus S_2^n$.

5) Our bounding technique can also be extended in the presence of noise. For the channel model

$$Y_1 = X \oplus S_1 \oplus Z_1$$
$$Y_2 = X \oplus S_2 \oplus Z_2$$

where Z_1 and Z_2 are mutually independent and identically distributed Bernoulli random variables and independent of all other variables, we can 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 and an upper bound is given by

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

Note that time sharing is again optimal in the special case when S_1 and S_2 are independent $\mathcal{B}(1/2)$ random variables.

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.

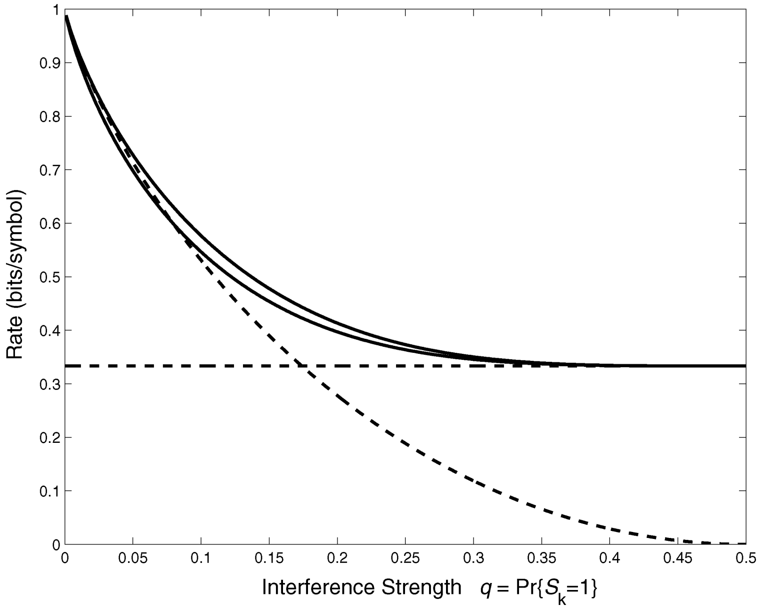

Fig. 3. Upper bound and lower bounds on the capacity of the three-user noiseless binary multicast channel, as a function of the strength of the interference. The solid curves depict the two bounds of (9). The horizontal dashed line indicates the performance of time sharing, while the other dashed curve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

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 distributed[4] is bounded according to

$$R_- \le C \le R_+ \tag{9a}$$

where

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

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

Proof: The upper bound (9b) is established in Appendix II. The lower bound (9c) is obtained via a direct generalization of the code construction (5) 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.[5] However, the associated rate gap is often small in practice. Fig. 3 illustrates the gap for the case $K = 3$.

The rate gap also decays to zero in the limit of large K, which follows readily from Theorem 2. In particular, $C \to 1 - H(S)$ as $K \to \infty$, 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.

$$\left(1 - \frac{1}{K}\right) H(S) \le \frac{1}{K} H(S_1 \oplus S_2, S_1 \oplus S_3, \dots, S_1 \oplus S_K) \le H(S). \tag{10}$$

As $K \to \infty$, 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 extension of Definition 1 and incorporate an average power constraint on the input. Unless otherwise stated, we restrict 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. We first focus on the case of independent interferences and consider the case of correlated interferences in Section IV-B. 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

$$Y_k^n = X^n + S_k^n + Z_k^n, \qquad k = 1, 2. \tag{11}$$

Our power constraint takes the form

$$\frac{1}{n} E\left[\sum_{i=1}^{n} X_i^2(W, S_1^n, S_2^n)\right] \le P \tag{12}$$

where the expectation is taken over the ensemble of messages and interference sequences. Finally, note that without loss of generality, we

[4] The upper bound does not require this assumption. The assumption is merely used to simplify the expression for the lower bound.

[5] A slightly improved lower bound appears in [14], but it, too, does not match the upper bound.

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 present the following bounds on the capacity.

Theorem 3: An upper bound on the Gaussian multicast channel capacity is

$$C \leq \min\{R_+^{\mathrm{I}}, R_+^{\mathrm{II}}\} \tag{13}$$

where [6],[7]

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

$$R_+^{\mathrm{II}} = \begin{cases} \frac{1}{2}\log\left(\frac{1+P+Q+2\sqrt{PQ}}{1+Q/2}\right), & Q \leq 2 \\ \frac{1}{2}\log\left(\frac{1+P+Q+2\sqrt{PQ}}{\sqrt{2Q}}\right) - \left[\frac{1}{4}\log\left(\frac{Q}{2P+2}\right)\right]^+, & Q > 2. \end{cases} \tag{15}$$

We have presented two different upper bounds denoted by R_+^{II} and R_+^{I} since neither bound dominates the other, over all values of (P, Q). The two bounds have been derived by slightly different methods. The bound R_+^{I} is obtained by observing that the channel is nontrivial even if we set one of the interferences (say S_1) to 0. Furthermore, it is possible to show that an upper bound on this modified channel is also an upper bound on the Gaussian multicast channel of interest. A complete derivation of this upper bound is presented in Appendix IV. The expression for R_+^{II} is obtained by directly applying a chain of inequalities on the Gaussian multicast channel and its derivation is presented in Appendix III.

We remark here that the upper bounds are explicit expressions of the following maximization:

$$R_+^{\mathrm{I}} = \min_{\rho\in[-1,1]} \frac{1}{4}\log\left(\frac{1+P}{1+\rho}\right) + \frac{1}{4}\log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1-\rho}\right) \tag{16}$$

$$\begin{aligned} R_+^{\mathrm{II}} = \min_{\rho\in[-1,1]} & \frac{1}{2}\log\left(\frac{P+Q+2\sqrt{PQ}+1}{\sqrt{(1+\rho)(Q+1-\rho)}}\right) \\ & - \left[\frac{1}{4}\log\left(\frac{Q}{2P+(1+\rho)}\right)\right]^+. \end{aligned} \tag{17}$$

Theorem 4: A lower bound on the Gaussian multicast channel capacity is

$$R_- = \begin{cases} \frac{1}{2}\log\left(1 + \frac{P}{Q/2+1}\right), & Q/2 < 1 \\ \frac{1}{2}\log\left(\frac{P+Q/2+1}{Q}\right) + \frac{1}{4}\log\left(\frac{Q}{2}\right), & 1 \leq Q/2 < P+1 \\ \frac{1}{4}\log(1+P), & Q/2 \geq P+1. \end{cases} \tag{18}$$

Proof: The lower bound[8] (18) is an explicit expression of the following maximization:

$$R_- = \max_{\{(P_A,P_D):P_A\geq 0, P_D \geq 0, P_A+P_D\leq P\}} R(P_A, P_D) \tag{19a}$$

with

$$R(P_A, P_D) \triangleq \frac{1}{2}\log\left(1+\frac{P_A}{P_D+Q/2+1}\right) + \frac{1}{4}\log\left(1+P_D\right). \tag{19b}$$

Accordingly, we show the achievability of (19b). The proposed scheme combines superposition coding, dirty paper coding, and time sharing, and exploits a representation of the interferences in the form

$$\begin{aligned} S_1^n &= A^n + D^n \\ S_2^n &= A^n - D^n \end{aligned} \tag{20}$$

where

$$\begin{aligned} A^n &= (S_1^n + S_2^n)/2 \\ D^n &= (S_1^n - S_2^n)/2. \end{aligned} \tag{21}$$

We list the main steps for codebook generation, encoding, and decoding. The probability of error analysis will be omitted as it is based on standard typicality arguments. See, e.g., [7].

Codebook Generation: The idea is to generate three codebooks. There is one common codebook which both the users share and two private codebooks which are intended for the corresponding user. More specifically we follow the following steps,

1) Decompose the message W into two submessages W_A and W_D and divide the power P into two powers P_A and P_D so that $P = P_A + P_D$. Message W_A will be decoded by both the receivers while message W_D will be decoded by only one receiver at a time. We will transmit it twice so that both the receivers can decode (see encoding and decoding rules below for a further description).
2) Generate a codebook $\mathcal{C}_A$ for W_A where the codewords U_A^n are sampled from i.i.d. a Gaussian distribution $U_A = X_A + \alpha_A A$. Here X_A is Gaussian $\mathcal{N}(0, P_A)$, independent of A, and $\alpha_A = P_A/(P+Q/2+1)$. A total of $2^{nI(U_A;Y_i)}$ codewords are thus generated and randomly partitioned into $2^{nI(U_A;A)}$ bins. The rate of this codebook, $I(U_A;Y_i) - I(U_A;A)$ can be shown to be[9]
$$R_A = \frac{1}{2}\log\left(1+\frac{P_A}{P_D+Q/2+1}\right). \tag{22}$$
3) Generate two codebooks $\mathcal{C}_D^{(1)}$ and $\mathcal{C}_D^{(2)}$ for W_D for the two receivers as follows. For $\mathcal{C}_D^{(1)}$, the codewords U_D^n are sampled from i.i.d. Gaussian distribution $U_D = X_D + \alpha_D((1-\alpha_A)A + D)$, where X_D is Gaussian $\mathcal{N}(0, P_D)$, independent of A and D, and $\alpha_D = P_D/(P_D+1)$. Generate $2^{nI(U_D;Y_1,U_A)}$ such codewords and partition them into $2^{nI(U_D;A,D)}$ bins. Follow analogous construction for codebook $\mathcal{C}_D^{(2)}$. The rate of each codebook[10] $I(U_D;Y_i,U_A) - I(U_D;A,D)$ can be shown to be
$$R_D = \frac{1}{2}\log(1+P_D). \tag{23}$$

Encoding: We transmit a superposition of two sequences corresponding to W_A and W_D as follows.

1) To encode a message W_A, find a codeword U_A^n in the bin of W_A, such that $X_A^n = U_A^n - \alpha_A A^n$ satisfies a power constraint of P_A. By construction, such a codeword exists with high probability.
2) To encode W_D, we decide whether to send it to user 1 or 2. The users are served alternately. When we decide to send it to user 1, we select a codeword U_D^n in the bin of codebook $\mathcal{C}_D^{(1)}$ corresponding to message W_D such that
$$X_D^n = U_D^n - \alpha_D\{(1-\alpha_A)A^n + D^n\}$$

[6] All logarithms are to the base 2 in this work. Also the notation $[f]^+$ refers to $\max(f, 0)$ in (15) and throughout the correspondence.

[7] The trivial upper bound of $\frac{1}{2}\log(1+P)$ is sometimes tighter than these two bounds, particular in the limit of very small P.

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

[9] Using a symmetry argument or otherwise, note that $I(U_A;Y_1) = I(U_A;Y_2)$, so we use the generic term $I(U_A;Y_i)$ to denote either of these.

[10] Notice that the codebooks can be the same for two users. For notational convenience while dealing with the two users we keep the codebooks separate.

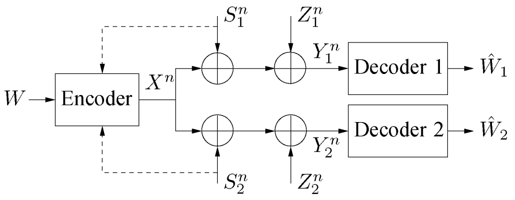

Fig. 4. Two-user Gaussian multicast channel model with additive interference. The encoder maps message W into codeword X^n. The state takes the form of interference sequences S_1^n and S_2^n. Each channel output $Y_k^n = X^n + S_k^n + Z_k^n$ is decoded to produce message estimate $\hat{W}_k$. The interference and noise sequences are i.i.d. and mutually independent. Furthermore, $S_1, S_2 \sim \mathcal{N}(0, Q)$ and $Z_1, Z_2 \sim \mathcal{N}(0, 1)$.

satisfies a power constraint of P_D. When we decide to transmit to user 2, we select a codeword U_D^n in the bin of codebook $\mathcal{C}_D^{(2)}$ corresponding to message W_D such that

$$X_D^n = U_D^n - \alpha_D\{(1-\alpha_A)A^n - D^n\}$$

satisfies the power constraint of P_D. Since there are $2^{nI(U_D;A,D)}$ codewords in each bin, such a codeword exists with high probability.

3) Send the superposition $X^n = X_A^n + X_D^n$, which has power P, over the channel.

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

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

$$\begin{aligned} Y_1^n &= X_A^n + A^n + (D^n + Z_1^n + X_D^n) \\ &= U_A^n + (1-\alpha_A)A^n + (D^n + Z_1^n + X_D^n) \\ Y_2^n &= X_A^n + A^n + (-D^n + Z_2^n + X_D^n) \\ &= U_A^n + (1-\alpha_A)A^n + (-D^n + Z_2^n + X_D^n). \end{aligned}$$

Since $D^n + Z_i^n + X_D^n$ is an i.i.d. Gaussian $\mathcal{N}(0, P_D + Q/2 + 1)$ sequence, independent of A^n, our choice of rate R_A in (22) ensures that the resulting $\hat{W}_A$ equals W_A with high probability at both the receivers.

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

$$\tilde{Y}_1^n = X_D^n + ((1-\alpha_A)A^n + D^n) + Z_1^n \tag{24}$$

$$\tilde{Y}_2^n = X_D^n + ((1-\alpha_A)A^n - D^n) + Z_2^n. \tag{25}$$

The rate R_D in (23) ensures that U_D^n can be decoded from either $\tilde{Y}_1^n$ or $\tilde{Y}_2^n$ so that the resulting $\hat{W}_D$ equals W_D with high probability at the corresponding receiver. Specifically, for the fraction of time that the transmitter encodes W_D for interference $(1-\alpha_A)A^n + D^n$, user 1 can recover W_D, while for the fraction of time that the transmitter encodes W_D for interference $(1-\alpha_A)A^n - D^n$, user 2 can recover W_D.

From this coding strategy, we see that the average rate delivered to each receiver is identical, i.e., $R_A + (1/2)R_D$. Maximizing this rate over the choices of P_A and P_D subject to the constraint $P = P_A + P_D$ optimizes the lower bound, whence (19a). □

From (18), 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 A^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 ($Q \ll 1$), Fig. 5 reflects the rather obvious fact that the side information can be ignored by the transmitter without sacrificing rate. Similarly, when the INR is large ($Q \gg 1$), Fig. 5 reflects that time sharing between the two users achieves the capacity. More generally

$$\lim_{Q\to\infty} C \leq \lim_{Q\to\infty} R_+^{\mathrm{I}} = \lim_{Q\to\infty} R_+^{\mathrm{II}} = \frac{1}{4}\log(1+P) \tag{26}$$

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 [24].

Perhaps more interestingly, our proposed achievable rate is optimal in the limit of high SINR. The behavior of the bounds as a function of SNR is depicted in Fig. 6 for a fixed INR of $Q = 15$ dB. We note that the expression for R_+^{II} coincides with R_- in this limit. Note that the baseline schemes (namely, time sharing and ignoring side information) do not achieve a rate particularly close to capacity, but the superposition dirty paper coding strategy corresponding to our lower bound does. More generally, we can show that

$$\lim_{P\to\infty}(C - R_-) \leq \lim_{P\to\infty}(R_+^{\mathrm{II}} - R_-) = 0. \tag{27}$$

To verify (27) for $Q \geq 2$, since $P \to \infty$, the middle case of the lower bound (18) applies which we can alternately express in the form

$$R_- = \frac{1}{2}\log\left(\frac{P + Q/2 + 1}{\sqrt{2Q}}\right). \tag{28}$$

Comparing (28) with the upper bound (15) we have

$$R_+^{\mathrm{II}} - R_- = \frac{1}{2}\log\frac{P + Q + 1 + 2\sqrt{PQ}}{\sqrt{2Q}} - \frac{1}{2}\log\frac{P + Q/2 + 1}{\sqrt{2Q}} \tag{29}$$

which in the limit $P \to \infty$ gives (27). The case $Q \leq 2$, can be similarly verified. We summarize the optimality properties in the following corollary.

Corollary 1: For the Gaussian multicast channel in Fig. 4, the proposed achievable rate in Theorem 4 is optimal in the limit of high SINR ($P \to \infty$, Q is fixed). For $Q > 2$ it can be expressed as $C(P) = \frac{1}{2}\log\left(\frac{P}{\sqrt{2Q}}\right) + o(1)$, where $o(1) \to 0$ as $P \to \infty$. For $Q \leq 2$ it can be expressed as $C(P) = \frac{1}{2}\log\left(\frac{P}{1+Q/2}\right) + o(1)$. Finally, for the case of fixed P and $Q \to \infty$, time sharing between the two users is optimal and the capacity can be expressed as $C(P) = \frac{1}{4}\log(1+P) + o(1)$, where $o(1) \to 0$ as $Q \to \infty$.

Finally, we show in Appendix III-C that a universal constant that bounds the difference between our upper and lower bounds is given by

$$\sup_{P,Q} R_+^{\mathrm{II}} - R_- = \frac{1}{2}\log\left(\frac{3}{2} + \sqrt{2}\right) = 0.7716. \tag{30}$$

We conclude this section with a few additional observations.

Some Further Remarks:

1) Extension to K receivers: Our upper-bounding technique for R_+^{II} in (15) can be extended to the case of K receivers each with independent interference. We show in Appendix III-D that the following upper bound holds for the case of K receivers:

$$\begin{aligned} R_+^K \leq \frac{1}{2}\log(P + Q + 1 + 2\sqrt{PQ}) - \frac{K-1}{2K}\log Q \\ -\frac{1}{2K}\log K - \left[\frac{1}{2K}\log\left(\frac{Q}{K(P+1)}\right)\right]^+. \end{aligned} \tag{31}$$

By taking the limit $Q \to \infty$ in (31), one can verify that time sharing is optimal for any number of users in the high-INR limit.

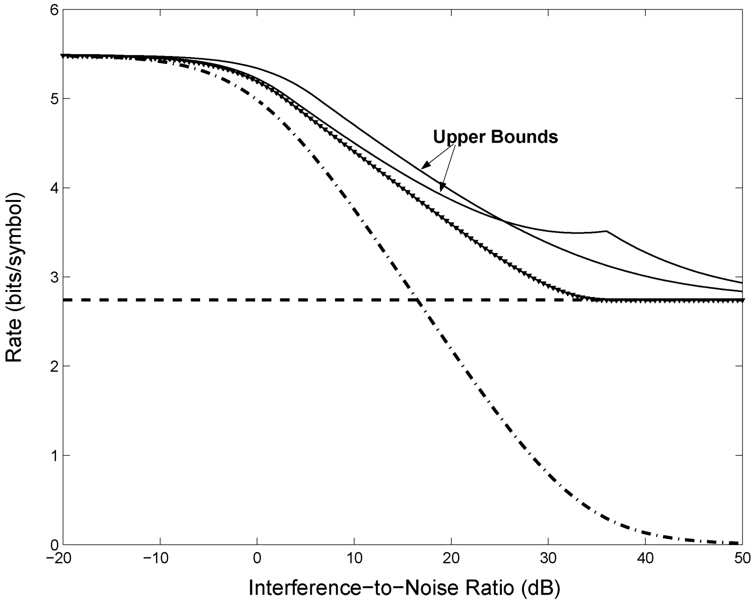

Fig. 5. Upper and lower bounds on the capacity of the two-user Gaussian multicast channel, as a function of INR Q for an SNR $P = 33$ dB. The upper two curves depict the two upper bounds from (14) and (15). The marked line is the achievable rate in (18). The horizontal dashed line indicates the performance 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.

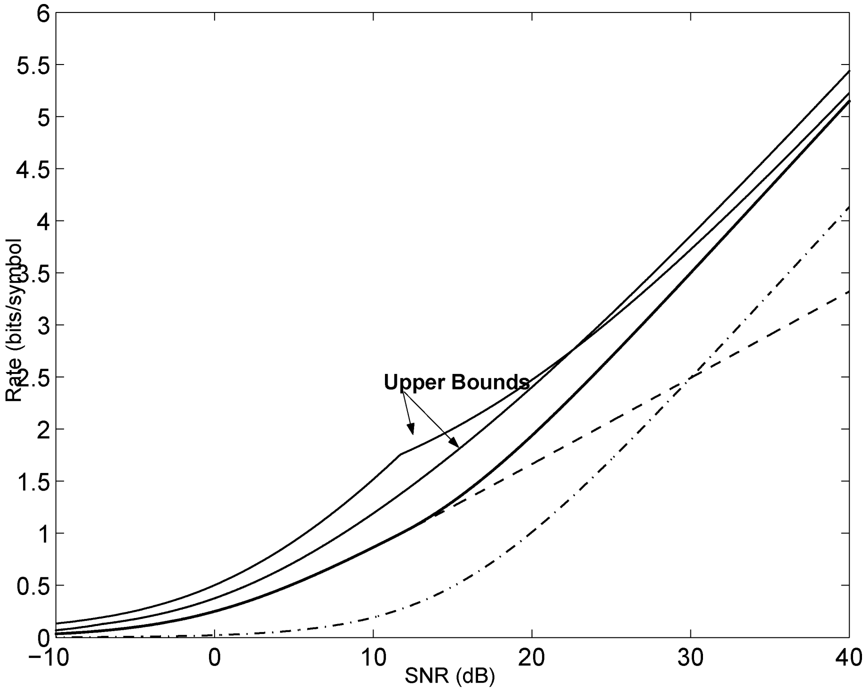

Fig. 6. Upper and lower bounds on the capacity of the two-user Gaussian multicast channel, as a function of SNR P for an INR $Q = 15$ dB. The upper two curves depict the two upper bounds in (14) and (15). The achievable rate in (18) is also shown. The dashed curve indicates the performance of time-sharing, 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.

2) Feedback does not help much. As discussed in Appendices III-B and IV-B, the expressions for R_+^{I} and R_+^{II} in (16) and (17) continue to hold in the presence of perfect causal feedback, provided we set ρ to equal the actual correlation between the noise terms—and do not optimize over it.

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 (19a) in Theorem 4, the obvious generalizations do not. In particular, strategies which precancel the interference in part of the codeword for each user achieved lower rates than our superposition dirty paper coding.

A. Correlated Interferences and Robust Dirty Paper Coding

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

$$Y^n = X^n + S^n + Z^n \tag{32}$$

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 $S_0^n \sim \mathcal{N}(0, Q\boldsymbol{I})$ 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 kth 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. In what follows, we focus on that case where there are two receivers i.e., $\beta \in \{\beta_1, \beta_2\}$. Extensions to the case of more than two receivers are possible, but will not be explored.

We first provide a general upper bound for the case of correlated, jointly Gaussian interference sequences and then specialize it to the case of scaled interferences. The general upper bound might be of independent interest and is derived in Appendix V.

Theorem 5: Consider a two-receiver channel model $Y_i^n = X^n + S_i^n + Z^n$ for $i = 1, 2$ when Z^n is i.i.d. $\mathcal{N}(0,1)$ noise, S_1^n and S_2^n are i.i.d. jointly Gaussian with marginal distributions $\mathcal{N}(0, Q_1)$ and $\mathcal{N}(0, Q_2)$, respectively, and suppose that the distribution of $S_1 - S_2$ is $\mathcal{N}(0, Q_d)$. An upper bound on the common message rate for this channel under a power constraint P at the transmitter is given by

$$R_+^{\mathrm{C}} = \sum_{i=1}^{2} \frac{1}{4} \log(P + Q_i + 1 + 2\sqrt{PQ_i}) - T(Q_d) \tag{33}$$

where

$$T(Q_d) = \begin{cases} \frac{1}{4}\log(Q_d), & Q_d > 4 \\ \frac{1}{2}\log\left(1 + \frac{Q_d}{4}\right), & Q_d \le 4. \end{cases} \tag{34}$$

We note that the upper bound is of most interest in the high-SINR limit i.e., when we fix Q_1, Q_2. and take $P \to \infty$.

Corollary 2: In the high-SINR limit (Q_1, Q_2 fixed, $P \to \infty$), the upper bound on the case of correlated interferences in Theorem 5 can be written as

$$R_+^{\mathrm{C}} = \frac{1}{2}\log(P) - T(Q_d) + o(1) \tag{35}$$

where the term $o(1)$ approaches 0 as $P \to \infty$ and Q_1, Q_2 fixed and $T(Q_d)$ is given in (34).

To establish an achievable rate, we will consider a modification to our lower bound in Theorem 4 which considers the case of independent interferences. To deal with the case of correlated interferences, we will require that the encoder and decoders have access to a common source of randomness which will be used as a dither sequence.

Consider a superposition dirty paper coding strategy analogous to that in the proof of the lower bound in Theorem 4, whereby we decompose the interferences according to (20). In this case, we have that (21) specializes to

$$\begin{aligned} A^n &= \beta_A S_0^n \\ D^n &= \beta_D S_0^n \end{aligned} \tag{36}$$

where

$$\begin{aligned} \beta_A &= (\beta_1 + \beta_2)/2 \\ \beta_D &= (\beta_1 - \beta_2)/2. \end{aligned} \tag{37}$$

When we turn to implement the encoding step in the proof of the lower bound of Theorem 4, in which A^n is treated as interference and D^n as additional noise, the results of [6] cannot be directly applied since the interference and noise sequences are now correlated. Fortunately, this correlation does not cost us in terms of achievable rate if we assume that the encoder and decoder(s) have access to a source of common randomness in the form of a dither sequence. In particular, for the lattice coding strategy in [8], correlation between the interference and noise sequences does not change the achievable rate relative to the case when the noise and interference sequences are independent.[11] With this scheme, we obtain the following lower bound.

Theorem 6: An achievable rate for our example multicast channel with correlated interferences and common randomness at the encoder and decoders is given by

$$C^\beta(P) \ge \max_{\{(P_A,P_D): P_A \ge 0, P_D \ge 0, P_A + P_D \le P\}} R^\beta(P_A, P_D) \tag{38a}$$

where

$$R^\beta(P_A, P_D) = \frac{1}{2}\log\left(1 + \frac{P_A}{1 + Q_d/4 + P_D}\right) + \frac{1}{4}\log(1 + P_D) \tag{38b}$$

where $Q_d \triangleq (\beta_1 - \beta_2)^2 Q$ is the variance of $S_1 - S_2$.

Optimizing over P_A and P_D, gives the following achievable rate:

$$R_-^\beta(P) = \begin{cases} \frac{1}{2}\log\left(1 + \frac{P}{1 + Q_d/4}\right), & Q_d < 4 \\ \frac{1}{2}\log\left(\frac{P + 1 + Q_d/4}{\sqrt{Q_d}}\right), & 4 \le Q_d \le 4(P+1) \\ \frac{1}{4}\log(1 + P), & Q_d \ge 4P + 4. \end{cases} \tag{39}$$

We note that in the limit of high SINR, our expression for R_-^β in (39) is given by $R_-^\beta = \frac{1}{2}\log(P) - T(Q_d) + o(1)$, where $T(Q_d)$ is given as in (34). This coincides with the upper bound in (35) and thus establishes the optimality of our scheme in the high-SINR limit.

Corollary 3: The proposed achievable rate in Theorem 6 is optimal in the limit of high SINR (fixed Q_1, Q_2, $P \to \infty$) i.e., $\lim_{P\to\infty} C^\beta(P) - R_-^\beta(P) = 0$.

V. Concluding Remarks

We introduced the multicast channel model and analyzed the special cases of binary and Gaussian channels with additive interference. Our main observation in this work is that unlike the single user case, the lack of side information at the receiver strongly limits capacity. We show that in both the binary and Gaussian cases if the interfering sequences are independent, time sharing is optimal in the limit of large

[11] The common dither sequence is necessary in [8] since the result is for an arbitrary interference sequence. We believe that in our case, common randomness may not be necessary—but this fact remains to be shown.

interference. Also certain achievable rates and their optimality properties have been discussed. The capacity has been established for the two-user noiseless binary case and for the Gaussian case in the high SINR limit.

It may be possible to extend the upper bounding techniques in this correspondence to more general channel models and perhaps also sharpen the results for the Gaussian and binary cases. We emphasize, however, that the proposed bounds indicate an important engineering insight that there is a significant loss in dealing with more than one interference sequence at the transmitter, even when they are correlated. An interesting direction of future work would be to investigate the connections of this result with a recent result on MIMO broadcast channel with imperfect channel state information at the transmitter [19], where again it was shown that lack of perfect channel state information strongly limits the broadcast channel capacity.

Appendix I
Proof of the Converse in Theorem 1

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

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

$$H(W|Y_k^N) \le n\varepsilon_n, \qquad \text{for } k = 1, 2 \tag{40}$$

where ε_n is a sequence that approaches 0 as $n \to \infty$. We can use Fano's inequality to bound the rate as

$$\begin{aligned} nR &= H(W) \\ &= H(W|Y_1^n) + I(W; Y_1^n) \\ &\le n\varepsilon_n + H(Y_1^n) - H(Y_1^n|W) && (41) \\ &\le n\varepsilon_n + \sum_{j=1}^{n} H(Y_{1j}) - H(Y_1^n|W) && (42) \\ &\le n\varepsilon_n + n - H(Y_1^n|W) && (43) \end{aligned}$$

where (41) follows by using the Fano inequality (40), (42) follows from the chain rule and the fact that conditioning reduces the entropy, and (43) follows from the fact that Y_{1j} is binary valued. We can similarly bound the rate on the second user's channel as

$$nR \le n\varepsilon_n + n - H(Y_2^n|W). \tag{44}$$

Combining (43) and (44), we obtain

$$\begin{aligned} nR &\le n - \max\{H(Y_1^n|W), H(Y_2^n|W)\} + n\epsilon_n \\ &\le n - \frac{1}{2}\{H(Y_1^n|W) + H(Y_2^n|W)\} + n\epsilon_n \\ &\le n - \frac{1}{2}H(Y_1^n, Y_2^n|W) + n\epsilon_n && (45) \\ &\le n - \frac{1}{2}H(Y_1^n \oplus Y_2^n|W) + n\epsilon_n && (46) \\ &= n - \frac{1}{2}H(S_1^n \oplus S_2^n) + n\epsilon_n && (47) \\ &= n\left(1 - \frac{1}{2}H(S_1 \oplus S_2) + \epsilon_n\right) && (48) \end{aligned}$$

where (45) follows from the fact that conditioning reduces entropy, (46) follows from the fact that $Y_1^n \oplus Y_2^n$ is a deterministic function of (Y_1^n, Y_2^n), (47) follows from the fact that $Y_1 \oplus Y_2 = S_1 \oplus S_2$, and (48) 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 (45), we have that any achievable rate satisfies

$$nR \le n - \frac{1}{K}H(Y_1^n, Y_2^n, \ldots, Y_K^n|W) + n\epsilon_n. \tag{49}$$

Proceeding from (49) we obtain

$$\begin{aligned} nR - n\epsilon_n &\le n - \frac{1}{K}H(Y_1^n, Y_2^n, \ldots, Y_K^n|W) \\ &= n - \frac{1}{K}H(Y_1^n, Y_1^n \oplus Y_2^n, \ldots, Y_1^n \oplus Y_K^n|W) && (50) \\ &= n - \frac{1}{K}H(X^n \oplus S_1^n, S_1^n \oplus S_2^n, \ldots, S_1^n \oplus S_K^n|W) \\ &= n - \frac{1}{K}H(S_1^n \oplus S_2^n, \ldots, S_1^n \oplus S_K^n|W) \\ &\quad - \frac{1}{K}H(X^n \oplus S_1^n|S_1^n \oplus S_2^n, \ldots S_1^n \oplus S_K^n, W) \\ &= n - \frac{n}{K}H(S_1 \oplus S_2, \ldots, S_1 \oplus S_K) \\ &\quad - \frac{1}{K}H(X^n \oplus S_1^n|S_1^n \oplus S_2^n, \ldots S_1^n \oplus S_K^n, W) \\ &\le n - \frac{n}{K}H(S_1 \oplus S_2, \ldots, S_1 \oplus S_K) && (51) \end{aligned}$$

where (50) follows from the fact that the mapping

$$(Y_1^n, Y_2^n, \ldots Y_K^n) \to (Y_1^n, Y_1^n \oplus Y_2^n, \ldots, Y_1^n \oplus Y_2^n)$$

is invertible, and (51) follows from the fact that $S_1^n, S_2^n, \ldots S_K^n$ are all i.i.d. and independent of W.

Appendix III
Proof of Upper Bound (15) in Theorem 3

We now derive (15) for R_+^{II}. We first note that the capacity of the channel only depends on the marginal distributions $p(Y_1^n|X^n, S_1^n, S_2^n)$ and $p(Y_2^n|X^n, S_1^n, S_2^n)$ and not on the joint distribution $p(Y_1^n, Y_2^n|X^n, S_1^n, S_2^n)$. Allowing correlation between the noise Z_1 and Z_2 does not change capacity. Specifically, we have the following.

Lemma 1: 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 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, p. 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

$$\begin{aligned} P_e^{1,n} &= \Pr\left(g_1(Y_1^n) \neq W\right) \\ P_e^{2,n} &= \Pr\left(g_2(Y_2^n) \neq W\right) \\ P_e^n &= \Pr\left(\bigcup_{k=1,2} \{g_k(Y_k^n) \neq W\}\right). \end{aligned}$$

Next, note that

$$\max\{P_e^{1,n}, P_e^{2,n}\} \le P_e^n \le P_e^{1,n} + P_e^{2,n} \tag{52}$$

where the left inequality in (52) follows from the fact that by definition $P_e^n \ge P_e^{k,n}$ 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 (52) 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 appendix 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 2: 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 show that a sequence of $(2^{nR}, n)$ codes that can be decoded by both the receivers with $P_e^n \to 0$ must satisfy $R \le R_+^{\mathrm{II}}$ in (17). Note that our power constraint is of the form $E[X_i^2] \le P_i$ with $\sum_{i=1}^n P_i \le nP$.

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 $R \le \min(R_1, R_2)$.

From Fano's inequality, we have that for some sequence ε_n, which approaches 0 as $n \to \infty$

$$H(W|Y_k^n) \le n\varepsilon_n, \qquad \text{for } k = 1, 2. \tag{53}$$

We first upper-bound R_1 as

$$\begin{aligned}
nR_1 &< I(W; Y_1^n) + n\epsilon_n \\
&= h(Y_1^n) - h(Y_1^n|W) + n\epsilon_n \\
&\le \sum_{i=1}^n h(Y_i) - h(Y_1^n|W) + n\epsilon_n && (54)\\
&\le \sum_{i=1}^n \frac{1}{2}\log 2\pi e(P_i + 1 + Q + 2\sqrt{P_i Q}) - h(Y_1^n|W) + n\epsilon_n && (55)\\
&\le \frac{n}{2}\log 2\pi e(P + 1 + Q + 2\sqrt{PQ}) - h(Y_1^n|W) + n\epsilon_n && (56)
\end{aligned}$$

where (54) follows from the chain rule and the fact that conditioning reduces entropy, and (55) follows from the fact that each Y_i has a variance no larger than $P_i + 1 + Q + 2\sqrt{P_i Q}$ and its differential entropy can be upper-bounded by that of a Gaussian random variable (RV). Finally, (56) is a consequence of Jensen's inequality.

Similarly, applying the above chain of inequalities on user 2, we have

$$nR_2 \le \frac{n}{2}\log 2\pi e(P + 1 + Q + 2\sqrt{PQ}) - h(Y_2^n|W) + n\epsilon_n. \tag{57}$$

Now we can find an upper bound on the common information rate using (56) and (57)

$$\begin{aligned}
nR &= n\min(R_1, R_2) \le \frac{n}{2}(R_1 + R_2) \\
&\le \frac{n}{2}\log 2\pi e(P + 1 + Q + 2\sqrt{PQ}) - \frac{1}{2}h(Y_1^n|W) \\
&\quad - \frac{1}{2}h(Y_2^n|W) + n\varepsilon_n \\
&\le \frac{n}{2}\log 2\pi e(P+1+Q+2\sqrt{PQ}) - \frac{1}{2}h(Y_1^n, Y_2^n|W) + n\varepsilon_n && (58)
\end{aligned}$$

where the last inequality (58) follows from the fact that conditioning reduces the differential entropy.

We now need to lower-bound $h(Y_1^n, Y_2^n|W)$. In what follows, we will also use the notation $S_+^n = \frac{S_1^n + S_2^n}{\sqrt{2}}$ and $S_-^n = \frac{S_1^n - S_2^n}{\sqrt{2}}$. Note that S_+ and S_- are mutually independent, Gaussian $\mathcal{N}(0, Q)$.

$$\begin{aligned}
&h(Y_1^n, Y_2^n|W) \\
&\quad =, h\left(\frac{Y_1^n - Y_2^n}{\sqrt{2}}, \frac{Y_1^n + Y_2^n}{\sqrt{2}} \,\middle|\, W\right) && (59)\\
&\quad = h(S_-^n + Z_-^n, \sqrt{2}X^n + S_+^n + Z_+^n|W) && (60)\\
&\quad = h(S_-^n + Z_-^n|W) + h(\sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) && (61)\\
&\quad = h(S_-^n + Z_-^n) + I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) \\
&\qquad + h(\sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n, S_+^n) && (62)\\
&\quad \ge h(S_-^n + Z_-^n) + I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) \\
&\qquad + h(\sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n, S_+^n, X^n) && (63)\\
&\quad = h(S_-^n + Z_-^n) + I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) \\
&\qquad + h(Z_+^n). && (64)
\end{aligned}$$

The preceding steps are justified as follows. In (59), we have used the fact that the differential entropy is invariant to a transformation of unit determinant. We substitute for Y_1 and Y_2 in (60). Equation (61) follows from the chain rule. In (62), we first drop the conditioning over W in the first term, since (S_-^n, Z_-^n) are jointly independent of W and expand the second term. Finally, (63) follows from the fact that conditioning on X^n further reduces the differential entropy while (64) is a consequence from Z_+^n being independent of $(X^n, S_+^n, S_-^n, Z_-^n, W)$.

Since S_-^n, Z_+^n, Z_-^n are all i.i.d. Gaussian with powers Q, $1+\rho$, and $1-\rho$, respectively, we have from (64)

$$\begin{aligned}
h(Y_1^n, Y_2^n|W) &\ge I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) \\
&\quad + \frac{n}{2}\log 2\pi e(Q + 1 - \rho) + \frac{n}{2}\log 2\pi e(1 + \rho). && (65)
\end{aligned}$$

It remains to lower-bound the mutual information term in (65). We first note that since S_+^n is independent of (W, S_-^n, Z_-^n) one can drop the conditioning in the mutual information expression.

Lemma 3: For each $n \ge 1$ and for any distribution $p(X^n|S_-^n, S_+^n, W)$ such that $\sum_{i=1}^n E[X_i^2] \le nP$. The mutual information term in (65) can be lower-bounded as

$$\begin{aligned}
&I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n) \\
&\quad \ge I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n) \ge \left[\frac{n}{2}\log\left(\frac{Q}{2P + 1 + \rho}\right)\right]^+. && (66)
\end{aligned}$$

Proof: The left-hand inequality follows immediately by expanding $I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n|W, S_-^n + Z_-^n)$ and using the fact that S_+^n is independent of (S_-^n, Z_-^n, W).

The right-hand side is a consequence of the rate-distortion theorem for i.i.d. Gaussian sources. Note that

$$E\left[\sum_{i=1}^n (\sqrt{2}X_i + Z_{+i})^2\right] \le n(2P + 1 + \rho).$$

Thus, if the right inequality were violated, for a certain distribution $p(X^n|S_+^n)$, we could use it as a test channel in quantizing a n-dimensional i.i.d. Gaussian source and do better than the rate distortion bound. Alternately, note that

$$\begin{aligned}
&I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n) \\
&\quad = h(S_+^n) - h(S_+^n|\sqrt{2}X^n + S_+^n + Z_+^n) \\
&\quad = h(S_+^n) - h(\sqrt{2}X^n + Z_+^n|\sqrt{2}X^n + S_+^n + Z_+^n) && (67)
\end{aligned}$$

$$
\begin{aligned}
&\geq h(S_+^n) - h(\sqrt{2}X^n + Z_+^n) && (68)\\
&\geq h(S_+^n) - \sum_{i=1}^{n} h(\sqrt{2}X_i + Z_{+,i}) && (69)\\
&\geq \frac{n}{2}\log Q - \sum_{i=1}^{n}\frac{1}{2}\log(2P_i + 1 + \rho) && (70)\\
&\geq \frac{n}{2}\log Q - \frac{n}{2}\log(2P + 1 + \rho) && (71)\\
&= \frac{n}{2}\log\left(\frac{Q}{2P+1+\rho}\right). && (72)
\end{aligned}
$$

Here (67) follows from the fact that $h(X|Y) = h(Y - X|Y)$, (68) from the fact that removing the conditioning on $\sqrt{2}X^n + S_+^n + Z_+^n$ only increases the differential entropy, (69) follows from the chain rule, (70) follows from the fact that the differential entropy with a fixed variance is maximized for a Gaussian distribution, and (71) follows from Jensen's inequality. Combining (72) with the trivial bound $I(S_+^n; \sqrt{2}X^n + S_+^n + Z_+^n) \geq 0$, we establish (66). □

Substituting, (66), (65) into (58), we get

$$
\begin{aligned}
R \leq \frac{1}{2}\log\left(\frac{P+Q+1+2\sqrt{PQ}}{\sqrt{(Q+1-\rho)(1+\rho)}}\right)&\\
- \left[\frac{1}{4}\log\left(\frac{Q}{2P+1+\rho}\right)\right]^+ + \varepsilon_n.& \qquad (73)
\end{aligned}
$$

Finally, since ρ is a free parameter of choice, we can select it to be the value that minimizes (73) and thus, (17) follows. To obtain the tightest possible bound we can optimize over the value of ρ. We obtain (15) by selecting the following choice for ρ:

$$
\rho^*(Q) = \begin{cases} Q/2, & \text{if } Q \leq 2\\ 1, & \text{if } Q > 2. \end{cases} \qquad (74)
$$

A. Gains From Feedback

In the presence of feedback, the transmitted symbol at time i depends on the past output, i.e., $x_i = f(w, y_1^{i-1}, y_2^{i-1}, s^n)$. In this situation, $Z_{+,i}$ is still independent of (W, Z_-^n, S^n, X_1^i). This condition suffices, for deriving the bounds in (58), (65), and (66). Lemma 1 does not hold, however, since now the joint distribution between noise sequences does matter in the probability of error. So while the expression (73) holds, one cannot optimize over ρ, but must select the value to be the actual correlation coefficient in the channel.

B. Universal Gap Between Upper and Lower Bounds

In this subsection, we verify (30), the gap between upper and lower bounds for all values of P and Q. We consider three different cases.

For $Q \leq 2$, we have

$$
R_+^{\mathrm{II}} - R_- = \frac{1}{2}\log\left(\frac{P+Q+1+2\sqrt{PQ}}{P+1+Q/2}\right). \qquad (75)
$$

It can be verified that the maximum for $P \geq 0$ and $0 \leq Q \leq 2$ occurs for $Q = 2$ and $P = 1/4(9 - \sqrt{17})$. The maximum value is $1/2\log((5+\sqrt{17})/4) \approx 0.5947$.

For the case $2 \leq Q \leq 2(P+1)$, the difference is also given by (75). The supremum is attained when we set $Q = 2(P+1)$ and let $P \to \infty$. The supremum value is $1/2\log((3+2\sqrt{2})/2) \approx 0.7716$.

Finally, for the case $Q \geq 2(P+1)$, the difference between the bounds is given by

$$
R_+^{\mathrm{II}} - R_- = \frac{1}{2}\log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q}\right).
$$

The supremum is obtained by taking $Q = 2(P+1)$ and letting $P \to \infty$ and again equals $1/2\log((3+2\sqrt{2})/2)$.

C. The Case of K Receivers

We consider the case where there are K receivers. To get an upper bound, we assume perfect correlation between the noise sequences, i.e., receiver $k = 1, 2, \ldots K$ gets $Y_k^n = X^n + S_k^n + Z^n$, where the interferences S_k^n are mutually independent and i.i.d. $\mathcal{N}(0, Q)$ and Z^n is i.i.d. $\mathcal{N}(0, 1)$.

To upper-bound the common rate for the case of K receivers, first note that the derivation that leads to (58) can be straightforwardly generalized to yield

$$
\begin{aligned}
nR \leq \frac{n}{2}\log 2\pi e(P+Q+1+2\sqrt{PQ})&\\
- \frac{1}{K}h(Y_1^n, Y_2^n, \ldots Y_K^n|W) + n\varepsilon_n.& \qquad (76)
\end{aligned}
$$

We now consider generalizing our derivation for (65) to lower-bound $h(Y_1^n, Y_2^n, \ldots Y_K^n|W)$. Let us consider a set of K orthogonal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots \boldsymbol{v}_K$, where $\boldsymbol{v}_1 = \frac{1}{\sqrt{K}}[1, 1, \ldots, 1]$ and $\boldsymbol{v}_2, \ldots \boldsymbol{v}_K$ are arbitrarily chosen. Let $\boldsymbol{Y}^n = (Y_1^n, Y_2^n, \ldots, Y_K^n)$ denote the K–tuple of received sequences.

Claim 1: The component-wise inner product of $\boldsymbol{Y}^n$ with $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ satisfies

$$
\begin{aligned}
\langle \boldsymbol{Y}^n, \boldsymbol{v}_1\rangle &= \sqrt{K}X^n + \sqrt{K}Z^n + T_1^n\\
\langle \boldsymbol{Y}^n, \boldsymbol{v}_j\rangle &= T_j^n, \qquad \text{for } j = 2, 3, \ldots K \qquad (77)
\end{aligned}
$$

where $T_1^n, T_2^n, \ldots T_K^n$ are mutually independent, i.i.d. Gaussian $\mathcal{N}(0, Q)$ sequences.

Proof: The expression for $\langle \boldsymbol{Y}^n, \boldsymbol{v}_1\rangle$ can be verified by direct substitution. Here $T_1^n = \frac{1}{\sqrt{K}}(S_1^n + S_2^n + \cdots + S_K^n)$. Since $\boldsymbol{v}_j$ and $\boldsymbol{v}_1$ are mutually orthogonal for $j \geq 2$, we have $\sum_{i=1}^{K} v_{ji} = 0$. Hence, $\langle \boldsymbol{Y}^n, \boldsymbol{v}_j\rangle = \sum_{i=1}^{K} v_{ji}S_i^n$. We denote $T_j^n = \sum_{i=1}^{K} v_{ji}S_i^n$. Since the S_j^n are mutually independent and i.i.d. and $\boldsymbol{v}_j$ are mutually orthogonal it follows that T_j^n are all mutually independent and i.i.d. $\mathcal{N}(0, Q)$. □

We can now lower-bound $h(Y_1^n, Y_2^n, \ldots Y_K^n|W)$ in a manner analogous to the derivation in (65)

$$
\begin{aligned}
&h(Y_1^n, Y_2^n, \ldots Y_K^n|W)\\
&= h(\langle \boldsymbol{Y}_1^n, \boldsymbol{v}_1\rangle, \langle \boldsymbol{Y}_2^n, \boldsymbol{v}_2\rangle, \ldots \langle \boldsymbol{Y}_K^n, \boldsymbol{v}_K\rangle\,|W) && (78)\\
&= h(\sqrt{K}X^n + \sqrt{K}Z^n + T_1^n, T_2^n, \ldots, T_K^n|W) && (79)\\
&= h(T_2^n) + \cdots + h(T_K^n)\\
&\quad + h(\sqrt{K}X^n + \sqrt{K}Z^n + T_1^n|T_2^n, \ldots, T_K^n, W) && (80)\\
&= \frac{n(K-1)}{2}\log 2\pi eQ + h(\sqrt{K}X^n + \sqrt{K}Z^n + T_1^n|W, \{T_j^n\}_{j=2}^K) && (81)\\
&= \frac{n(K-1)}{2}\log 2\pi eQ + h(\sqrt{K}X^n + \sqrt{K}Z^n + T_1^n|W, \{T_j^n\}_{j=1}^K)\\
&\quad + I(T_1^n; \sqrt{K}X^n + \sqrt{K}Z^n + T_1^n|T_2^n \ldots T_K^n, W) && (82)\\
&\geq \frac{n(K-1)}{2}\log 2\pi eQ + \frac{n}{2}\log 2\pi eK\\
&\quad + I(T_1^n; \sqrt{K}X^n + \sqrt{K}Z^n + T_1^n|T_2^n \ldots T_K^n, W) && (83)\\
&\geq \frac{n(K-1)}{2}\log 2\pi eQ + \frac{n}{2}\log 2\pi eK + \left[\frac{n}{2}\log\left(\frac{Q}{K(P+1)}\right)\right]^+. && (84)
\end{aligned}
$$

The justification for the preceding steps is as follows. In (78) we have used the fact that the differential entropy is invariant to a rotation, while (79) follows from Claim 1. In (80) and (81) we have used the fact that T_j^n are mutually independent, i.i.d., and independent of W. Equation (83) follows by additionally conditioning the entropy term in (82) with

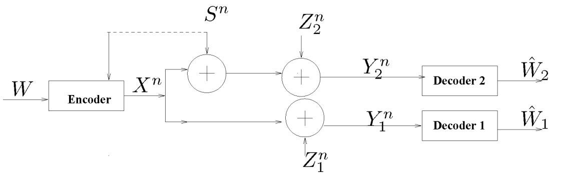

Fig. 7. Two-user Gaussian Channel with one-interference sequences. We derive the upper bound on the capacity of this channel and show that this is also an upper bound for the two-interference channel in Fig. 4. Here only receiver 2 experiences additive white Gaussian interference of variance Q.

X^n and using the fact that Z^n is independent of $(W, X^n, T_1^n, \dots T_K^n)$. Finally, (84) follows from fact that since T_1^n is independent of $\{T_j^n\}_{j=2}^K$ and W we can use an argument analogous to that in Lemma 3 to have

$$I(T_1^n; \sqrt{K}X^n + \sqrt{K}Z^n + T_1^n | T_2^n \dots T_K^n, W) \geq \left[\frac{n}{2}\log\left(\frac{Q}{K(P+1)}\right)\right]^+.$$

Finally, substituting (84) in (76), we obtain (31).

Appendix IV
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 capacity of the two-interference channel of interest in Theorem 3 cannot be higher.

As shown in Fig. 7, the single-interference channel is one in which $S_1^n = 0$ and $S_2^n = S^n$. Only the second receiver experiences interference.

The subsequent two lemmas establish that an upper bound on the capacity of the single-interference channel is also an upper bound on the capacity of the two-interference channel in Fig. 4.

Lemma 4: Suppose that for the single-interference channel model in Fig. 7, the encoder and decoder 1 have access to a source of common randomness Θ, which is independent of the message W and (S, Z_1, Z_2). Then, the capacity of the single-interference Gaussian channel is at least as large as the channel with two independent interferences in Fig. 4.

Proof: The proof follows by observing that using the source of common randomness Θ, we can generate an i.i.d. Gaussian $\mathcal{N}(0, Q)$ sequence S_C^n, for any value of n. This sequence is independent of all other channel parameters and is known to both the encoder and decoder 1. It is used to simulate the two independent interference channel as follows. Decoder 1 simply adds this sequence to the received output, and ignores its knowledge in decoding. The encoder has to deal with two sequences (S_C^n, S^n), both i.i.d. Gaussian $\mathcal{N}(0, Q)$. With this transformation, any coding scheme for the two-interference channel in Fig. 4 can be used over this channel with arbitrarily small probability of error. □

Lemma 5: A source of common randomness Θ, which is independent of the message W and the channel parameters (S, Z_1, Z_2) cannot increase the capacity of the single-interference channel in Fig. 7.

Proof: Our proof is analogous to the proof that common randomness does not increase the capacity in the single-user case in [8]. We argue that for any sequence of codes, given a stochastic encoder and decoder that depends on the shared random variable Θ, there exists a deterministic encoder and decoder with a smaller probability of error.

Given the message m and state sequence s^n, and a realization θ of the shared random variable, the encoding function (cf. Definition 1) is given by $x^n = f(m, s^n, \theta)$. Similarly, the decoding functions are given by $\hat{m}_k = g_k(y_k^n, \theta)$ for $k = 1, 2, \dots, K$. The average probability of error for the rate R randomized code is then defined by

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

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 Θ, 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 θ for which the term inside the expectation is minimized. We can design the encoding and decoding function for this deterministic value of θ 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 channel of interest. □

Lemmas 4 and 5 imply that an upper bound on the capacity of the single-interference channel in Fig. 7 is also an upper bound on the two independent-interference channel in Fig. 4. So we will derive an upper bound for the former.

Invoking the result of Lemma 1, we can let $E[Z_1Z_2] = \rho$, where $\rho \in [-1, 1]$ will be optimized later. As in Appendix III, define $Z_- \triangleq (Z_1 - Z_2)/\sqrt{2}$ and $Z_+ \triangleq (Z_1 + Z_2)/\sqrt{2}$.

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 $R \leq \min(R_1, R_2)$. Similar to our derivation in Appendix III, we use Fano's inequality to bound R_1 and R_2 as

$$nR_1 \leq \frac{n}{2}\log 2\pi e(P+1) - h(Y_1^n|W) + n\epsilon_n \tag{85}$$

$$nR_2 \leq \frac{n}{2}\log 2\pi e(P+1+Q+2\sqrt{PQ}) - h(Y_2^n|W) + n\epsilon_n. \tag{86}$$

Our bound for R follows the derivation analogous to that for (58) and is given by

$$nR \leq \frac{n}{4}\log 2\pi e(P+1+Q+2\sqrt{PQ}) + \frac{n}{4}\log 2\pi e(P+1) - \frac{1}{2}h(Y_1^n, Y_2^n|W) + 2n\varepsilon_n. \tag{87}$$

It remains to lower-bound the joint-entropy term in (87)

$$\begin{aligned}
&h(Y_1^n, Y_2^n|W)\\
&= h\left(\frac{Y_1^n + Y_2^n}{\sqrt{2}}, \frac{Y_1^n - Y_2^n}{\sqrt{2}}\,\middle|\, W\right) \qquad (88)\\
&= h\left(\sqrt{2}X^n + Z_+^n + \frac{1}{\sqrt{2}}S^n, -\frac{1}{\sqrt{2}}S^n + Z_-^n\,\middle|\, W\right)\\
&= h\left(-\frac{1}{\sqrt{2}}S^n + Z_-^n\,\middle|\, W\right)
\end{aligned}$$

$$
\begin{aligned}
&+ h\left(\sqrt{2}X^n + Z_+^n + \frac{1}{\sqrt{2}}S^n \,\middle|\, W, -\frac{1}{\sqrt{2}}S^n + Z_-^n\right)\\
&\geq h\left(-\frac{1}{\sqrt{2}}S^n + Z_-^n\right)\\
&\quad + h\left(\sqrt{2}X^n + Z_+^n + \frac{1}{\sqrt{2}}S^n \,\middle|\, W, -\frac{1}{\sqrt{2}}S^n + Z_-^n, S^n, X^n\right) \qquad (89)\\
&= h\left(-\frac{1}{\sqrt{2}}S^n + Z_-^n\right) + h(Z_+^n) \qquad (90)\\
&= \frac{n}{2}\log 2\pi e\left(\frac{Q}{2} + 1 - \rho\right) + \frac{n}{2}\log 2\pi e\,(1+\rho). \qquad (91)
\end{aligned}
$$

In the preceding steps, (88) follows from the fact that differential transformation is invariant under a pure rotation, (89) follows from the fact that the pair (S^n, Z_-^n) is independent of W and conditioning on additional terms only reduces the second term, while (90) follows from the fact that Z_+^n is independent of all other variables in the second term.

Substituting (91) into (87) and rearranging, we get

$$
R \leq \frac{1}{4}\log\left(\frac{1+P}{1+\rho}\right) + \frac{1}{4}\log\left(\frac{P+Q+1+2\sqrt{PQ}}{Q/2+1-\rho}\right) + \varepsilon_n. \qquad (92)
$$

Thus, we have shown the expression for (16). To obtain the tightest bound we minimize the right-hand side of the above over ρ. The tightest bounds is obtained with the choice

$$
\rho^*(Q) = \begin{cases} Q/4, & \text{if } Q \leq 4 \\ 1, & \text{if } Q > 4. \end{cases} \qquad (93)
$$

Substituting this value of ρ, in (92) yields (14).

A. Gains From Feedback

As noted in Appendix III-B, in the presence of causal feedback it still holds that $Z_{+,i}$ is independent of (W, Z_-^n, S^n, X_1^i). It can be verified that with this condition, the derivation that leads to (91) continues to hold and the upper bound in (92) remains valid. One cannot however optimize over ρ in the presence of feedback as Lemma 1 fails to hold in the presence of feedback.

Appendix V
Case of Correlated Interferences

In this appendix, we present the derivation of the upper bound in Theorem 5. The derivation is a minor modification of the derivation for the case of independent interferences. So only the steps that need to be modified will be presented. As in the statement of the theorem, we assume that $S_1 \sim \mathcal{N}(0, Q_1)$, $S_2 \sim \mathcal{N}(0, Q_2)$, and $S_1 - S_2 \sim \mathcal{N}(0, Q_d)$.

We first note that using Fano's inequality and the steps that lead to (58) in Appendix III, an upper bound on the common rate can be shown to be

$$
nR \leq \frac{1}{2}h(Y_1^n) + \frac{1}{2}h(Y_2^n) - \frac{1}{2}h(Y_1^n, Y_2^n|W) + n\varepsilon_n. \qquad (94)
$$

Using the power constraint, we upper-bound

$$
h(Y_i^n) \leq \frac{n}{2}\log 2\pi e(P + Q_i + 1 + 2\sqrt{PQ_i}), \qquad \text{for } i = 1, 2.
$$

It remains to lower-bound the joint entropy term. In what follows, we denote $Z_+^n = \frac{Z_1^n+Z_2^n}{2}$ and $Z_-^n = Z_1^n - Z_-^n$. Note that Z_+^n and Z_-^n are mutually independent and i.i.d. samples from $\mathcal{N}(0, (1+\rho)/2)$ and $\mathcal{N}(0, 2(1-\rho))$, respectively.

$$
\begin{aligned}
&h(Y_1^n, Y_2^n|W)\\
&= h\left(Y_1^n - Y_2^n, \frac{Y_1^n + Y_2^n}{2}|W\right) \qquad (95)\\
&= h\left(S_1^n - S_2^n + Z_-^n, X^n + \frac{S_1^n + S_2^n}{2} + Z_+^n|W\right)\\
&= h(S_1^n - S_2^n + Z_-^n)\\
&\quad + h\left(X^n + \frac{S_1^n + S_2^n}{2} + Z_+^n|W, S_1^n - S_2^n + Z_-^n\right) \qquad (96)\\
&\geq h(S_1^n - S_2^n + Z_-^n) + h(Z_+^n)\\
&= \frac{n}{2}\log 2\pi e(Q_d + 2(1-\rho)) + \frac{n}{2}\log 2\pi e\left(\frac{1+\rho}{2}\right). \qquad (97)
\end{aligned}
$$

Here (95) follows from the fact that the transformation $\begin{bmatrix} 1 & -1 \\ 1/2 & 1/2 \end{bmatrix}$ has unit determinant and the differential entropy is invariant to this transformation, (96) from the fact that $S_1^n - S_2^n + Z_-^n$ is independent of W and (97) from the fact that Z_+^n is independent of all other variables. The optimal value of ρ, which yields the largest value for the lower bound is given by $\rho^* = \min(1, Q_d/4)$ and the corresponding lower bound is given by

$$
h(Y_1^n, Y_2^n) \geq \begin{cases} n\log(2\pi e)\left(1 + \frac{Q_d}{4}\right), & \text{if } Q_d \leq 4 \\ \frac{n}{2}\log(2\pi e)^2 Q_d, & \text{if } Q_d > 4. \end{cases} \qquad (98)
$$

Finally, substituting (98) in (94) gives us the expression in (33).

Acknowledgment

The authors thank two anonymous reviewers for their insightful comments which helped to improve the quality of the correspondence.

References

[1] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[2] Y. Cemal and Y. Steinberg, "The multiple-access channel with partial state information at the encoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, 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. 47, no. 4, pp. 1423–1443, May 2001.

[4] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[5] M. H. M. Costa, private communication.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

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

[8] U. Erez, S. Shamai (Shitz), and R. Zamir, "Capacity and lattice strategies for cancelling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

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

[10] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Pered. Inform. (Probl. Inf. Transm.)*, vol. 9, no. 1, pp. 19–31, 1980.

[11] S. I. Gel'fand and M. S. Pinsker, "On Gaussian channels with random parameters," in *Proc. Int. Symp. Information Theory*, Tashkent, U.S.S.R., Sep. 1984, pp. 247–250.

[12] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 5, pp. 731–739, Sep. 1983.

[13] S. A. Jafar, "Capacity with causal and noncausal side information—A unified view," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5468–5475, Dec. 2006.

[14] A. Khisti, "Coding Techniques for Multicasting," M.S. thesis, MIT, Cambridge, MA, 2004.

[15] A. Khisti, U. Erez, and G. Wornell, "Writing on many pieces of dirty paper at once: The binary case," in *Proc. Int. Symp. Information Theory*, Chicago, IL, Jun./Jul 2004, p. 535.

[16] Y.-H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in *Proc. Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 534.

[17] S. Kotagiri and J. N. Laneman, "Achievable rates for multiple access channels with state information known at one encoder," in *Proc. 42nd Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.

[18] A. V. Kuznetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform. (Probl. Inf. Transm.)*, vol. 10, pp. 52–60, Apr.-Jun. 1974.

[19] A. Lapidoth, S. Shamai (Shitz), and M. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proc. 43rd Annual Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sep. 2005.

[20] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 563–593, Mar. 2003.

[21] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Devel.*, vol. 2, pp. 289–293, Oct. 1958.

[22] S. Sigurjonsson and Y. H. Kim, "On multiple user channels with causal state information at the transmitters," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 72–76.

[23] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2867–2877, Aug. 2005.

[24] Y. Steinberg and S. Shamai (Shitz), "Achievable rates for the broadcast channel with states known at the transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, 2005, pp. 2184–2188.

[25] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3926–3964, Sep. 2006.

[26] J. Wolfowitz, *Coding Theorems of Information Theory*. New York: Springer-Verlag, 1964.

Achievable Error Exponents for the Private Fingerprinting Game

Anelia Somekh-Baruch, *Member, IEEE*, and
Neri Merhav, *Fellow, IEEE*

***Abstract*—Fingerprinting systems in the presence of *collusive* attacks are analyzed as a game between a fingerprinter and a decoder on the one hand, and a coalition of two or more attackers on the other hand. The fingerprinter distributes, to different users, different fingerprinted copies of a host data (*covertext*), drawn from a memoryless stationary source, embedded with different fingerprints. The coalition members create a forgery of the data while aiming at erasing the fingerprints in order not to be detected. Their action is modeled by a multiple-access channel (MAC). We analyze the performance of two classes of decoders, associated with different kinds of error events. The decoder of the first class aims at detecting the *entire* coalition, whereas the second is satisfied with the detection of *at least one* member of the coalition. Both decoders have access to the original covertext data and observe the forgery in order to identify member(s) of the coalition. Motivated by a worst case approach, we assume that the coalition of attackers is informed of the hiding strategy taken by the fingerprinter and the decoder, while they are uninformed of the attacking scheme. Achievable single-letter expressions for the two kinds of error exponents are obtained. Single-letter lower bounds are also derived for the subclass of constant composition codes. These lower and the upper bounds coincide for the error exponent of the first class. Further, for the error of the first kind, a decoder that is optimal is introduced, and the worst case attack channel is characterized.**

***Index Terms*—Coding with side information, error exponents, information hiding, fingerprinting, private watermarking, randomized code, steganography, universal decoding, watermarking.**

I. Introduction

In fingerprinting systems, several copies of the same host data are embedded with different fingerprints (that designate, e.g., the different digital signatures or serial numbers of the copies they are provided with) and distributed to different users. The fingerprints identify one of many users in order to enable copyright protection. In this situation, two or more users can form a coalition, and collusive attacks on the fingerprinting system are possible and have to be taken into account in the code design. Each of the coalition members contributes his distinct fingerprinted copy in order to create a better forgery. Hence, the fingerprinting problem can be thought of as a game between the fingerprinter and the coalition of attackers.

As mentioned in [33], the fingerprinting game is closely related to (and is actually an extension of) the watermarking game, that in turn can be modeled as a coded communication system equipped with side information, for a single user as opposed to one of many users. Watermarking systems have been studied from the information-theoretic

Manuscript received February 10, 2005; revised June 7, 2006. The material in this correspondence was presented in part at the Canadian Workshop on Information Theory 2005, Montreal, QC, Canada, June 2005. This work was performed while A. Somekh-Baruch was with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel.

A. Somekh-Baruch was with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel. She is now with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544 USA (e-mail: anelia@princeton.edu).

N. Merhav is with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2007.894632