

Cooperative Encoding with Asymmetric State Information at the Transmitters

Anelia Somekh-Baruch Shlomo Shamai (Shitz) Sergio Verdú

Abstract—We generalize the Gel’fand-Pinsker model with the setup of a memoryless multiple-access channel where there is a single message source fed to both encoders. Only one of the encoders knows the state of the channel (non-causally), which is also unknown to the receiver. We find an explicit characterization of the capacity of this single-user channel. An explicit characterization of the capacity is also provided for the same channel with causal channel state information. Further, we apply the general formula to the Gaussian case with non-causal channel state information, in which capacity is achievable by a generalized writing-on-dirty-paper scheme.

I. INTRODUCTION

Communication over state-dependent channels has become a widely investigated research area. The framework of channel states available at the transmitter dates back to Shannon [1], who characterized the capacity of a state-dependent memoryless channel whose states are i.i.d. and available causally to the transmitter. In their celebrated paper [2], Gel’fand and Pinsker established a single-letter formula for the capacity of the same channel under the conceptually different setup where the transmitter observes the channel states non-causally. The main tool in proving achievability in this setup is the binning encoding principle [2]. Costa [3] applied Gel’fand Pinsker’s (GP) result to the Gaussian case, where there are two additive Gaussian noise sources, one of which, the interference, takes the role of the channel state. Costa originated the term “writing on dirty paper” which stands for an application of GP’s binning encoding scheme that adapts the transmitted signal to the channel states sequence rather than attempting to cancel it. This results in a surprising phenomenon - the operative upper bound, of a channel having no interference, can be attained, even though the channel states are not known to the receiver. It was shown in [4],[5], that this principle continues to hold even if the interference is not Gaussian. Extensions of these channel models to the multi-user case were performed by Gel’fand and Pinsker in [6] who showed that interference cancelation is also possible in the Gaussian broadcast channel, and the Gaussian Multiple Access Channel (MAC). Kim et al. [7] showed that a similar thing happens for the physically

degraded Gaussian relay channel. Steinberg and Shamai [8] provided achievable rates for the broadcast channel with states known non-causally at the transmitter. Another multi-user extension, where the channel state information (CSI) is causally available at the transmitters [1], was made by Steinberg [9] for the capacity region of the degraded broadcast channel. In [10], the capacity of the physically degraded relay channel with causal CSI was found. For other related work see [11], [12], [13], [14].

Much research interest has been devoted to applications of these channel models, for example, watermarking, [15], [16], [17], [18], [19], multi-input-multi-output (MIMO) broadcast channels, [20], and cooperative networks, [21].

In [22] and [23], the problem of a two-user GP MAC with CSI known non-causally to only one of the encoders, and each encoder transmitting a separate message, is addressed. While the symmetric (interference known to all the encoders) setup of [6] enables interference cancelation and the capacity region is characterized fully, here, only inner and outer bounds on the capacity of the additive white Gaussian MAC as well as the general discrete channel are derived. The informed encoder uses a generalized dirty paper coding (DPC) scheme that allows arbitrary correlation between the codeword and the known CSI.

In this paper we consider the GP memoryless two-user MAC, with CSI available non-causally to one of the encoders but not to the other encoder nor to the receiver. The problem considered here specializes to the two users transmitting a common message. In a way, this can be regarded as a single-user channel, in the sense that there is one message source and one destination. We refer to this channel as a Generalized GP (GGP) channel. We characterize the capacity for the general finite-alphabet case with a single-letter expression. This is enabled by a generalized binning coding scheme. It is argued that the single-letter expression remains the capacity even if one allows feedback at the informed encoder, but not at the uninformed encoder (similarly to the single-user GP channel setting [24]). While feedback does not increase the ultimate rate, it simplifies considerably the signalling technique which is capable of approaching capacity [24]. We also generalize [1] by providing a single-letter expression for the same setup considered in the GGP channel with the exception that the CSI is available *causally* to one encoder only. The channel, in this case, is referred to as an asymmetric causal state-dependent channel. Further, we consider the Gaussian channel with non-causal CSI. In contrast to Costa’s setup and to the symmetric Gaussian MAC [7], where the very trivial operative upper bound of a channel having no interference

This research was supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme. The work of S. Shamai and S. Verdú has also been supported by the Binational US-Israel Scientific Foundation, Grant 2004140.

Anelia Somekh-Baruch and Sergio Verdú are with the Department of Electrical Engineering, Princeton University, Princeton, New-Jersey 08544, USA {anelia,verdu}@princeton.edu

Shlomo Shamai (Shitz) is with the Department of Electrical Engineering, Technion I.I.T., Haifa 32000, Israel sshlomo@ee.technion.ac.il

is achievable, in our setup one cannot hope for complete interference cancelation. This renders the converse part of the theorem a more ambitious task. We define therefore an equivalent notion of interference cancelation that is adequate to our setup. We present an operative upper bound on the highest achievable rate and point out the loss due to the asymmetric side information. The resulting upper bound is shown to be achievable in the Gaussian case, yielding a closed-form expression for the capacity. We characterize the optimal strategy of the informed encoder balancing the trade-off between enhancing the signal of the uninformed encoder, decreasing the interference, and transmitting an additional information about the message that is not transmitted by the uninformed encoder. Another interesting insight derived from the proof is the capacity of a class of finite alphabet and Gaussian parallel channels with non-causal side information at the transmitter.

A Cognitive Radio (see, [25], [26], [27], [28]) is a device, added to an existing system having licensed users, that is capable of sensing its environment and making use of that knowledge to increase the spectral efficiency of the system. A useful model for the cognitive radio is as a transmitter with side information about the primary (licensed) transmission. An assumption taken in the models considered in [25], [28] is that the cognitive radio has non-causal knowledge of the codeword of the primary user. In our setup, the informed encoder can be thought of as a cognitive radio, which identifies the channel states (that can stand for other interfering signals) and helps the licensed user to transmit the message, exploiting its side information. The model applies also to cooperative transmission in the realm of the cognitive paradigm (that is one of the nodes is cognizant of the channel state which stands for information transmitted in the system). Another application of our results is to watermarking, where two encoders are jointly embedding the watermark. The first performs the embedding in a generic way, i.e., independently of the actual coverttext, and the second embeds information in a coverttext-dependent method. Our work accounts also for other scenarios of cooperative communication used to increase performance [21], [29].

The rest of this paper is organized as follows. In Section II we state the problem more explicitly and define some notation that will be used throughout the paper. Section III is devoted to establishing a single-letter expression for the GGP channel capacity in the discrete case, an upper bound on the capacity, and the capacity of a special class of GGP channels, referred to as degenerate parallel channels. The causal case is treated in Section IV where we provide the capacity formula for the asymmetric causal state-dependent channel. In Section V we apply the single-letter expression of the GGP channel to the Gaussian channel with non-causal CSI, and establish an explicit closed-form formula. Section VI concludes with a summary of the main contributions of this paper.

II. NOTATION AND STATEMENT OF THE PROBLEM

Throughout the paper, random variables will be denoted by capital letters, while deterministic realizations thereof will be denoted by lower-case letters. We shall use the short-hand notation x_i^j to abbreviate $(x_i, x_{i+1}, \dots, x_j)$, and $x^n = (x_1, \dots, x_n)$. For convenience, the n -vector x^n will occasionally be denoted by the boldface notation \mathbf{x} as well. The probability law of a random variable X will be denoted by P_X , and the conditional probability distribution of Y given X will be denoted by $P_{Y|X}$.

A stationary memoryless state-dependent Multiple-Access Channel is defined by a distribution Q_S on the set \mathcal{S} and the channel conditional probability distribution $W_{Y|S, X_1, X_2}$ from $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2$ to \mathcal{Y} . Let $X_{(1)}^n = (X_1(1), \dots, X_1(n))$ and $X_{(2)}^n = (X_2(1), \dots, X_2(n))$ designate the inputs of transmitters 1 and 2 to the channel, respectively. The output of the channel will be denoted by Y^n . The stationarity and memorylessness assumptions imply that

$$P_{Y^n | S^n, X_{(1)}^n, X_{(2)}^n} (y^n | s^n, x_{(1)}^n, x_{(2)}^n) = \prod_{i=1}^n W_{Y|S, X_1, X_2} (y_i | s_i, x_{i1}, x_{i2}).$$

The symbols $S_i, X_1(i), X_2(i)$ and Y_i represent the channel state, the channel inputs produced by two distinct encoders, and the channel output, at time index i , respectively. We assume that the channel states S^n are i.i.d., each distributed according to Q_S . As can be seen in Figure 1, the setup we consider is asymmetric in the sense that it is only encoder 2 which produces $X_{(2)}^n$ that is informed of the channel states, while the other encoder, which produces $X_{(1)}^n$, as well as the decoder are oblivious. Unlike the ordinary MAC, there is a single message source fed to both encoders. In the case where encoder 2 observes the CSI non-causally, we shall refer to this channel as a Generalized Gel'fand-Pinsker (GGP) channel, when encoder 2 observes the states causally, the channel will be referred to as an asymmetric causal state-dependent channel.

A sub-class of GGP channels that will be of special interest is the following. A *memoryless parallel channel with non-causal asymmetric side information* is a GGP channel with $Y = (Y_1, Y_2)$ and

$$W_{Y_1, Y_2 | S, X_1, X_2} = W_{Y_1 | X_1, S} W_{Y_2 | X_2, S}. \quad (1)$$

In words, this is a GGP channel with two outputs $Y_1(1), \dots, Y_1(n)$ and $Y_2(1), \dots, Y_2(n)$ that are both observed by the receiver. If, in addition, one has

$$W_{Y_2 | X_2, S} = W_{Y_2 | X_2} \quad (2)$$

we shall say that the parallel channel is degenerate.

A message, \mathcal{W} , is a random variable uniformly distributed over the set $\{1, \dots, M\}$ where $M = \lfloor e^{nR} \rfloor$. A rate- R code for the GGP channel is composed of two encoders $\varphi_n^{(1)}, \varphi_n^{(2)}$ and a decoder ψ_n : the first encoder is unaware of the CSI and thus is defined by a mapping

$$\varphi_n^{(1)}: \{1, \dots, M\} \rightarrow \mathcal{X}_1^n. \quad (3)$$

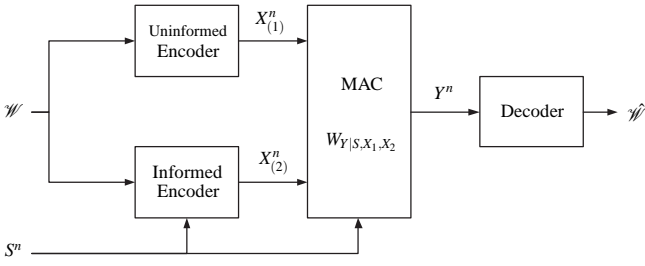


Fig. 1. Asymmetric state-dependent MAC with a common message.

The second encoder, observes the CSI non-causally, and is defined by a mapping

$$\varphi_n^{(2)} : \{1, \dots, M\} \times \mathcal{S}^n \rightarrow \mathcal{X}_2^n. \quad (4)$$

The decoder is a mapping

$$\psi_n : \mathcal{Y}^n \rightarrow \{1, \dots, M\}. \quad (5)$$

A rate- R code for the asymmetric causal state-dependent channel is defined similarly to that of the GGP channel, with the exception that the second encoder is defined by a sequence of mappings

$$\varphi_{n,i}^{(2)} : \{1, \dots, M\} \times \mathcal{S}^i \rightarrow \mathcal{X}_2 \quad i = 1, \dots, n, \quad (6)$$

and at time index i , the channel input is given by $X_2(i) = \varphi_{n,i}^{(2)}(\mathcal{W}, S^i)$.

An (ε, n, R) -code for the GGP channel is a code $(\varphi_n^{(1)}, \varphi_n^{(2)}, \psi_n)$ having average probability of error not exceeding ε , i.e.,

$$\Pr(\mathcal{W} \neq \psi_n(Y_1^n)) \leq \varepsilon. \quad (7)$$

A rate R is said to be achievable if there exists a sequence of (ε_n, n, R) -codes with $\lim_{n \rightarrow \infty} \varepsilon_n = 0$. The capacity of the GGP channel is defined as the supremum of all achievable rates. The definitions of an (ε, n, R) -code, an achievable rate and the capacity of the asymmetric causal state-dependent channel are similar.

III. A SINGLE-LETTER EXPRESSION FOR THE CAPACITY - FINITE INPUT ALPHABET GGP CHANNEL

Before we state the main result of this section, we note that when the two encoders are informed non-causally of the CSI, the single-letter expression for the capacity is given by

$$\max_{P_{X_1, X_2, U|S}} [I(U; Y) - I(U; S)] \quad (8)$$

with $|\mathcal{U}| \leq |\mathcal{S}| |\mathcal{X}_1| |\mathcal{X}_2|$, as a direct application of the GP channel to the case of input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$.

The following theorem provides a single-letter expression for the capacity of the finite-input-alphabet GGP channel, that is, when the alphabets $\mathcal{S}, \mathcal{X}_1, \mathcal{X}_2$ are finite.

Theorem 1 *The capacity of the finite input alphabet GGP channel is given by*

$$C_{GGP} = \max [I(U, X_1; Y) - I(U, X_1; S)], \quad (9)$$

where the maximum is over all the joint measures $P_{S, X_1, U, X_2, Y}$ on $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{U} \times \mathcal{X}_2 \times \mathcal{Y}$ having the form

$$P_{S, X_1, U, X_2, Y} = Q_S P_{X_1} P_{U|S, X_1} P_{X_2|S, U} W_{Y|S, X_1, X_2}, \quad (10)$$

where $|\mathcal{U}| \leq |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2|$.

We note that the theorem continues to hold if in (10) we replace $P_{X_2|S, U}$ by $P_{X_2|S, X_1, U}$, i.e., $X_1 \leftrightarrow (S, U) \leftrightarrow X_2$ does not have to be a Markov chain. Moreover, there exists a maximizing measure with X_2 that is a deterministic function of (S, X_1, U) . The fact that we can replace $P_{X_2|S, U}$ by $P_{X_2|S, X_1, U}$ yields the following corollary which provides an alternative expression for C_{GGP} .

Corollary 1 *The capacity of the finite input alphabet GGP channel is given by*

$$C_{GGP} = \max [I(V; Y) - I(V; S)], \quad (11)$$

where the maximum is over all the joint measures $P_{S, X_1, V, X_2, Y}$ on $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{V} \times \mathcal{X}_2 \times \mathcal{Y}$ having the form

$$P_{S, X_1, V, X_2, Y} = Q_S P_{X_1} P_{V, X_2|S, X_1} W_{Y|S, X_1, X_2}, \quad (12)$$

and X_1 is a deterministic function of V . The alphabet cardinality of V satisfies $|\mathcal{V}| \leq |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2|$.

It is noted that this result remains intact if we allow for feedback to the informed encoder, i.e., if, before producing the i -th channel input symbol, the informed encoder observes the previous channel outputs, Y^{i-1} , that is, while the uninformed encoder is a mapping of the form (3), the informed encoder is actually sequences of mappings $\varphi_n^{(2,i)} = \{\varphi_n^{(2,i)}\}_{i=1}^n$ with

$$\varphi_n^{(2,i)} : \{1, \dots, M\} \times \mathcal{S}^n \times \mathcal{Y}^{i-1} \rightarrow \mathcal{X}_2. \quad (13)$$

We now give a description of a capacity achieving random coding scheme. The proof of the converse part in Appendix A followed by the proof of the direct part of Theorem 1 which relies on this scheme and appears in Appendix B.

Note that for any measure of the form (10), one has

$$\begin{aligned} & I(X_1, U; Y) - I(X_1, U; S) \\ &= I(X_1; Y) + I(U; Y|X_1) - I(U; S|X_1). \end{aligned} \quad (14)$$

This r.h.s. expression provides the intuition that a capacity-achieving random coding scheme has the following two-stage structure. The message \mathcal{W} is split into two parts. The uninformed encoder encodes the first part of the message, and assuming it will become available to the receiver, the informed encoder encodes the second part of the message using a binning scheme similar to the one used in [2].

Encoding: Fix a measure $P_{S, X_1, U, X_2, Y}$ satisfying (10), and denote

$$\begin{aligned} L &= e^{n[I(X_1; Y) - \varepsilon]} \\ K &= e^{n[I(U; Y|X_1) - I(U; S|X_1) - \varepsilon]} \\ J &= e^{n[I(S; U|X_1) + 2\varepsilon]} \\ M &= e^{n[I(X_1, U; Y) - I(X_1, U; S) - 2\varepsilon]}. \end{aligned} \quad (15)$$

Note that from (15) we have $M = L \cdot K$.

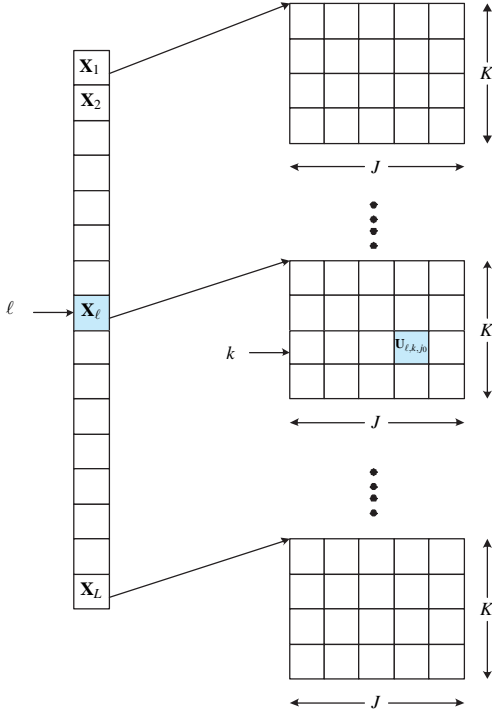


Fig. 2. A generalized binning coding scheme

The random encoders operate as follows: The uninformed encoder draws L i.i.d. vectors, $\{\mathbf{x}_\ell\}_{\ell=1}^L$, each with i.i.d. components drawn subject to P_{X_1} . The ordered collection of the drawn vectors constitutes the codebook used by the uninformed encoder.

For each codeword, \mathbf{x}_ℓ , the informed encoder draws $K \times J$ auxiliary vectors, denoted $\{\mathbf{u}_{\ell,k,j}\}$, $k = 1, \dots, K$, $j = 1, \dots, J$, independently and with i.i.d. components given \mathbf{x}_ℓ . Hence, each codeword in the uninformed user codebook is associated with a codebook of auxiliary codewords.

Since $M = LK$, transmitting a message, $m \in \{1, \dots, M\}$, is equivalent to transmitting $\ell \in \{1, \dots, L\}$ and $k \in \{1, \dots, K\}$. The uninformed encoder transmits \mathbf{x}_ℓ , and the informed encoder, which observes \mathbf{s} , is responsible for transmitting k . Transmission of k is done by searching for the lowest $j_0 \in \{1, \dots, J\}$ such that \mathbf{u}_{ℓ,k,j_0} is jointly typical with $(\mathbf{x}_\ell, \mathbf{s})$. Denote this j by $j(\mathbf{s}, \ell, k)$. If such j_0 is not to be found, or if the observed state sequence \mathbf{s} is non-typical, an error is declared and $j(\mathbf{s}, \ell, k)$ is set to $j = 1$.

Finally, the output of the second (informed) encoder is some vector $\tilde{\mathbf{x}}$ that is jointly typical with $(\mathbf{s}, \mathbf{u}_{\ell,k,j(\mathbf{s}, \ell, k)}, \mathbf{x}_\ell)$.

Decoding: Upon observing \mathbf{y} , the decoder searches for a pair $(\hat{\ell}, \hat{k})$ such that $\mathbf{x}_{\hat{\ell}}, \mathbf{u}_{\hat{\ell}, \hat{k}, j}$ are jointly typical with \mathbf{y} and outputs $\hat{m} = (\hat{\ell}, \hat{k})$. If there is no such pair, or it is not unique, an error is declared.

The analysis of the probability of error of this scheme is performed in Section B establishing the direct part of Theorem 1.

We now state a lemma (see [30] for the proof) that provides an upper bound on the capacity of the GGP channel.

It is the generalization of the trivial bound $\max_{P_{X_1|S}} I(X; Y|S)$ on the capacity of the ordinary GP channel. This lemma will be of great significance in the proof of the converse part of the coding theorem for the Gaussian GGP channel, since this upper bound is achievable in the Gaussian case.

Lemma 1 *The capacity of the finite input alphabet GGP channel is upper bounded by*

$$\max_{P_{X_1} P_{X_2|S, X_1}} [I(X_1, X_2; Y|S) - I(S; X_1|Y)], \quad (16)$$

where $P_{S, X_1, X_2, Y} = Q_S P_{X_1} P_{X_2|S, X_1} W_{Y|S, X_1, X_2}$.

In the following lemma, whose proof appears in [30], we find the capacity of the degenerate parallel GGP channel, and establish the fact that the CSI does not help.

Lemma 2 *The capacity of the degenerate parallel GGP channel is given by the sum of the capacities obtained without transmitter CSI.*

IV. THE CAUSAL ASYMMETRIC STATE-DEPENDENT CHANNEL

In this section we consider the causal asymmetric state-dependent channel. As a side note, we mention that if the CSI is available to both of the encoders, the single-letter expression for the capacity is deduced as a direct application of the formula derived by Shannon [1] for a channel with input alphabet $\mathcal{X}_1 \times \mathcal{X}_2$.

Theorem 2 *The capacity of the finite input alphabet causal asymmetric state-dependent channel is given by*

$$C_{causal} = \max I(U; Y), \quad (17)$$

where the maximum is over all the joint measures $P_{S, X_1, U, X_2, Y}$ on $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{U} \times \mathcal{X}_2 \times \mathcal{Y}$ having the form

$$P_{S, X_1, U, X_2, Y} = Q_S P_U P_{X_1|U} P_{X_2|S, U, X_1} W_{Y|S, X_1, X_2}, \quad (18)$$

where X_1 is a deterministic function of U , and $|\mathcal{U}| \leq |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2|$.

The proof can be found in [30] and is omitted due to space limitations.

V. THE GAUSSIAN GGP CHANNEL

In this section we analyze the additive Gaussian GGP channel. Based on the results obtained in Section III we derive a closed-form formula for the capacity, discuss it and provide numerical results.

A. Channel Model

The Gaussian GGP channel is given by

$$Y_i = X_1(i) + X_2(i) + S_i + N'_i. \quad (19)$$

The transmitted signals $X_{(1)}^n$ and $X_{(2)}^n$ have powers not exceeding P_1 and P_2 , respectively, that is

$$\frac{1}{n} \sum_{i=1}^n X_1^2(i) \leq P_1, \quad \frac{1}{n} \sum_{i=1}^n X_2^2(i) \leq P_2. \quad (20)$$

As before, the message is available to both encoders. Only the second encoder knows the realization of the interference S^n (non-causally). The noise processes, S^n and N^n , are assumed to be Gaussian i.i.d. zero-mean with $E(S_i^2) = Q$ and $E(N_i^2) = N$. The process N^n is independent of $(X_{(1)}^n, X_{(2)}^n, S^n)$.

It should be noted that had we been dealing with a constraint on the total received power, it is evident that all the power should have been assigned to the informed encoder, and the problem would have degenerated to the ordinary "dirty paper" Costa setup [3].

B. Capacity Formula and Discussion

We are interested in finding the highest reliably transmitted rate of a common message observed by the two encoders. To this end, using standard techniques [31], an application of the single-letter expression derived for the finite alphabet case to the Gaussian GGP channel $W_{Y|S, X_1, X_2}(y|s, x, x') = \frac{1}{\sqrt{2\pi N}} e^{-(y-s-x-x')^2/2N}$ gives

$$C_{GGP}(P_1, P_2, Q, N) = \sup_P [I(X_1, U; Y) - I(X_1, U; S)], \quad (21)$$

where the allowed joint distribution of S, X_1, U, X_2, Y satisfies (10), and $E(X_2^2) \leq P_2$, $E(X_1^2) \leq P_1$.

The following theorem provides an explicit formula for the capacity of this channel.

Theorem 3 *The capacity $C_{GGP}(P_1, P_2, Q, N)$ of the Gaussian GGP is given by the following formula*

$$C_{GGP}(P_1, P_2, Q, N) = \begin{cases} \frac{1}{2} \log \left(1 + \frac{P_1}{Q} \right) + \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right) & \text{if } \frac{P_1(P_2+N)^2}{(P_1+Q)^2} \leq P_2Q \\ \max_{-1 \leq \rho \leq 0} \frac{1}{2} \log \left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2} \sqrt{1-\rho^2})^2}{(\sqrt{Q} + \sqrt{P_2} \rho)^2 + N} \right) & \text{o.w.} \end{cases} \quad (22)$$

where, in fact, the maximization over ρ can be limited to either $\rho = -1$, $\rho = 0$ or $\rho = \frac{x^*}{\sqrt{P_2 Q}}$ where x^* is any real root of the 4th order polynomial

$$g(x) = (P_1 + Q)x^4 + 2Q(P_1 + P_2 + Q + N)x^3 + Q(Q(P_1 + 4P_2 + Q + 2N) - P_1P_2 + (N + P_2)^2)x^2 + 2Q^2P_2(-P_1 + P_2 + Q + N)x + Q^3P_2(P_2 - P_1) \quad (23)$$

that satisfies $-\sqrt{P_2 Q} \leq x^* \leq 0$.

The proof of Theorem 3 involves lower and upper bounding of (21) with bounds that coincide. The proof of Theorem 3 appears in [30] and is omitted due to space limitations.

In Costa's channel model [3], the GP capacity formula for the Gaussian channel was calculated explicitly. The proof relies on a capacity-achieving binning scheme which is shown to achieve the same reliably transmitted rate as if the interference S^n were not there. Hence, Costa's problem, as well as its multiuser counterpart [7], were special in that the trivial operative upper bound is achievable. The upper bound of Lemma 1 plays the role of the operative bound

and constitutes the core of the converse part. So, in fact, the generalization of interference cancellation to the GGP asymmetric setup is that the upper bound of Lemma 1 is achievable, a phenomenon that happens in the Gaussian GGP. Recalling (16), this implies that the subtracted term, $I(S; X_1|Y)$, can be interpreted as the rate loss incurred due to the fact that S is known only to the second transmitter (and not to both). Indeed, any information that X_1 conveys to Y about S is an inevitable waste of resources in terms of rate.

The main goal of the proof is to show that the maximizing distribution in (21) is such that (S, X_1, X_2) are jointly Gaussian, and an optimal choice for U is

$$U = X_2 + \alpha_{opt} S \quad (24)$$

with

$$\alpha_{opt} = \frac{P_2 P_1 Q - P_1 \sigma_{2s}^2 - P_1 N \sigma_{2s} - \sigma_{12}^2 Q}{P_2 P_1 Q + P_1 N Q - P_1 \sigma_{2s}^2 - \sigma_{12}^2 Q}, \quad (25)$$

where $\sigma_{12} = E(X_1 X_2)$ and $\sigma_{2s} = E(X_2 S)$ are chosen to maximize (21). The allowable values for the covariances, σ_{12} and σ_{2s} are such that the resulting covariance matrix $\Lambda_{X_1, X_2, S, N'}$ of (X_1, X_2, S, N') ,

$$\Lambda_{X_1, X_2, S, N'} = \begin{pmatrix} P_1 & \sigma_{12} & 0 & 0 \\ \sigma_{12} & P_2 & \sigma_{2s} & 0 \\ 0 & \sigma_{2s} & Q & 0 \\ 0 & 0 & 0 & N \end{pmatrix} \quad (26)$$

satisfies the semi-positive-definiteness condition

$$\det(\Lambda_{X_1, X_2, S, N'}) = P_1(P_2 Q N - \sigma_{2s}^2 N) - \sigma_{12}^2 Q N \geq 0, \quad (27)$$

i.e., for all $Q > 0$,

$$P_1 \sigma_{2s}^2 + Q \sigma_{12}^2 \leq P_1 P_2 Q, \quad (28)$$

or, in terms of correlation coefficients $\rho_{12} = \frac{\sigma_{12}}{\sqrt{P_1 P_2}}$ and $\rho_{12} = \frac{\sigma_{2s}}{\sqrt{P_2 Q}}$,

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \quad (29)$$

For reasons that will become clear in the sequel, we introduce the following terminology.

Definition 1 *The region of parameters P_1, P_2, Q, N such that*

$$\frac{P_1(P_2 + N)^2}{P_1 + Q} \geq P_2 Q \quad (30)$$

will be referred to as the silent region and its complement will be referred to as the active region.

Since Q, P_1, P_2, N take only non-negative values, the active region is equivalent to

$$\begin{aligned} P_1 &\leq \frac{P_2 Q^2}{(P_2 + N)^2 - P_2 Q} \\ \Leftrightarrow Q &\geq -\frac{P_1}{2} + \frac{\sqrt{P_1(P_1 P_2 + 4(N + P_2)^2)}}{2\sqrt{P_2}} \\ \Leftrightarrow N &\leq \sqrt{\frac{P_2 Q(P_1 + Q)}{P_1}} - P_2. \end{aligned} \quad (31)$$

so, in a sense, in the silent region the interference predominates, and in the active region the noise predominates.

In the sequel, we separate the discussion on the capacity formula to the two complementary regions, the silent region and the active region.

Silent Region: It is shown that in the silent region, the optimal values of σ_{12} and σ_{2s} are such that the condition (28) is met with equality, i.e.,

$$P_1\sigma_{2s}^2 + Q\sigma_{12}^2 = P_1P_2Q \quad (32)$$

or equivalently,

$$\rho_{12}^2 + \rho_{2s}^2 = 1. \quad (33)$$

This is also equivalent to

$$E\left(X_2 - \hat{X}_2^{lin}(X_1, S)\right)^2 = 0, \quad (34)$$

where $\hat{X}_2^{lin}(X_1, S)$ is the optimal linear estimator (in the MMSE sense) of X_2 given X_1 and S

$$\hat{X}_2^{lin}(X_1, S) = \frac{\sigma_{12}}{P_1}X_1 + \frac{\sigma_{2s}}{Q}S. \quad (35)$$

Eq. (34) implies that in the silent region

$$X_2 = \hat{X}_2^{lin}(X_1, S) = \frac{\sigma_{12}}{P_1}X_1 + \frac{\sigma_{2s}}{Q}S, \quad (36)$$

and thus,

$$Y = X_1 \left(1 + \frac{\sigma_{12}}{P_1}\right) + S \left(1 + \frac{\sigma_{2s}}{Q}\right) + N', \quad (37)$$

calculating the optimal value of α (25) while accounting for (32), yields

$$\begin{aligned} \alpha_{opt}^{silent} &= -\frac{\sigma_{2s}}{Q} \\ U_{opt}^{silent} &= X_2 - \frac{\sigma_{2s}}{Q}S = \frac{\sigma_{12}}{P_1}X_1 \end{aligned} \quad (38)$$

and hence, in the silent region of parameters, the capacity (22) formula is equal to

$$\begin{aligned} &\max_{\sigma_{12}, \sigma_{2s}} I(U, X_1; Y) - I(U, X_1; S)|_{U=\frac{\sigma_{12}}{P_1}X_1} \\ &= \max_{\sigma_{12}, \sigma_{2s}} I\left(X_1; X_1 \left(1 + \frac{\sigma_{12}}{P_1}\right) + S \left(1 + \frac{\sigma_{2s}}{Q}\right) + N'\right) \end{aligned} \quad (39)$$

with σ_{12}, σ_{2s} satisfying (33). Inspecting (39), it is easy to verify that an alternative selection of U , that is of more elegance

$$U_{opt}^{silent} = 0 \quad (40)$$

yields¹ the same achievable rate and hence is also optimal.

The fact that in the silent region the capacity is equal to (39), suggests that at this range, the informed encoder puts all its power into decreasing the interference and enhancing the signal of the uninformed encoder. No power is put into transmission of additional information, and hence, we refer to this region as silent.

¹This selection is legitimate due to the comment following Theorem 1, i.e., X_2 may depend on X_1 given (S, U) .

A useful geometrical interpretation to the capacity formula in the silent region can be attained by substituting $\cos \phi = \rho$ in (22), this yields

$$\max_{\phi} \frac{1}{2} \log \left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2} \cdot \sin \phi)^2}{(\sqrt{Q} + \sqrt{P_2} \cdot \cos \phi)^2 + N} \right), \quad (41)$$

where it is obvious that one should maximize over $\phi \in [\pi/2, \pi]$ to obtain a non-negative sine and a non-positive cosine. The larger ϕ is in $[\pi/2, \pi]$, a larger portion of user 2's power is devoted to reducing the interference and less to enhancing X_1 , and achieving the capacity in the silent region amounts to optimizing over ϕ (or ρ in (22)).

The maximizing ρ of the capacity formula in the silent region (see (22)) is either 0, -1 or any real root of the 4th order polynomial $g(x)$ (23) multiplied by $\sqrt{P_2Q}$. For example, when $P_1 = P_2 = P > 0$ and $N = Q > 0$, the parameters lie in the silent region, and finding the roots of $g(x)$ (23) degenerates to finding the roots of a 3rd order polynomial. It turns out that the optimal value of ρ corresponding to the real root of $g(x)$ is given by

$$\rho = \left(A^{1/3} + \frac{4 - 5\eta}{A^{1/3}(\eta + 1)} - 4 \right) \frac{1}{3\sqrt{\eta}} \quad (42)$$

with

$$A = 8 + 3\sqrt{3} \left(\frac{7\eta^3 - 4\eta^2 + 16\eta}{(\eta + 1)^3} \right), \quad \eta = \frac{P}{Q}. \quad (43)$$

Active Region: In the active region, the informed encoder balances the tradeoff among three goals: decreasing the interference, enhancing the signal of the uninformed encoder, and transmitting additional information (as opposed to the silent region where no additional information is transmitted). Therefore, this region of parameters is referred to as active. Keeping the other parameters fixed, the higher the interference Q is, the portion of the power that the informed user allocates to the additional information becomes larger at the expense of interference reduction and enhancement of the uninformed user's signal. In this region too, the maximizing (X_1, X_2, S) is Gaussian, with

$$\sigma_{12}^{active} = -\sigma_{2s}^{active} = \frac{P_1(P_2 + N)}{P_1 + Q}, \quad (44)$$

i.e.,

$$\rho_{12}^{active} = \frac{P_1(P_2 + N)}{\sqrt{P_1P_2}(P_1 + Q)}, \quad \rho_{2s}^{active} = \frac{P_1(P_2 + N)}{\sqrt{QP_2}(P_1 + Q)}. \quad (45)$$

The resulting α_{opt} (see (25)) when using the correlations (44) is given by

$$\alpha_{opt}^{active} = \frac{P_2}{P_2 + N}, \quad (46)$$

which is equal to the optimal α in Costa's setup [3] when the uninformed user is not present. The choice of correlations (44), results in a surprising phenomenon which happens only in the active region. The achievable rate is $\frac{1}{2} \log \left(1 + \frac{P_1}{Q} \right) + \frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$, the same as that of a decoder that observes both $Y_1 = X_1 + S$ and $Y_2 = X_2 + N'$ rather than $Y = X_1 + X_2 +$

$S + N'$. In other words, the upper bound of the Gaussian degenerate parallel channel with asymmetric non-causal CSI (see Lemma 2) can actually be achieved, even if the decoder is constrained to see only the sum of the channel outputs.

C. Capacity Achieving Coding Scheme

In this subsection we describe the capacity achieving scheme for the Gaussian GGP resulting (using standard techniques [31]) from that of the finite alphabet GGP.

Silent Region: Due to (36) and (40), here, no binning is needed, or in other words, this is a degenerate binning scheme with bin size 1. The uninformed encoder generates a random codebook consisting of $M = \lfloor \exp\{n(C_{GGP}(P_1, P_2, Q, N) - \varepsilon)\} \rfloor$ codewords $\{\mathbf{x}_m\}_{m=1}^M$ with i.i.d. symbols, each distributed according to $\mathcal{N}(0, P_1)$. Given a message m to be transmitted, which corresponds to the codeword $\mathbf{x}_m = (x_m(1), x_m(2), \dots, x_m(n))$, and a state-sequence \mathbf{s} , the informed encoder simply transmits the n -vector $\tilde{\mathbf{x}}$ whose i -th symbol is given by

$$\tilde{x}_i = x_m(i) \frac{\sigma_{12}}{P_1} + s_i \frac{\sigma_{2s}}{Q}, \quad (47)$$

where $\sigma_{2s} = \sqrt{P_2 Q} \cdot \rho$, with ρ being the maximizer in (22) and $\rho_{12} = \sqrt{1 - \sigma_{2s}^2}$. Either an ordinary Maximum Likelihood (ML) decoder or a typicality decoder can be used to achieve capacity.

Active Region: As stated earlier, in this region the informed encoder spends energy to interference reduction and enhancement of the uninformed user's signal as well as transmission of additional information. So as opposed to the silent region, the binning scheme is not void. The random scheme is as described in Section III, with Gaussian P_{S, X_1, X_2} with the covariance matrix

$$\begin{pmatrix} Q & 0 & \sigma_{2s}^{active} \\ 0 & P_1 & \sigma_{12}^{active} \\ \sigma_{2s}^{active} & \sigma_{12}^{active} & P_2 \end{pmatrix}$$

where $\sigma_{12}^{active}, \sigma_{2s}^{active}$ are defined in (44), and

$$U = X_2 + \alpha_{opt}^{active} S \quad (48)$$

(see (46)). The quantities that determine the sizes of the codebooks (15) used in the encoding scheme are given by

$$\begin{aligned} I(X_1; Y) &= \frac{1}{2} \log \left(\frac{(P_1 + Q)^2}{(Q(P_1 + Q) + P_1(P_2 + N))} \right) \\ I(U; Y|X_1) - I(U; S|X_1) &= \frac{1}{2} \log \left(\frac{(P_2 + N)(Q(P_1 + Q) + P_1(P_2 + N))}{(P_1 + Q)NQ} \right) \\ I(U; S|X_1) &= \frac{1}{2} \log \left(\frac{Q}{P_2 - \frac{2P_1 P_2}{P_1 + Q} + \frac{P_2^2 Q}{(P_2 + N)^2} - \frac{P_1(P_2 + N)^2}{(P_1 + Q)^2}} \right). \end{aligned}$$

D. Extreme Case Analysis

Table I summarizes the behavior of the capacity $C_{GGP}(P_1, P_2, Q, N)$ that can be deduced from (22) in several extreme cases. As expected, for infinite Q , the capacity degenerates to that of Costa's channel, that is, when the

TABLE I
EXTREME CASE ANALYSIS.

regime	Behavior of $C_{GGP}(P_1, P_2, Q, N)$
$Q \rightarrow \infty$	$\frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$
$P_1 = 0$	$\frac{1}{2} \log \left(1 + \frac{P_2}{N} \right)$
$Q = 0$	$\frac{1}{2} \log \left(1 + \frac{(\sqrt{P_1} + \sqrt{P_2})^2}{N} \right)$
$P_2 = 0$	$\frac{1}{2} \log \left(1 + \frac{P_1}{N + Q} \right)$

uninformed user is not present. This is because the amount of information that can be reliably transmitted by the uninformed user becomes negligible. A similar phenomenon happens when $P_1 = 0$.

For $Q = 0$, the only noise present is N' and thus there is no side information and the encoders transmit coherently.

If the power of the informed encoder, P_2 , is zero, then it cannot transmit information, nor help the uninformed user by partially canceling the interference and thus the capacity is as though the effective noise is $S + N'$ and the power used for transmission is P_1 .

E. Numerical Results

In Figure 3 the capacity is plotted as a function of Q for fixed values of P_1, P_2, N which, in turn, were chosen in two groups (the first group consists of $(P_1 = 2, P_2 = N = 1)$, $(P_1 = 4, P_2 = N = 2)$, and $(P_1 = 6, P_2 = N = 3)$ and the second has $(P_1 = 5, P_2 = 2, N = 1)$, $(P_1 = 10, P_2 = 4, N = 2)$, and $(P_1 = 20, P_2 = 8, N = 4)$). The capacity values for both $Q = 0$ and for $Q \rightarrow \infty$ are equal for all the members of each of these groups. The transition points between the silent region and the active region $Q = -\frac{P_1}{2} + \frac{\sqrt{P_1(P_1 P_2 + 4(N + P_2)^2)}}{2\sqrt{P_2}}$ (see (31)) are indicated with diamonds.

In Figure 4, the capacity and the optimal values of ρ_{2s} and ρ_{12} (the correlation coefficients between X_2 and S , and X_2 and X_1 , respectively) are depicted as a function of Q . Again, the transition points of the capacity curves from the silent region to the active region are indicated with diamonds. In the silent region, ρ_{2s} is, in fact, the maximizer of (22) and $\rho_{12} = \sqrt{1 - \rho_{2s}^2}$. In the active region, the optimal ρ_{12}, ρ_{2s} are given in (45). While ρ_{12} is a monotonically decreasing function of Q , $|\rho_{2s}|$ is increasing in the silent region and decreasing in the active region.

In Figure 5, the capacity is plotted as a function of P_1 for fixed values of P_2, Q, N . The diamonds indicate the points at which there are transitions from the active region to silent region, i.e., $P_1 = \frac{P_2 Q^2}{(P_2 + N)^2 - P_2 Q}$ (see (31)). The upper thick solid line stands for the plot of $Q = N = \frac{1}{2}$ and $P_2 = 1$, for which the transition occurs at $P_1 = 16$, a point which does not appear within the range depicted in this figure. The curves that meet at $P_1 = 0$ correspond to equal $\frac{P_2}{N}$ ratios, because the capacity is $\frac{1}{2} \log(1 + \frac{P_2}{N})$ for $P_1 = 0$.

In Figure 6, the capacity is plotted as a function of P_2 for fixed values of P_1, Q, N . The diamonds signify the points at which there are transitions from the active region to silent

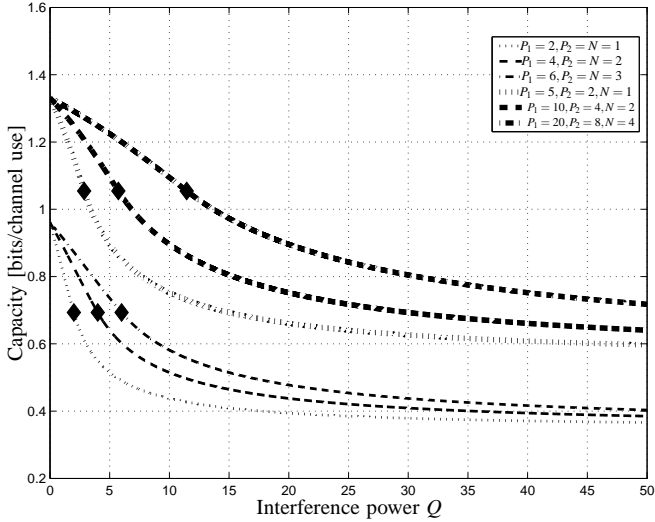


Fig. 3. Capacity as a function of the interference power Q .

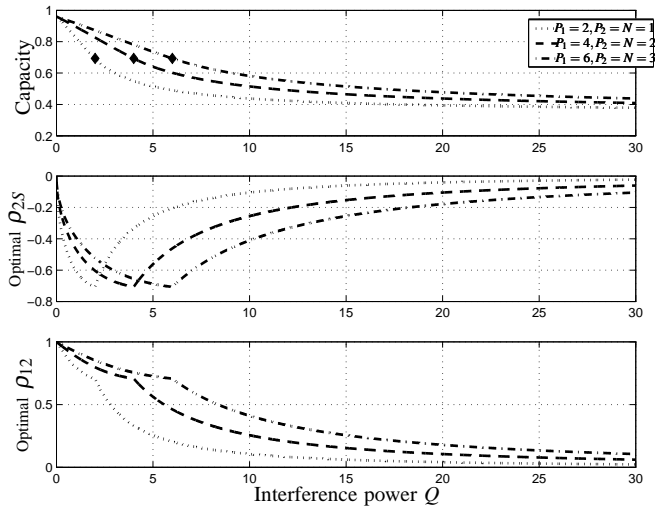


Fig. 4. Capacity and the optimal correlation coefficients, ρ_{12} and ρ_{2s} , as a function of the interference power Q .

region, i.e., $P_2 = \frac{Q(P_1+Q)-2P_1N+\sqrt{(Q(P_1+Q)-2P_1N)^2-4P_1^2N^2}}{2P_1}$. For the parameters $Q = 1, P_1 = 6, N = 3$, the entire curve is in the silent region.

VI. CONCLUSIONS

In this work we analyze a setup of cooperative communication over the GP MAC with a common message, referred to as the GGP channel, where the channel states are non-causally available to one user only. We assume that the users transmit a common message, and characterize the capacity of this channel for the general finite input-alphabet two-encoder case. Key to the characterization of the capacity is a generalized binning coding scheme. The message is split into two parts A and B . The uninformed encoder encodes part A of the message, and the informed encoder creates a codebook of auxiliary codewords for each codeword of the uninformed encoder using a binning scheme, and uses it to transmit the

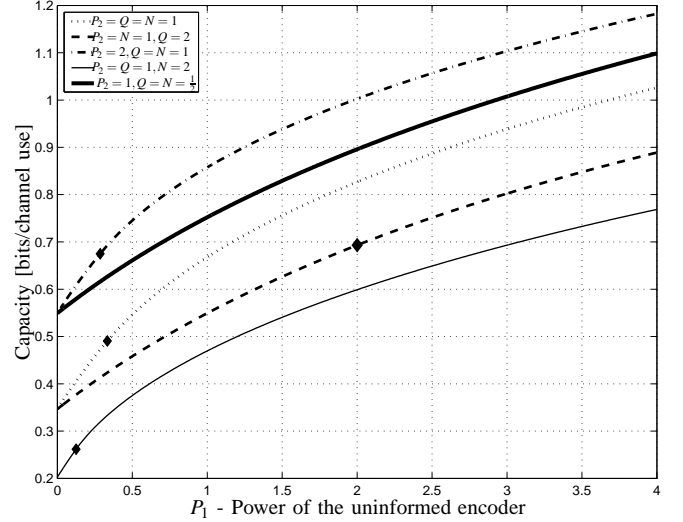


Fig. 5. Capacity as a function of P_1 .

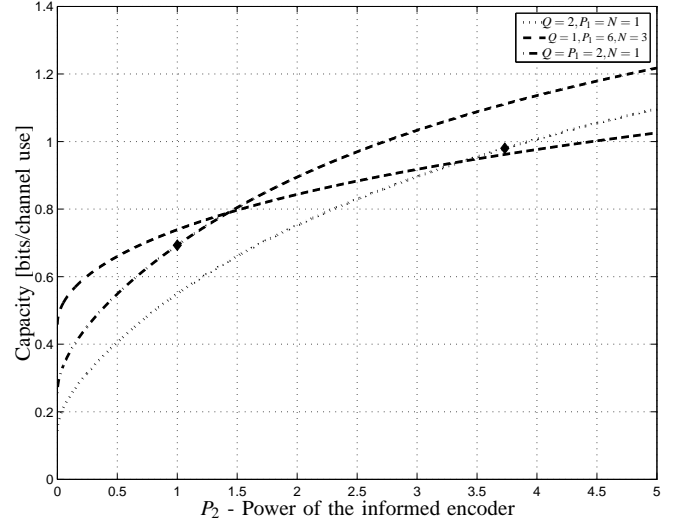


Fig. 6. Capacity as a function of P_2 .

part B of the message. The results are extendable to a general multiuser setup under the common message regime [30]. Another important extension carried out in [30] is to the capacity region of the case where in addition to the common message, the informed user is allowed to transmit a private message as well. Further, we establish two useful results for the general finite-input alphabet case. The first is a useful upper bound on the capacity of the GGP channel, referred to as an operative bound. This bound is the equivalent of a genie aided decoder observing the state information in the ordinary single-user GP channel. The second result relates to the special case of a GGP channel, a degenerate parallel GGP channel. We demonstrate that the knowledge of the CSI at the informed transmitter does not help in the degenerate parallel case, and derive the capacity formula of this channel as a special case of the general GGP channel capacity

formula. We also characterize the capacity of an asymmetric causal state-dependent channel which is the same channel as the GGP channel, with the exception that the CSI is available causally. Additionally, we focus on the two-encoder Gaussian GGP channel case, modeling the CSI as an additive Gaussian interference. By proving that in the Gaussian case the operative bound is achievable, we establish a closed-form formula for the capacity of this channel. Technically speaking, this upper bound enables proving that the maximizing distribution of the single-letter expression is Gaussian. Four parameters determine the capacity: the powers available to the two encoders, the interference power and the noise power. We partition the four dimensional space of all possible values of these parameters into two regions, a silent region and an active region. The capacity formula as a function of these four parameters takes on two different forms depending on whether the parameters lie in the active region or the silent region. In the silent region the informed encoder allocates a portion of its power to interference cancelation and the remaining power to enhancing the uninformed user's signal. In the active region, the encoder has the additional task of transmitting a part of the message that is not transmitted by the uninformed encoder. Surprisingly, we show that in the active region, the capacity is equal to that of a channel whose decoder observes two outputs (the first being the sum of the uninformed user's signal and the interference and the second being the sum of the informed user's signal and the noise).

APPENDIX

A. Converse Part of Theorem 1

The proof of the converse part of Theorem 1 is a quite straightforward extension of its GP counterpart [2]. To realize this, let an (ϵ_n, n, R) -code be given. Thus, we have

$$\begin{aligned}
nR &= H(\mathcal{W}) \\
&\leq I(\mathcal{W}; Y_1^n) + 1 + nR\epsilon_n \\
&\stackrel{(*)}{\leq} \sum_{i=1}^n [I(\mathcal{W}, Y^{i-1}, S_{i+1}^n; Y_i) - I(\mathcal{W}, Y^{i-1}, S_{i+1}^n; S_i)] + 1 + nR\epsilon_n \\
&= \sum_{i=1}^n I(\mathcal{W}, Y^{i-1}, S_{i+1}^n, X_1(i); Y_i) \\
&\quad - \sum_{i=1}^n I(\mathcal{W}, Y^{i-1}, S_{i+1}^n, X_1(i); S_i) + 1 + nR\epsilon_n \quad (49)
\end{aligned}$$

where the first inequality is Fano's Inequality, $(*)$ follows exactly as in the derivation of the converse part of the proof of the capacity formula for the ordinary GP channel [2], and the last equality follows from $X_1(i)$ being a function of \mathcal{W} .

Therefore, defining $\bar{U}_i = (\mathcal{W}, Y^{i-1}, S_{i+1}^n)$ one has

$$R \leq \frac{1}{n} \sum_{i=1}^n I(\bar{U}_i, X_1(i); Y_i) - I(\bar{U}_i, X_1(i); S_i) + \frac{1}{n} + R\epsilon_n. \quad (50)$$

Now, we introduce a time-sharing random variable, T , distributed uniformly over $\{1, \dots, n\}$, and denote

the collection of random variables $(S, X_1, \bar{U}, X_2, Y) = (S_T, X_1(T), \bar{U}_T, X_2(T), Y_T)$, to obtain

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n I(\bar{U}_i, X_1(i); Y_i) - I(\bar{U}_i, X_1(i); S_i) \\
&= I(\bar{U}, X_1; Y|T) - I(\bar{U}, X_1; S|T) \\
&= I(T, \bar{U}, X_1; Y) - I(T; Y) - I(T, \bar{U}, X_1; S) + I(T; S) \\
&\leq I(T, \bar{U}, X_1; Y) - I(T, \bar{U}, X_1; S), \quad (51)
\end{aligned}$$

where the last step follows by the stationarity of S_i . Substituting $U = (T, \bar{U})$ one gets

$$R \leq I(X_1, U; Y) - I(X_1, U; S) + \frac{1}{n} + R\epsilon_n. \quad (52)$$

Now, to verify that the appropriate Markovian conditions hold, we note that

- since $X_1(i)$ is the i -th output symbol of the uninformed encoder, it is independent of the channel state symbol S_i ,
- by definition of U_i , and by the fact that $X_2(i)$ is a function of (S_i^n, \mathcal{W}) , and since $S_i^n, X_1(i)$ are independent, we have $P_{X_2(i)|S_i, X_1(i), U_i} = P_{X_2(i)|S_i, U_i}$.

The above constitutes the proof that for every (ϵ_n, n, R) -code, there exists a measure of the form (10) with essentially $R \leq I(X_1, U; Y) - I(X_1, U; S)$.

It remains to show that the alphabet of the random variable U can be limited without loss of generality to $|\mathcal{U}| < |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2|$. This is done by a standard application of the support Lemma. Fix a distribution Q of (S, X_1, X_2, Y) and a conditional distribution μ on the Borel σ -algebra of $\mathcal{P}(\mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})$. Note that

$$\begin{aligned}
&I(X_1, U; Y) - I(X_1, U; S) \\
&= H(Y) - H(S) \\
&\quad - H(Y|U) + H(S|U) + I(X_1; Y|U) + I(X_1; S|U) \quad (53)
\end{aligned}$$

and $H(Y)$ and $H(S)$ are unaffected by U if Q is fixed, so due to the Markovity $U \leftrightarrow (S, X_1, X_2) \leftrightarrow Y$, it is only required that the expectation w.r.t. μ of the following functionals of a distribution Q on the set $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ are preserved

$$\begin{aligned}
f_{s,x,\bar{x}}(Q) &= Q(s, x, \bar{x}) \quad \forall (s, x, \bar{x}) \in \mathcal{S} \times \mathcal{X}_1 \times \mathcal{X}_2 \quad (54) \\
f_0(Q) &= H_Q(Y) - H_Q(S) + I_Q(X_1; Y) + I_Q(X_1; S) \quad (55)
\end{aligned}$$

To satisfy this condition, according to the support Lemma, since there are $A = |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2|$ functionals², the cardinality of the alphabet of \mathcal{U} can be taken to be A without loss of generality.

B. Direct Part of Theorem 1

Since the error probability analysis of the random coding scheme presented in Section III is also a rather straightforward extension of the proof of the GP direct, we shall state it in brevity.

²In (54), there are in fact only $A-1$ degrees of freedom.

Error probability analysis: For a measure P , let $T_\varepsilon(P)$ stand for the set of ε -typical sequences. One has

$$\begin{aligned} \Pr(\text{error}) &= \sum_{(\mathbf{s}, \mathbf{x}) \in T_\varepsilon^c(Q_S \times P_{X_1})} \Pr(\mathbf{s}, \mathbf{x}) \\ &+ \sum_{(\mathbf{x}, \mathbf{s}) \in T_\varepsilon(Q_S \times P_{X_1})} \Pr(\mathbf{s}, \mathbf{x}) \Pr(\text{error}|\mathbf{s}, \mathbf{x}) \end{aligned} \quad (56)$$

Due to the AEP, the probability that (\mathbf{s}, \mathbf{x}) are not jointly typical vanishes exponentially, thus, it is sufficient to upper bound the second term on the r.h.s. of (56). Assume that the transmitted message is $m = (\ell, k)$, and that \mathbf{s} is the state sequence, and \mathbf{x} is the output of the uninformed encoder. The error event is contained in the union of the following events

$$\begin{aligned} E_1(\mathbf{s}, \mathbf{x}) &= \{\exists j \text{ s.t. } (\mathbf{s}, \mathbf{x}, \mathbf{u}_{\ell, k, j}) \in T_\varepsilon(P_{S, X_1, U})\} \\ E_2(\mathbf{x}) &= \{(\mathbf{x}, \mathbf{y}) \notin T_\varepsilon(P_{X_1, Y})\} \\ E_3 &= \{\exists \ell' \neq \ell \text{ s.t. } (\mathbf{x}_{\ell'}, \mathbf{y}) \in T_\varepsilon(P_{X_1, Y})\} \\ E_4(\mathbf{s}) &= \{(\mathbf{u}_{\ell, k, j(s, \ell, k)}, \mathbf{y}) \notin T_\varepsilon(P_{U, Y})\} \\ E_5(\mathbf{x}) &= \{\exists k' \neq k, j', \text{ s.t. } (\mathbf{x}, \mathbf{u}_{\ell, k', j'}) \in T_\varepsilon(P_{X_1, U, Y})\}. \end{aligned}$$

One can easily realize as an immediate extension of [2] that $\Pr(E_1(\mathbf{s}, \mathbf{x}))$ behaves essentially like $[1 - 2^{-n(I(S;U|X)+\varepsilon)}]^J \leq \exp(2^{-n\varepsilon})$. Given that $E_1(\mathbf{s}, \mathbf{x})$ does not occur, we have that (\mathbf{x}, \mathbf{s}) are jointly typical with the output of the informed encoder $\tilde{\mathbf{x}}$ and $\mathbf{u}_{\ell, k, j(s, \ell, k)}$, that is, $(\mathbf{s}, \mathbf{x}, \mathbf{u}_{\ell, k, j(s, \ell, k)}, \tilde{\mathbf{x}}) \in T_\varepsilon(Q_S \times P_{X_1} \times P_{U, X_2|S, X_1})$.

For $\mathbf{s}, \mathbf{x}, \mathbf{u}_{\ell, k, j(s, \ell, k)}, \tilde{\mathbf{x}}$ jointly typical, the probabilities of the events $E_2(\mathbf{x}), E_4(\mathbf{s})$ vanish also due to the AEP.

Further, using the union bound, it is easily argued that for all $\delta > 0$ there exists n sufficiently large such that,

$$\begin{aligned} \Pr(E_3) &\leq L2^{-n(I(X;Y)+\varepsilon)} + \delta \leq 2^{-2n\varepsilon} + \delta \\ \Pr(E_5(\mathbf{x})) &\leq KJ2^{-n(I(U;Y|X_1)+\varepsilon)} + \delta \leq 2^{-2n\varepsilon} + \delta \end{aligned} \quad (57)$$

Taking the limit of $\delta \rightarrow 0$ yields the desired result.

REFERENCES

- [1] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, pp. 289–293, 1958.
- [2] S. Gelfand and M. Pinsker, "Coding for channels with random parameters," *Problems of control and information theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [3] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. IT-29, pp. 439–441, May 1983.
- [4] A. Cohen and A. Lapidoth, "Generalized writing on dirty paper," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'02)*, (Lausanne, Switzerland), 2002.
- [5] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, pp. 1250–1276, June 2002.
- [6] S. Gelfand and M. Pinsker, "On Gaussian channel with random parameters," in *6th International Symposium on Information Theory*, (Tashkent), 1984.
- [7] Y. H. Kim, A. Sutivong, and S. Sigurjónsson, "Multiple user writing on dirty paper," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT'04)*, (Chicago, IL, USA), 2004.
- [8] Y. Steinberg and S. Shamai (Shitz), "Achievable rates of the broadcast channel with states known at the transmitter," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'05)*, (Adelaide, Australia), 2005.
- [9] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *IEEE Transactions on Information Theory*, vol. 51, pp. 2867–2877, August 2005.
- [10] S. Sigurjónsson and Y. H. Kim, "On multiple user channels with causal state information at the transmitters," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'05)*, (Adelaide, Australia), pp. 72–76, 2005.
- [11] A. Sutivong, M. Chiang, T. M. Cover, and Y.-H. Kim, "Channel capacity and state estimation for state-dependent Gaussian channels," *IEEE Transactions on Information Theory*, vol. IT-51, pp. 1486–1495, April 2005.
- [12] N. Merhav and S. Shamai (Shitz), "Information rates subjected to state masking," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'06)*, (Seattle, WA, USA), July 2006.
- [13] N. Merhav and S. Shamai (Shitz), "On joint sourcechannel coding for the WynerZiv source and the GelfandPinsker channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 2844–2855, November 2003.
- [14] A. Somekh-Baruch and N. Merhav, "On the random coding error exponents of the single-user and the multiple-access Gelfand-Pinsker channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'04)*, (Chicago, IL, USA), p. 448, June-July 2004.
- [15] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, pp. 563–593, March 2003.
- [16] A. S. Cohen and A. Lapidoth, "The Gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, pp. 1639–1667, June 2002.
- [17] A. Somekh-Baruch and N. Merhav, "On the capacity game of public watermarking systems," *IEEE Transactions on Information Theory*, vol. 50, pp. 511–524, March 2004.
- [18] N. Merhav, "On joint coding for watermarking and encryption," *IEEE Transactions on Information Theory*, vol. 52, pp. 190–205, January 2006.
- [19] A. Maor and N. Merhav, "On joint information embedding and lossy compression in the presence of a stationary memoryless attack channel," *IEEE Transactions on Information Theory*, vol. 51, pp. 3166–3175, September 2005.
- [20] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [21] A. Høst-Madsen, "On the capacity of cooperative diversity in slow fading channels," in *Proc. Allerton Conf. Communications, Control, and Computing*, (Monticello, IL, USA), October 2002.
- [22] S. Kotagiri and J. N. Laneman, "Achievable rates for multiple access channels with state information known at one encoder," in *Proc. Allerton Conf. Communications, Control, and Computing*, (Monticello, IL, USA), October 2004.
- [23] S. Kotagiri and J. N. Laneman, "Multiple access channels with state information known to some encoders," preprint, 2006.
- [24] T. Weissman and N. Merhav, "Coding for the feedback Gelfand-Pinsker channel and the feedforward Wyner-Ziv source," *IEEE Transactions on Information Theory*, vol. 52, pp. 4207–4211, September 2006.
- [25] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1813–1827, 2006.
- [26] N. Devroye, P. Mitran, and V. Tarokh, "Limits on communications in a cognitive radio channel," *IEEE Communications Magazine*, vol. 44, pp. 44–49, June 2006.
- [27] W. Wu, S. Vishwanath, and A. Arapostathis, "On the capacity of Gaussian weak interference channels with degraded message sets," in *Conference on Information Sciences and Systems (CISS2006)*, (Princeton, NJ, USA), March 2006.
- [28] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective," preprint, 2006.
- [29] A. Høst-Madsen, "Capacity bounds for cooperative diversity," *IEEE Transactions on Information Theory*, vol. 52, pp. 1522–1544, April 2006.
- [30] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, "Cooperative encoding with asymmetric state information at the transmitters," In preparation, 2006.
- [31] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.