CCIT Report #653 November 2007

Cooperative Multiple Access Encoding with States Available at One Transmitter

Anelia Somekh-Baruch Shlomo Shamai (Shitz) Sergio Verdú

Abstract

We generalize the Gel'fand-Pinsker model to encompass the setup of a memoryless multiple-access channel. According to this setup, only one of the encoders knows the state of the channel (non-causally), which is also unknown to the receiver. Two independent messages are transmitted: a common message and a message transmitted by the informed encoder. We find an explicit characterization of the capacity region of this channel. An explicit characterization of the capacity region 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, under an individual power constraint as well as a sum power constraint. In this case, the capacity region 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

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

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], where dirty-paper coding happens to be a central ingredient in achieving the capacity region, 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) Gaussian setup of [6] enables interference cancelation and the capacity region is characterized fully, here, only inner and outer bounds on the capacity region of the additive white Gaussian MAC as well as the general discrete channel are derived. In the asymmetric case, even in the Gaussian model the capacity region, yet unknown, is degraded in general as compared to the no interference case. The inner bound on the capacity region is attained by a generalized dirty paper coding (DPC) scheme used by the informed encoder, that allows arbitrary correlation between the codeword and the known CSI. Another paper in which asymmetric CSI at the transmitters is studied is [24], where full CSI at the decoder is assumed.

In this paper, which generalizes a former conference version [25], 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 is that of two users transmitting a common message, and the informed encoder transmitting a private message. We refer to this channel as a Generalized GP (GGP) channel. We characterize the capacity region 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 characterization remains the same even if one allows feedback at the informed encoder, but not at the uninformed encoder (similarly to the single-user GP channel setting [26]). While feedback does not increase the ultimate rate, it simplifies considerably the signalling technique which is capable of approaching capacity [26]. We specialize the expression of the capacity region to the case where it is only a common message that is being transmitted. Transmission of a single common message source can be regarded as a single-user channel, in the sense that there is one message source and one destination. The capacity in this case is referred to as a *common message capacity*. 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 under both an individual power constraint and a sum power constraint. 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 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 outer bound on the achievable

rate pairs and point out the loss due to the asymmetric side information. The resulting outer bound is shown to be achievable in the Gaussian case, yielding a closed-form expression for the capacity region. We further specialize the results to the common message capacity case, and we characterize the optimal strategy of the informed encoder balancing the tradeoff 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. We point out the optimal power allocation between the two users when the sum power constraint is concerned. Another interesting insight derived from the proof is the capacity region of a class of finite alphabet and Gaussian parallel channels with non-causal side information at the transmitter.

A Cognitive Radio (see, [27], [28], [29], [30]) 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 made in the models considered in [27], [30] 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) helps the licensed user to transmit the message, exploiting its side information, and transmits an additional message. 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 covertext, and the second embeds information in a covertext-dependent method. Our work accounts also for other scenarios of cooperative communication used to increase performance [21], [31], and this will be detailed in [32].

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 region in the discrete case, an outer bound on the capacity region, and the capacity region 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 region 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 expression. Section VI concludes with a summary of the main contributions of this paper.

II. Notation and Problem SETUP

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 $(x_i, x_{i+1}, ..., x_j)$, and $x^n = (x_1, ..., x_n)$. 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 $P_{Y|X}$. The alphabet of a scalar RV, X, will be designated by the

corresponding caligraphic letter $\mathcal{X}$. The set of probability distributions defined on an alphabet $\mathcal{X}$, will be denoted by $\mathcal{P}(\mathcal{X})$. The cardinality of a set $\mathcal{A}$ will be denoted by $|\mathcal{A}|$.

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 $X_{(1)}^n=(X_1(1),...,X_1(n))$ and $X_{(2)}^n=(X_2(1),...,X_2(n))$ designate the inputs of transmitters 1 and 2 to the channel, respectively. The output of the channel is denoted by Y^n. The stationarity and memorylessness assumptions imply that

$$P_{Y^n|S^n,X_{(1)}^n,X_{(2)}^n}(y^n|s^n,x^n,\tilde{x}^n)=\prod_{i=1}^{n}W_{Y|S,X_1,X_2}(y_i|s_i,x_i,\tilde{x}_i).$$

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 is Figure 1, the setup we consider is asymmetric in the sense that only encoder 2 is informed of the channel states, while neither the other encoder nor the decoder know the channel states. Unlike the ordinary MAC, with partially known state information, we allow a common message source fed to both encoders, and an independent message that is to be transmitted by the informed encoder. When 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

$$W_{Y_1,Y_2|S,X_1,X_2} \;=\; W_{Y_1|X_1,S}W_{Y_2|X_2,S}. \tag{1}$$

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

$$W_{Y_2|X_2,S} \;=\; W_{Y_2|X_2} \tag{2}$$

we shall say that the parallel channel is degenerate.

The common message, $\mathcal{W}_c$, and the private message, $\mathcal{W}_2$, are independent random variables uniformly distributed over the sets $\{1,...,M_c\}$ and $\{1,...,M_2\}$, respectively, where $M_c=\lfloor e^{nR_c}\rfloor$ and $M_2=\lfloor e^{nR_2}\rfloor$. An (e^{R_c},e^{R_2},n)-code for the GGP channel consists of two encoders $\varphi_n^{(1)},\varphi_n^{(2)}$ and a decoder ψ_n: the first encoder, unaware of the CSI is defined by a mapping

$$\varphi_n^{(1)}:\ \{1,...,M_c\}\to\mathcal{X}_1^n. \tag{3}$$

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

$$\varphi_n^{(2)}:\ \{1,...,M_c\}\times\{1,...,M_2\}\times\mathcal{S}^n\to\mathcal{X}_2^n. \tag{4}$$

The decoder is a mapping

$$\psi_n:\ \mathcal{Y}^n\to\{1,...,M_c\}\times\{1,...,M_2\}. \tag{5}$$

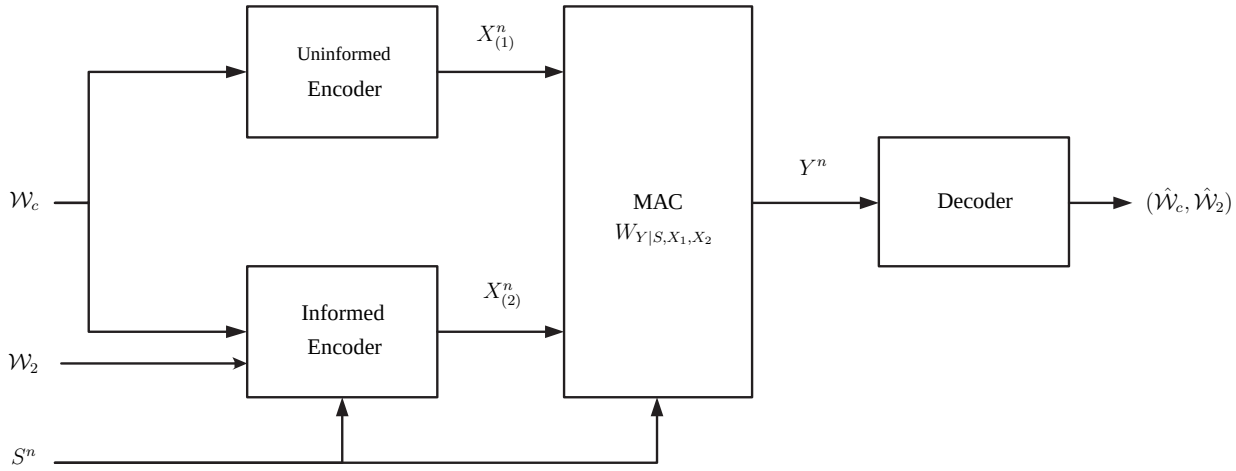

Fig. 1. Asymmetric state-dependent MAC with a common message.

An (e^{R_c}, e^{R_2}, n)-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

$$\varphi^{(2)}_{n,i} : \{1, ..., M_c\} \times \{1, ..., M_2\} \times \mathcal{S}^i \to \mathcal{X}_2 \quad i = 1, ..., n, \tag{6}$$

and at time index i, the channel input is given by $X_2(i) = \varphi^{(2)}_{n,i}(\mathcal{W}_c, \mathcal{W}_2, S^i)$.

An (ϵ, n, R_c, R_2)-code for the GGP channel is a code $(\varphi^{(1)}_n, \varphi^{(2)}_n, \psi_n)$ having average probability of error not exceeding ϵ, i.e.,

$$\Pr\left((\mathcal{W}_c, \mathcal{W}_2) \neq \psi_n(Y^n_1)\right) \leq \epsilon. \tag{7}$$

A rate pair (R_c, R_2) is said to be achievable if there exists a sequence of $(\epsilon_n, n, R_c, R_2)$-codes with $\lim_{n\to\infty} \epsilon_n = 0$. The capacity region of the GGP channel is defined as the closure of the set of achievable (R_c, R_2) rate pairs. The definitions of an (ϵ, n, R_c, R_2)-code, an achievable rate pair and the capacity region of the asymmetric causal state-dependent channel are similar.

III. Capacity Region - Finite Input Alphabet GGP Channel

The following theorem provides a single-letter expression for the capacity region 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 region of the finite input alphabet GGP channel, $\mathcal{C}$, is the closure of all rate pairs, (R_c, R_2), satisfying*

$$\begin{aligned} R_2 &\leq I(U;Y|X_1) - I(U;S|X_1) \\ R_c + R_2 &\leq I(U, X_1; Y) - I(U, X_1; S), \end{aligned} \tag{8}$$

for some joint measure $P_{S,X_1,U,X_2,Y}$ *on* $\mathcal{S}\times\mathcal{X}_1\times\mathcal{U}\times\mathcal{X}_2\times\mathcal{Y}$ *having the form*

$$P_{S,X_1,U,X_2,Y}=Q_S P_{X_1} P_{U|S,X_1} P_{X_2|S,X_1,U} W_{Y|S,X_1,X_2}, \tag{9}$$

where

$$|\mathcal{U}| \;\leq\; |\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|. \tag{10}$$

The proof of Theorem 1 appears in Appendix A, the direct part of the proof involves error probability analysis of a coding scheme and is described in the sequel (after Corollary 2). The proof is a quite straightforward extension of its GP counterpart [2] and an immediate extension of the proof of Theorem 1 in [25].

It is noted that Theorem 1 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 $\varphi_n^{(2)}=\{\varphi_n^{(2,i)}\}_{i=1}^n$ with

$$\varphi_n^{(2,i)}:\ \{1,...,M_c\}\times\{1,...,M_2\}\times\mathcal{S}^n\times\mathcal{Y}^{i-1}\to\mathcal{X}_2. \tag{11}$$

It is easily verified that for the case of a channel which does not depend on the states, i.e., $W_{Y|S,X_1,X_2}=W_{Y|X_1,X_2}$, the expression for the capacity region reduces to the collection of rate pairs (R_c,R_2) satisfying

$$\begin{aligned} R_2 &\leq I(X_2;Y|X_1,Z),\\ R_c+R_2 &\leq I(X_1,X_2;Y), \end{aligned} \tag{12}$$

for some $P_{Z,X_1,X_2,Y}=P_Z P_{X_1|Z} P_{X_2|Z} W_{Y|X_1,X_2}$. This expression coincides with that of [33], which studies a MAC with two private messages $\mathcal{W}_1,\mathcal{W}_2$ and a common message [33], degenerated to the case of no message $\mathcal{W}_1$ ($R_1=0$).

We now specialize Theorem 1 to the important case where only the common message is transmitted.

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

$$C=\max\left[I(U,X_1;Y)-I(U,X_1;S)\right], \tag{13}$$

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*

$$P_{S,X_1,U,X_2,Y}=Q_S P_{X_1} P_{U,X_2|S,X_1} W_{Y|S,X_1,X_2}, \tag{14}$$

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

Corollary 1 follows from Theorem 1 by relaxing the constraint on R_2 in (8). Also, there exists a maximizing measure with X_2 that is a deterministic function of (S,X_1,U). The following corollary provides an alternative expression for C.

Corollary 2 *The common message capacity of the finite input alphabet GGP channel is given by*

$$C = \max\left[I(Z;Y) - I(Z;S)\right], \tag{15}$$

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

$$P_{S,X_1,Z,X_2,Y} = Q_S P_{X_1} P_{Z,X_2|S,X_1} W_{Y|S,X_1,X_2}, \tag{16}$$

and X_1 is a deterministic function of Z. The alphabet cardinality of Z satisfies $|\mathcal{Z}| \leq |\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|+1$.

We note that the condition that X_1 is a deterministic function of Z can be replaced by $X_1 \leftrightarrow Z \leftrightarrow S$. The proof appears in Appendix B.

We now give a description of a random coding scheme that is used to prove the achievability part of Theorem 1. It is based on the following principle: For a measure P_{S,X_1,U,X_2} satisfying (9), the uninformed user transmits at rate $R_1 = I(X_1;Y)$. Now the informed user can transmit using a Gel'fand Pinsker-like scheme, at rate $R_2' = I(U;Y,X_1) - I(U;S,X_1) = I(U;Y|X_1) - I(U;S|X_1)$, and that is by treating X_1 as part of the state, and accounting for the fact that X_1 is available at the decoder already as the information send by the uninformed encoder has been decoded first. Now the information sent by the informed encoder at rate R_2' can be shared between the private message $\mathcal{W}_2$ and the common message $\mathcal{W}_c$.

Encoding:

Fix a measure $P_{S,X_1,U,X_2,Y}$ satisfying (9). The following scheme is used to show (see Appendix A.2) that the rate pair

$$\begin{aligned} R_2^* &= I(U;Y|X_1) - I(U;S|X_1) \\ R_c^* &= I(X_1;Y) - I(X_1;S) = I(X_1;Y) \end{aligned} \tag{17}$$

is achievable.

Denote

$$\begin{aligned} M_1 &= e^{n[I(X_1;Y)-\epsilon]} \\ M_2 &= e^{n[I(U;Y|X_1)-I(U;S|X_1)-\epsilon]} \\ J &= e^{n[I(U;S|X_1)+2\epsilon]}. \end{aligned} \tag{18}$$

The random encoders operate as follows: The uninformed encoder draws M_1 i.i.d. vectors, $\{\mathbf{x}_\ell\}_{\ell=1}^{M_1}$, 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, $\mathbf{x}_\ell$, the informed encoder draws $M_2 \times J$ auxiliary vectors, denoted $\{\mathbf{u}_{\ell,k,j}\}$, $k = 1,...,M_2$, $j = 1,...,J$, independently and with i.i.d. components given $\mathbf{x}_\ell$. Hence, each codeword in the uninformed user codebook is associated with a codebook of auxiliary codewords.

To transmit ℓ, the uninformed encoder transmits the vector $\mathbf{x}_\ell$. Transmission of k is done by the informed encoder who searches for the lowest $j_0 \in \{1,...,J\}$ such that $\mathbf{u}_{\ell,k,j_0}$ is jointly

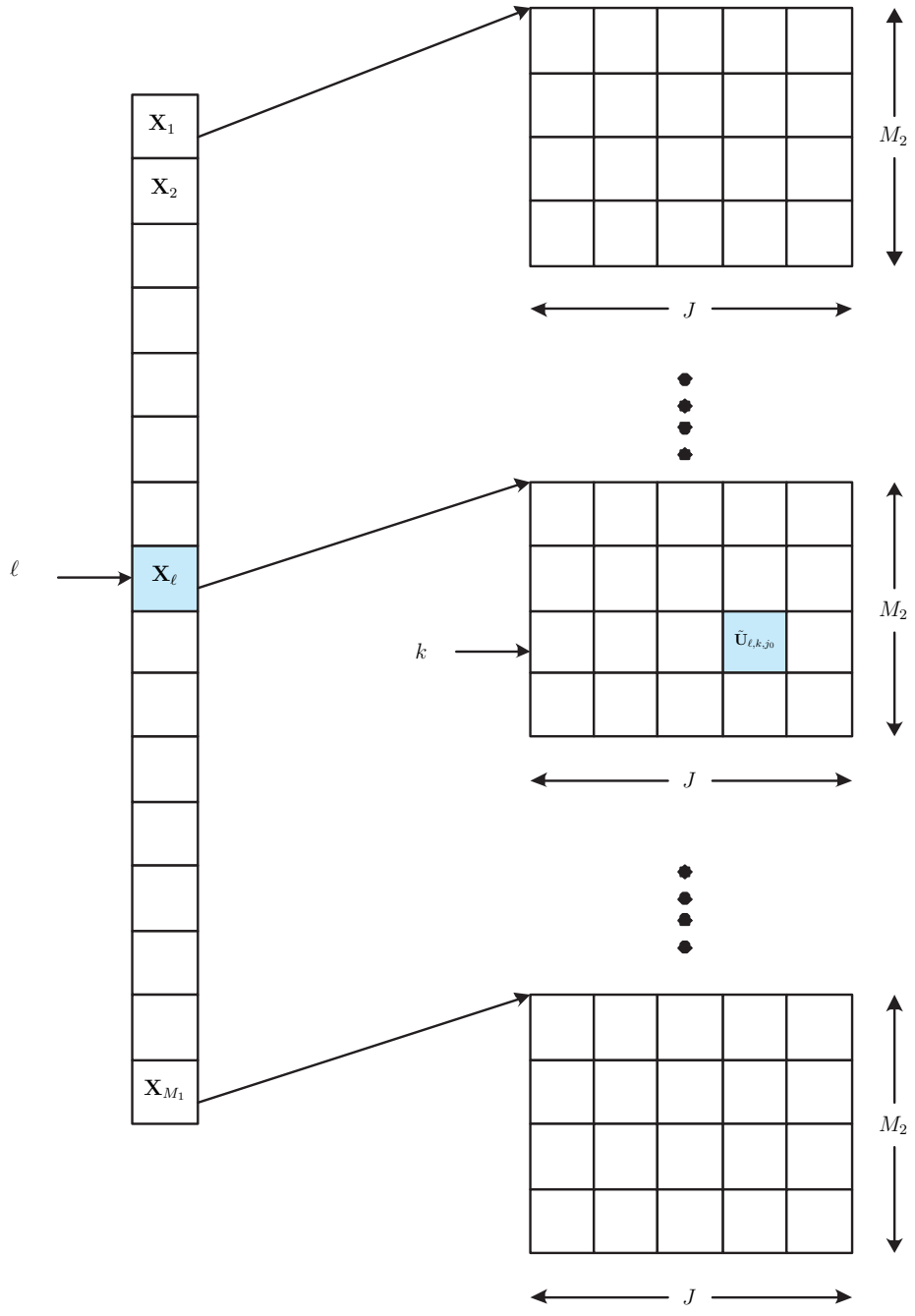

Fig. 2. A generalized binning coding scheme

typical with $(\mathbf{x}_\ell, \mathbf{s})$. Denote this j by $j(\mathbf{s}, \ell, k)$. If such j_0 is not 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 a vector $\tilde{\mathbf{x}}$ that is drawn i.i.d. conditionally given $(\mathbf{s}, \mathbf{u}_{\ell,k,j(\mathbf{s},\ell,k)}, \mathbf{x}_\ell)$ (using conditional measure that is the appropriate marginal of $\mathcal{Q}_S P_{X_1} P_{U,X_2|X_1,S}$).

Decoding: Upon observing $\mathbf{y}$, the decoder searches for a pair of indices, $(\hat{\ell}, \hat{k})$, such that $\mathbf{x}_{\hat{\ell}}, \mathbf{u}_{\hat{\ell},\hat{k},j}$ are jointly typical with $\mathbf{y}$ and outputs them. 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 A.2 establishing the achievability of the rate pair (R_2^*, R_c^*) (17). As mentioned earlier, the proof of the converse part of Theorem 1 can be found in Appendix A.1.

The following claim completes the proof of the direct part of Theorem 1.

Claim 1 *If (R_2, R_c) is achievable, then so is $(0, R_c + R_2)$.*

Proof: Bits that are attributed to $\mathcal{W}_2$ can instead be attributed to $\mathcal{W}_c$. ■

This proves that also $(0, R_2^* + R_c^*)$ is achievable and thus the entire trapezoid (8) is an achievable region.

It should be noted, that when the common message capacity is concerned, the above encoding scheme can be applied by attributing the bits assigned to $\mathcal{W}_2$ in the above described scheme, to the common message $\mathcal{W}_c$. The output of decoder is the pair $\hat{m} = (\hat{\ell}, \hat{k})$.

We now state a theorem that provides an outer bound on the capacity region of the GGP channel. It is a generalization of the trivial bound $\max_{P_{X|S}} I(X;Y|S)$ on the capacity of the ordinary single-user GP channel. This theorem 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.

Theorem 2 *The closure of the set of rate pairs satisfying*

$$\begin{aligned} R_2 &\leq I(X_2; Y|S, X_1) \\ R_c + R_2 &\leq I(X_1, X_2; Y|S) - I(S; X_1|Y) \end{aligned} \tag{19}$$

for some measure $P_{S,X_1,X_2,Y} = Q_S P_{X_1} P_{X_2|S,X_1} W_{Y|S,X_1,X_2}$ is an outer bound on the capacity region of the GGP channel.

Proof: Recall that the capacity region of the finite input alphabet GGP channel is given by (8). Now,

$$\begin{aligned} & I(U, X_1; Y) - I(U, X_1; S) \\ = \;& I(U, X_1; Y|S) - I(U, X_1; S|Y) \\ \leq \;& I(X_1, X_2; Y|S) - I(U, X_1; S|Y) && (20) \\ = \;& I(X_1, X_2; Y|S) - I(S; X_1|Y) - I(U; S|X_1, Y) \\ \leq \;& I(X_1, X_2; Y|S) - I(S; X_1|Y), && (21) \end{aligned}$$

where the first inequality holds since $U \leftrightarrow (X_2, X_1, S) \leftrightarrow Y$ is a Markov chain.

Further,

$$\begin{aligned} & I(U;Y|X_1) - I(U;S|X_1) \\ = \; & I(U;Y|X_1,S) - I(U;S|X_1,Y) \\ \leq \; & I(X_2;Y|S,X_1), \end{aligned} \tag{22}$$

which concludes the proof of Theorem 2. ■

In the following theorem, we find the capacity region of the degenerate parallel GGP channel, for which we establish the fact that the CSI does not help.

Theorem 3 *The capacity region of the degenerate parallel GGP channel is equal to the capacity region obtained without transmitter CSI, i.e.,*

$$\begin{aligned} R_2 &\leq C_2 \\ R_c + R_2 &\leq C_1 + C_2, \end{aligned} \tag{23}$$

where C_1 is the capacity of the channel $W_{Y_1|X_1,S}$ obtained without transmitter CSI, and C_2 is the capacity of the channel $W_{Y_2|X_2}$.

The proof of Theorem 3 appears in Appendix C.

IV. The Causal Asymmetric State-Dependent Channel

In this section we consider the causal asymmetric state-dependent channel (see (6)).

Theorem 4 *The capacity region of the finite input alphabet causal asymmetric state-dependent channel is given by the closure of the set of rate pairs (R_2, R_c) satisfying*

$$\begin{aligned} R_2 &\leq I(U;Y|X_1) \\ R_c + R_2 &\leq I(U, X_1; Y), \end{aligned} \tag{24}$$

for some joint measure $P_{S,X_1,U,X_2,Y}$ on $\mathcal{S} \times \mathcal{X}_1 \times \mathcal{U} \times \mathcal{X}_2 \times \mathcal{Y}$ having the form

$$P_{S,X_1,U,X_2,Y} = Q_S P_{X_1,U} P_{X_2|S,X_1,U} W_{Y|S,X_1,X_2}, \tag{25}$$

where $|\mathcal{U}|$ satisfies

$$|\mathcal{U}| \leq |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1. \tag{26}$$

The proof can be found in Appendix D.

The expression for the capacity region of Theorem 4 can be interpreted as a special case of Theorem 1, where U is independent of S. This is similar to the relation between the expression for the capacity of state dependent channel with causal CSI introduced by Shannon [1], and its non-causal counterpart, the Gel'fand-Pinsker [2] channel [34].

Specializing Theorem 4 to the case where there is only a transmission of a common message, we get the following.

Corollary 3 *The common message capacity of the finite input alphabet causal asymmetric state-dependent channel is given by*

$$C_{causal} = \max I(U;Y), \tag{27}$$

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

$$P_{S,X_1,U,X_2,Y} = Q_S P_U P_{X_1|U} P_{X_2|S,U,X_1} W_{Y|S,X_1,X_2}, \tag{28}$$

where X_1 is a deterministic function of U, and $|\mathcal{U}| \leq |\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|+2$.

We note that an alternative expression for C_{causal} with $I(U,X_1;Y)$ replacing $I(U;Y)$ in (27), and where X_1 does not have to be deterministic given U with $|\mathcal{U}| \leq |\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|+1$ also holds. The proof is similar to that of Theorem 4 and thus is omitted. As a side note, we mention that if the CSI is available to both of the encoders, the single-letter expression for the common message capacity is deduced as a direct application of the formula derived by Shannon [1] for a channel with input alphabet $\mathcal{X}_1\times\mathcal{X}_2$.

V. The Gaussian GGP Channel

In this section we analyze the additive Gaussian GGP channel. In Subsection V-A we present the channel model, and the power constraints that we analyze (individual and sum power constraints). Based on the results obtained in Section III, we derive a closed-form formula for the capacity region under individual power constraints in Subsection V-B, and discuss it. The results are then specialized in Subsection V-C to the common message capacity under individual power constraints, where we also provide several numerical results. We conclude this section in Subsection V-D where sum power constraints are addressed.

A. Channel Model

The Gaussian GGP channel is given by

$$Y_i \;=\; X_1(i) + X_2(i) + S_i + N'_i. \tag{29}$$

As before, the message $\mathcal{W}_c$ is available to both encoders. Only the second encoder knows the realization of the interference S^n (non-causally), and the message $\mathcal{W}_2$ to be transmitted. The noise processes, S^n and N'^n, are assumed to be zero-mean Gaussian i.i.d. with $E(S_i^2)=Q$ and $E(N'^2_i)=N$. The process N'^n is independent of $\left(X^n_{(1)}, X^n_{(2)}, S^n\right)$. Several power constraints can be considered:

a) Individual power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} X_1^2(i) \leq P_1 \;,\; \frac{1}{n}\sum_{i=1}^{n} X_2^2(i) \leq P_2. \tag{30}$$

b) Sum power constraint:

$$\frac{1}{n}\sum_{i=1}^{n} X_1^2(i) + \frac{1}{n}\sum_{i=1}^{n} X_2^2(i) \leq P. \tag{31}$$

c) Total received power constraint:

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_1(i)+X_2(i)\right)^2 \leq P, \tag{32}$$

where here it is evident that all the power should be assigned to the informed encoder, and the problem degenerates to the ordinary "dirty paper" Costa setup [3], where the informed transmitter can assign bits of the transmitted information to either $\mathcal{W}_c$ or $\mathcal{W}_2$, that is, the capacity region in this case is a triangle whose vertex (R_c, R_2) points are $(0,0)$, $\left(0, \frac{1}{2}\log\left(1+\frac{P}{N}\right)\right)$, and $\left(\frac{1}{2}\log\left(1+\frac{P}{N}\right), 0\right)$.

We are interested in finding the capacity regions for the individual power constraint and the sum power constraint. To this end, using standard techniques [35], an application of the single-letter expression derived for the finite alphabet case to the Gaussian GGP channel $W_{Y|S,X_1,X_2}(y|s,x,x') = \frac{1}{\sqrt{2\pi N}}e^{-(y-s-x-x')^2/2N}$ with the individual power constraint gives the capacity region $\mathcal{C}(P_1, P_2, Q, N)$ as the closure of the union of rate pairs (R_c, R_2) satisfying (8) where the allowed joint distribution of S, X_1, U, X_2, Y satisfies (9), and

$$E(X_1^2) \leq P_1,\ E(X_2^2) \leq P_2. \tag{33}$$

When the sum power constraint is considered, the expression for the capacity region, denoted $\mathcal{C}(P, Q, N)$, remains the same with the exception that (33) is replaced with $E(X_1^2)+E(X_2^2) \leq P$.

B. Capacity Region under Individual Power Constraints

Before establishing the capacity region of the GGP channel under individual power constraints, we provide an outer bound which takes on a very simple form. Then, we establish the tightness of this bound for a certain range of rates. In order to present the outer bound we need the following definition of a special case of a degenerate parallel channel (see (1)-(2)).

Definition 1 *A Gaussian degenerate parallel channel with non-causal asymmetric CSI is a GGP channel whose i-th output is given by $Y_i = (Y_1(i), Y_2(i))$ with*

$$\begin{aligned} Y_1(i) &= X_1(i) + S_i \\ Y_2(i) &= X_2(i) + N_i', \end{aligned} \tag{34}$$

where S^n and N'^n are i.i.d. Gaussian independent noise processes.

Theorem 5 *The capacity region of the Gaussian degenerate parallel channel with non-causal asymmetric CSI under individual power constraints is given by the set of rate pairs (R_c, R_2) satisfying*

$$R_2 \leq \frac{1}{2}\log\left(1+\frac{P_2}{N}\right) \tag{35}$$

$$R_c + R_2 \leq \frac{1}{2}\log\left(1+\frac{P_1}{Q}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{N}\right). \tag{36}$$

Proof: This is an immediate consequence of Theorem 3 applied to the Gaussian case. ■
The capacity region of this degenerate parallel channel contains that of the GGP channel, as the decoder has more information, $(Y_1(i), Y_2(i))$ rather than $Y(i) = Y_1(i) + Y_2(i)$, and yields the following outer bound to $\mathcal{C}(P_1, P_2, Q, N)$.

Corollary 4 *$C(P_1, P_2, Q, N)$ is contained in the set of rate pairs (R_c, R_2) satisfying (35) and (36).*

The following theorem provides an explicit characterization for the capacity region of this channel for the individual power constraints.

Theorem 6 *The capacity region $\mathcal{C}(P_1, P_2, Q, N)$ of the Gaussian GGP channel under individual power constraints is given by the union of the rate pairs satisfying*

$$\begin{aligned} R_2 &\leq \frac{1}{2}\log\left(1 + \frac{P_2(1-\rho_{12}^2-\rho_{2s}^2)}{N}\right) \\ R_c + R_2 &\leq \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1}+\rho_{12}\sqrt{P_2}\right)^2}{P_2(1-\rho_{12}^2-\rho_{2s}^2)+\left(\sqrt{Q}+\rho_{2s}\sqrt{P_2}\right)^2+N}\right) \\ &\quad + \frac{1}{2}\log\left(1 + \frac{P_2(1-\rho_{12}^2-\rho_{2s}^2)}{N}\right) \end{aligned} \tag{37}$$

for some $\rho_{12} \in [0,1], \rho_{2s} \in [-1,0]$ such that

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{38}$$

We note that the expression for the capacity region can be simplified for certain ranges of rates, this will be done in the sequel (see Proposition 1), moreover, in Corollary 5 to follow, we specify the (ρ_{12}, ρ_{2s})-pairs that yield vertex points which lie on the border of the capacity region. The proof of Theorem 6 appears in Appendix E.

The following Corollary provides a more explicit characterization of the capacity region, it follows from Theorem 6, by substituting $\Delta = 1 - \rho_{12}^2 - \rho_{2s}^2$ and $\rho = \rho_{2s}$.

Corollary 5 *The capacity region of the Gaussian GGP channel under individual power constraints, $\mathcal{C}(P_1, P_2, Q, N)$, is given by the union of the rate pairs satisfying*

$$\begin{aligned} R_2 &\leq \frac{1}{2}\log\left(1 + \frac{P_2\Delta}{N}\right) \\ R_c + R_2 &\leq \max_{\rho\in[-\sqrt{1-\Delta},0]} \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1}+\sqrt{1-\Delta-\rho^2}\sqrt{P_2}\right)^2}{P_2\Delta+\left(\sqrt{Q}+\rho\sqrt{P_2}\right)^2+N}\right) + \frac{1}{2}\log\left(1+\frac{P_2\Delta}{N}\right) \end{aligned} \tag{39}$$

for some $\Delta \in [\Delta_{min}, 1]$, with $\Delta_{min} = \min\left\{0, 1 - \frac{P_1(P_2+N)^2}{P_2Q(P_1+Q)}\right\}$, where, if $\Delta_{min} > 0$, the maximizing ρ corresponding to Δ_{min} is $\rho = -\frac{P_1(P_2+N)}{\sqrt{QP_2}(P_1+Q)}$ and for $\Delta > \Delta_{min}$, the maximization

over ρ can be limited to either $\rho = -\sqrt{1-\Delta}$, $\rho = 0$ or any real root of $g_\Delta(\rho)$ that satisfies $\rho \in [-\sqrt{1-\Delta}, 0]$, with

$$\begin{aligned} g_\Delta(\rho) &= -P_2(P_1+Q)\rho^4 \\ &\quad -2\sqrt{QP_2}(P_2+Q+N+P_1)\rho^3 \\ &\quad +\left[-2P_2(1-\Delta)Q-(P_2+Q+N)^2+(1-\Delta)P_1P_2-P_1Q\right]\rho^2 \\ &\quad +2\sqrt{P_2Q}(1-\Delta)\left[P_1-(P_2+Q+N)\right]\rho \\ &\quad +(1-\Delta)Q\left[P_1-P_2(1-\Delta)\right]. \end{aligned} \tag{40}$$

The proof of Corollary 5 appears in Appendix F.

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 [2], were special in that the trivial operative upper bound is achievable. The upper bound of Theorem 2 plays the role of the operative bound and constitutes the core of the converse part. So, in fact, the generalization of interference cancelation to the GGP channel asymmetric setup is that the upper bound of Theorem 2 is achievable, a phenomenon that happens in the Gaussian GGP channel. Recalling (19), this implies that the subtracted term, $I(S;X_1|Y)$, can be interpreted as the inevitable 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 for the Gaussian channel, in (8), one can restrict attention to jointly Gaussian (S, X_1, X_2) without loss of generality, and an optimal choice for U is

$$U = X_2 + \alpha_{opt} S \tag{41}$$

with

$$\alpha_{opt} = \frac{P_2P_1Q - P_1\sigma_{2s}^2 - P_1N\sigma_{2s} - \sigma_{12}^2Q}{P_2P_1Q + P_1NQ - P_1\sigma_{2s}^2 - \sigma_{12}^2Q}, \tag{42}$$

where different values of $\sigma_{12} = E(X_1X_2)$ and $\sigma_{2s} = E(X_2S)$ are chosen to achieve different points that lie in (or, on the border of) the capacity region. 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'),

$$\Lambda_{X_1,X_2,S,N'} = \begin{pmatrix} P_1 & \sigma_{12} & 0 & 0 \\ \sigma_{12} & P_2 & \sigma_{2s} & 0 \\ 0 & \sigma_{2s} & Q & 0 \\ 0 & 0 & 0 & N \end{pmatrix} \tag{43}$$

satisfies the nonnegative-definiteness condition

$$\det\left(\Lambda_{X_1,X_2,S,N'}\right) = P_1(P_2QN - \sigma_{2s}^2N) - \sigma_{12}^2QN \geq 0, \tag{44}$$

i.e.,

$$P_1\sigma_{2s}^2 + Q\sigma_{12}^2 \leq P_1P_2Q, \tag{45}$$

or, in terms of correlation coefficients,

$$\rho_{12} \;=\; \frac{\sigma_{12}}{\sqrt{P_1P_2}},\ \rho_{12} = \frac{\sigma_{2s}}{\sqrt{P_2Q}}, \tag{46}$$

$$\rho_{12}^2 + \rho_{2s}^2 \leq 1. \tag{47}$$

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

Definition 2 *The set of parameters P_1, P_2, Q, N such that*

$$\frac{P_1(P_2+N)^2}{P_1+Q} \geq P_2Q \tag{48}$$

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

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

$$\begin{aligned} & P_1 \leq \frac{P_2Q^2}{(P_2+N)^2 - P_2Q} \\ \Leftrightarrow\quad & Q \geq -\frac{P_1}{2} + \frac{\sqrt{P_1(P_1P_2 + 4(N+P_2)^2)}}{2\sqrt{P_2}} \\ \Leftrightarrow\quad & N \leq \sqrt{\frac{P_2Q(P_1+Q)}{P_1}} - P_2. \end{aligned} \tag{49}$$

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

The following proposition simplifies the capacity region expression for certain ranges of rates, by setting values of vertex points. It indicates a certain range of rates for which the outer bound given in Theorem 5 is tight.

Proposition 1 *1. For any P_1, P_2, Q, N, the segment connecting the following points (a) and (b) in the $R_c - R_2$ plane lies on the boundary of $\mathcal{C}(P_1, P_2, Q, N)$*

(a) $\quad (R_c, R_2) = \left(0, \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)\right)$

(b) $\quad (R_c, R_2) = \left(\frac{1}{2}\log\left(1+\frac{P_1}{Q+P_2+N}\right), \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)\right)$.

2. If P_1, P_2, Q, N lie in the active regime, the segment connecting the following points (c) and (d) also lies on the boundary of $\mathcal{C}(P_1, P_2, Q, N)$

(c) $\quad (R_c, R_2) = \left(\frac{1}{2}\log\left(\frac{(P_1+Q)^2}{(Q(P_1+Q)-P_1(P_2+N))}\right), \frac{1}{2}\log\left(\frac{(P_2+N)(Q(P_1+Q)-P_1(P_2+N))}{(P_1+Q)NQ}\right)\right)$,

(d) $\quad (R_c, R_2) = \left(\frac{1}{2}\log\left(1+\frac{P_1}{Q}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{N}\right), 0\right)$.

TABLE I

EXTREME CASE ANALYSIS.

regime	Behavior of $\mathcal{C}(P_1, P_2, Q, N)$
$Q \to \infty$	$R_2 \le \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)$, $R_c + R_2 \le \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)$
$P_1 = 0$	$R_2 \le \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)$, $R_c + R_2 \le \frac{1}{2}\log\left(1+\frac{P_2}{N}\right)$
$Q = 0$	$R_2 \le \frac{1}{2}\log\left(1+\frac{P_2\Delta}{N}\right)$, $R_c + R_2 \le \frac{1}{2}\log\left(1+\frac{(\sqrt{P_1}+\sqrt{(1-\Delta)P_2})^2}{N}\right) + \frac{1}{2}\log\left(1+\frac{P_2\Delta}{N}\right)$
$P_2 = 0$	$R_2 = 0$, $R_c \le \frac{1}{2}\log\left(1+\frac{P_1}{N+Q}\right)$

Proof: It is easily verified that the line segment connecting (a) and (b) (first segment) and the one connecting (c) and (d) (second segment) both lie on the boundary of the outer bound specified in Corollary 4. It therefore remains to show that the points $(a), (b)$ stand for achievable rate pairs, and the points $(c), (d)$ are also achievable provided that P_1, P_2, Q, N lie in the active regime. To establish this claim, it suffices to substitute in (37), appropriate values of ρ_{12}, ρ_{2s} that will yields the desired segments. This is done by taking $\rho_{12} = \rho_{2s} = 0$ to get the first segment, and

$$\rho_{12}^* = \frac{P_1(P_2+N)}{\sqrt{P_1P_2}(P_1+Q)}, \quad \rho_{2s}^* = -\frac{P_1(P_2+N)}{\sqrt{QP_2}(P_1+Q)} \tag{50}$$

(which is a legitimate choice only when $\rho_{12}^2 + \rho_{2s}^2 \le 1$, as stated in (38) or in other words, $\frac{P_1(P_2+N)^2}{P_1+Q} \le P_2Q$) to get the second segment. ■

The capacity region for the parameters $(P_1, P_2, Q, N) = \left(\frac{1}{2}, 1, \frac{3}{2}, 1\right)$ which lies in the active regime, $\mathcal{C}(\frac{1}{2}, 1, \frac{3}{2}, 1)$, as well as the points $(a), (b), (c), (d)$ discussed in Proposition 1, and the outer bound of Corollary 4, are plotted in Figure 3. Each dotted trapezoids express achievable regions attained by choosing a specific Δ value in (39). The segment connecting (c) and (d) meets the R_c axis at -45^o, because it lies on the boundary of the outer bound of Theorem 5.

The points (a) and (b) discussed in Proposition 1 as well as the capacity region $\mathcal{C}(\frac{3}{2}, 1, \frac{1}{2}, 1)$ can be seen in Figure 4 for $P_1 = \frac{3}{2}, P_2 = N = 1, Q = \frac{1}{2}$ which lie in the silent regime. The figure also shows the corresponding outer bounds of Corollary 4.

Extreme Case Analysis

Table I summarizes the behavior of the capacity region $\mathcal{C}(P_1, P_2, Q, N)$ that can be deduced from Corollary 5 in several extreme cases. As expected, for infinite Q, the common message capacity degenerates to that of Costa's channel, that is, when the uninformed user is not present. The capacity region is triangular 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. The informed encoder therefore can decide which portion of its power, ΔP_1, to devote to transmission of its own message, and the remaining power, $(1-\Delta)P_1$, is allocated to coherent transmission (with the uninformed encoder) of the common message.

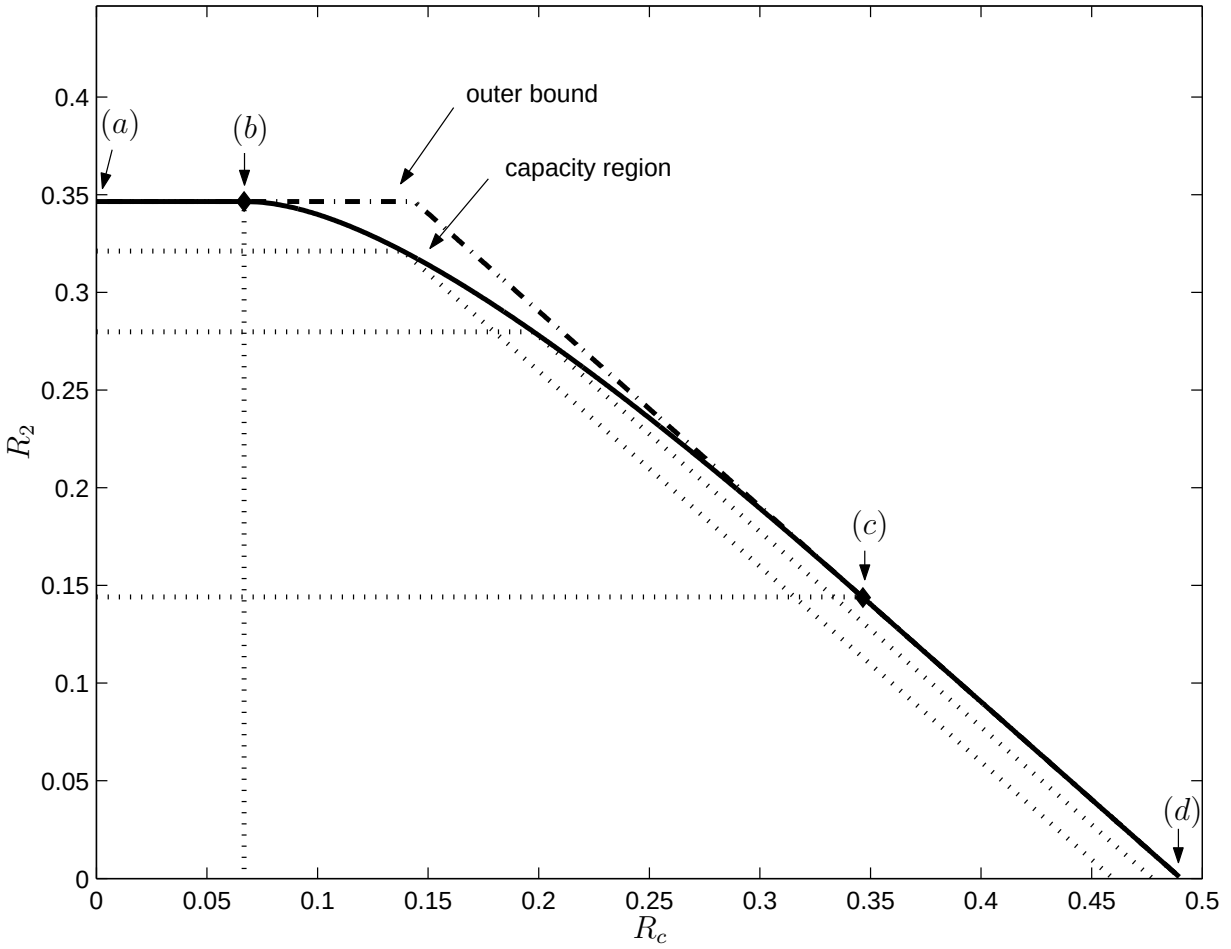

Fig. 3. Capacity region for $P_1 = \frac{1}{2}, P_2 = N = 1, Q = \frac{3}{2}$, $\mathcal{C}(\frac{1}{2}, 1, \frac{3}{2}, 1)$, and outer bound.

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 common message capacity is as though the effective noise is $S + N'$ and the power used for transmission is P_1.

C. The Common Message Capacity with Individual Power Constraints

This subsection is devoted to specializing the results pertaining to the Gaussian channel to the case where it is only a common message that is transmitted. The application of the tight upper bound of Theorem 2 (whose tightness in the Gaussian case was established in Theorem 6) to the common message capacity, yields that the common message capacity $C(P_1, P_2, Q, N)$ of the Gaussian GGP channel under individual power constraints is given by

$$\max_{\rho_{12}, \rho_{2s}} \left[I(X_1, X_2; Y|S) - I(X_1; S|Y) \right]. \tag{51}$$

where (X_1, X_2, S) are jointly Gaussian with $E(X_i^2) = P_i$, $i = 1, 2$.

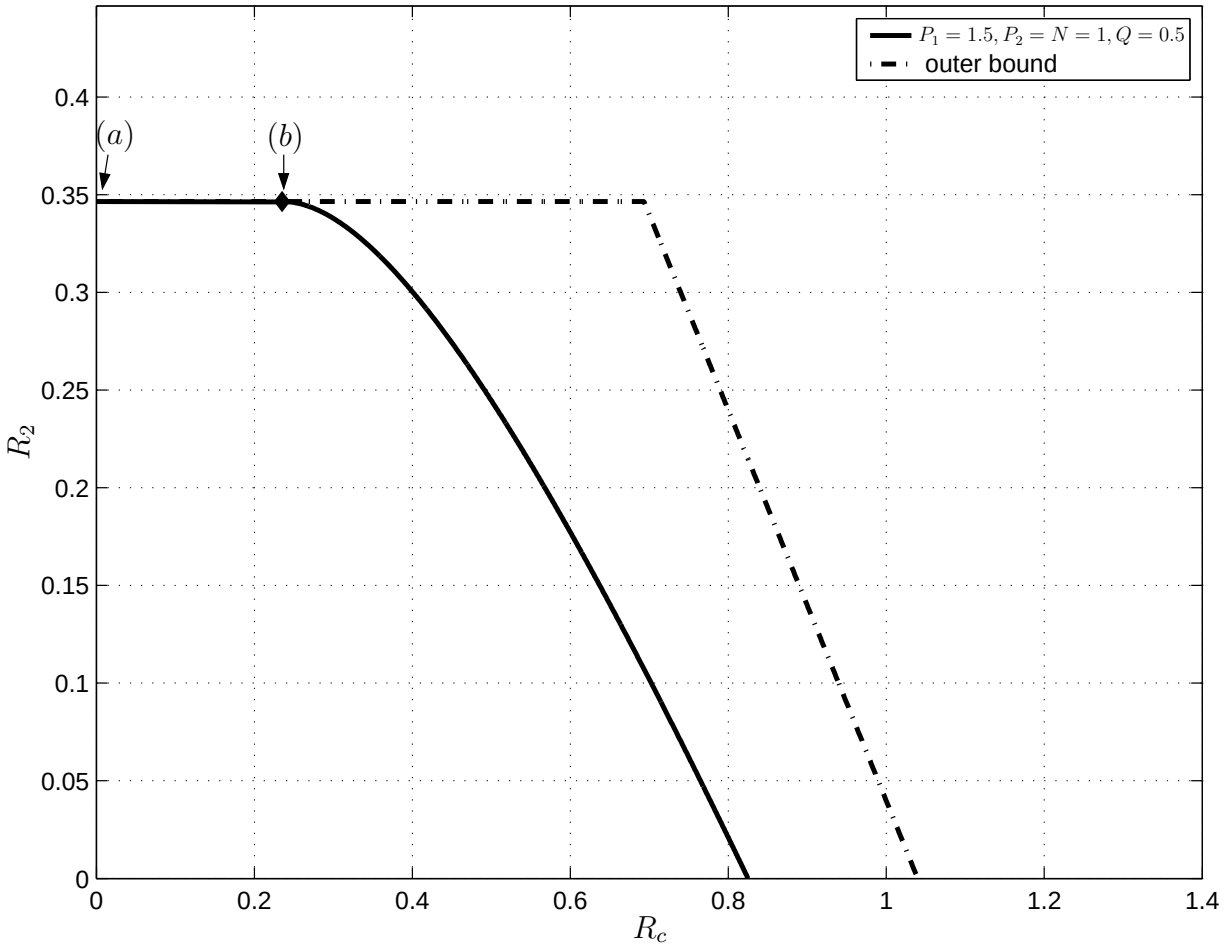

Fig. 4. Capacity region for $P_1 = \frac{3}{2}, P_2 = N = 1, Q = \frac{1}{2}$, $\mathcal{C}(\frac{1}{2}, 1, \frac{3}{2}, 1)$, and the outer bound.

Theorem 7 *The common message capacity of the Gaussian GGP channel under individual power constraints is given by the following formula*

$$C(P_1, P_2, Q, N) = \begin{cases} \frac{1}{2}\log\left(1+\frac{P_1}{Q}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{N}\right) & \text{if } \frac{P_1(P_2+N)^2}{(P_1+Q)} \leq P_2 Q \\ \max_{\rho\in[-1,0]} \frac{1}{2}\log\left(1+\frac{\left(\sqrt{P_1}+\sqrt{P_2}\sqrt{1-\rho^2}\right)^2}{\left(\sqrt{Q}+\sqrt{P_2}\cdot\rho\right)^2+N}\right) & \text{o.w.} \end{cases} \tag{52}$$

where, in fact, the maximization over ρ can be limited to either $\rho = -1$, $\rho = 0$ or any real root ρ of the 4th order polynomial $g_0(\rho)$ (see (40)) that satisfies $\rho \in [-1, 0]$.

The proof of Theorem 7 appears in Appendix G.

The rest of this subsection is devoted to a discussion on the common message capacity results, comments on the capacity achieving scheme and numerical results.

1) Discussion: In the sequel, we separate the discussion on the common message capacity formula to the two complementary regimes of parameters (P_1, P_2, Q, N), the silent regime and the active regime (see Definition 2).

Silent Regime: It is shown that in the silent regime, the optimal values of σ_{12} and σ_{2s} as far as the common message capacity is concerned, are such that the condition (45) is met with equality, i.e.,

$$P_1\sigma_{2s}^2 + Q\sigma_{12}^2 = P_1P_2Q \tag{53}$$

or equivalently,

$$\rho_{12}^2 + \rho_{2s}^2 = 1. \tag{54}$$

This is also equivalent to

$$E\left(X_2 - \hat{X}_2^{lin}(X_1, S)\right)^2 = 0, \tag{55}$$

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

$$\hat{X}_2^{lin}(X_1, S) \;=\; \frac{\sigma_{12}}{P_1}X_1 + \frac{\sigma_{2s}}{Q}S. \tag{56}$$

Eq. (55) implies that in the silent regime

$$X_2 = \hat{X}_2^{lin}(X_1, S) = \frac{\sigma_{12}}{P_1}X_1 + \frac{\sigma_{2s}}{Q}S, \tag{57}$$

and thus,

$$Y = X_1\left(1 + \frac{\sigma_{12}}{P_1}\right) + S\left(1 + \frac{\sigma_{2s}}{Q}\right) + N', \tag{58}$$

calculating the optimal value of α (42) while accounting for (53), yields

$$\begin{aligned} \alpha_{opt}^{silent} &= -\frac{\sigma_{2s}}{Q} \\ U_{opt}^{silent} &= X_2 - \frac{\sigma_{2s}}{Q}S = \frac{\sigma_{12}}{P_1}X_1 \end{aligned} \tag{59}$$

and hence, in the silent regime of parameters, the common message capacity (52) formula is equal to

$$\begin{aligned} &\max_{\sigma_{12},\sigma_{2s}} I(U, X_1; Y) - I(U, X_1; S)|_{U=\frac{\sigma_{12}}{P_1}X_1} \\ &= \max_{\sigma_{12},\sigma_{2s}} I\left(X_1; X_1\left(1 + \frac{\sigma_{12}}{P_1}\right) + S\left(1 + \frac{\sigma_{2s}}{Q}\right) + N'\right) \end{aligned} \tag{60}$$

with σ_{12}, σ_{2s} satisfying (54). Inspecting (60), it is easy to verify that a simpler selection of U,

$$U_{opt}^{silent} \;=\; 0 \tag{61}$$

yields the same achievable rate and hence is also optimal.

The fact that in the silent regime the common message capacity is equal to (60), suggests that in this regime, in order to achieve capacity, the informed encoder can devote all its power to decreasing the interference and enhancing the signal of the uninformed encoder. No power is devoted to transmission of additional information, and hence, we refer to this region as silent.

A useful geometrical interpretation to the common message capacity formula in the silent regime can be attained by substituting $\cos\phi = \rho$ in (52), this yields

$$\max_{\phi} \frac{1}{2} \log \left(1 + \frac{\left(\sqrt{P_1} + \sqrt{P_2} \cdot \sin\phi\right)^2}{\left(\sqrt{Q} + \sqrt{P_2} \cdot \cos\phi\right)^2 + N} \right), \tag{62}$$

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 regime amounts to optimizing over ϕ (or ρ in (52)).

The maximizing ρ of the common message capacity formula in the silent regime (see (52)) is either 0, -1 or any real root of the 4th order polynomial $g_0(\rho)$ (40). For example, when $P_1 = P_2 = P > 0$ and $N = Q > 0$, the parameters lie in the silent regime, and finding the roots of $g_0(\rho)$ (40) 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 $g_0(\rho)$ is given by

$$\rho = \left(A^{1/3} + \frac{4 - 5\eta}{A^{1/3}(\eta + 1)} - 4 \right) \frac{1}{3\sqrt{\eta}} \tag{63}$$

with

$$A = 8 + 3\sqrt{3} \left(\frac{7\eta^3 - 4\eta^2 + 16\eta}{(\eta + 1)^3} \right), \quad \eta = \frac{P}{Q}. \tag{64}$$

Active Regime: In the active regime, 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 regime where no additional information is transmitted). Therefore, this regime 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 regime too, the maximizing (X_1, X_2, S) is Gaussian, but with

$$\sigma_{12}^{active} = -\sigma_{2s}^{active} = \frac{P_1(P_2 + N)}{P_1 + Q}, \tag{65}$$

i.e.,

$$\rho_{12}^{active} = \frac{P_1(P_2 + N)}{\sqrt{P_1 P_2}(P_1 + Q)}, \rho_{2s}^{active} = -\frac{P_1(P_2 + N)}{\sqrt{Q P_2}(P_1 + Q)}. \tag{66}$$

The resulting α_{opt} (see (42)) when using the correlations (65) is given by

$$\alpha_{opt}^{active} = \frac{P_2}{P_2+N}, \tag{67}$$

which is equal to the optimal α in Costa's setup [3] when the uninformed user is not present. As mentioned earlier, the choice of correlations (65), results in a surprising phenomenon which happens only in the active regime. The highest achievable common message 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 Theorem 3) can actually be achieved, even if the decoder is constrained to see only the sum of the channel outputs.

2) Comments on the Capacity Achieving Scheme: Next, we elaborate on the common message capacity achieving scheme for the Gaussian GGP channel resulting (using standard techniques [35]) from that of the finite alphabet GGP channel.

Silent Regime: Due to (57) and (61), 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 $M = \lfloor \exp\{n(C(P_1,P_2,Q,N)-\epsilon)\}\rfloor$ 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 $\mathbf{x}_m = (x_m(1), x_m(2), ..., x_m(n))$, and a state-sequence $\mathbf{s}$, the informed encoder simply transmits the n-vector $\tilde{\mathbf{x}}$ whose i-th symbol is given by

$$\tilde{x}_i = x_m(i)\frac{\sigma_{12}}{P_1} + s_i\frac{\sigma_{2s}}{Q}, \tag{68}$$

where $\sigma_{2s} = \sqrt{P_2 Q}\cdot\rho$, with ρ being the maximizer in (52) and $\rho_{12} = \sqrt{1-\rho_{2s}^2}$. Either an ordinary Maximum Likelihood (ML) decoder or a typicality decoder can be used to achieve the common message capacity.

Active Regime: As stated earlier, in this regime 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 regime, 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 (65), and

$$U = X_2 + \alpha_{opt}^{active} S \tag{69}$$

(see (67)).

3) Numerical Results: In Figure 5 the common message 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 common message capacity values for both $Q = 0$ and for $Q \to \infty$ are equal for all the members of each of these groups. The transition points between the silent regime and the active regime $Q = -\frac{P_1}{2} + \frac{\sqrt{P_1(P_1P_2+4(N+P_2)^2)}}{2\sqrt{P_2}}$ (see (49)) are indicated with diamonds.

In Figure 6, the common message 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 regime to the active regime are indicated with diamonds. In the silent regime, ρ_{2s} is, in fact, the maximizer of (52) and $\rho_{12} = \sqrt{1-\rho_{2s}^2}$. In the active regime, the optimal ρ_{12}, ρ_{2s} are given in (66). While ρ_{12} is a monotonically decreasing function of Q, $|\rho_{2s}|$ is increasing in the silent regime and decreasing in the active regime.

In Figure 7, the common message 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 regime to silent regime, i.e., $P_1 = \frac{P_2Q^2}{(P_2+N)^2-P_2Q}$ (see (49)). 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 common message capacity is $\frac{1}{2}\log(1+\frac{P_2}{N})$ for $P_1 = 0$.

In Figure 8, the common message 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 regime to silent regime, 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 regime.

D. The Capacity Region and the Common Capacity under Sum Power Constraints

Next, we state a closed form characterization of the capacity region under a sum power constraint (31). We denote by ζ the portion of the power that is used by the informed user.

Theorem 8 *The capacity region of the Gaussian GGP channel under sum power constraints, $\mathcal{C}(P, Q, N)$, is given by the union of the rate pairs satisfying*

$$(R_c, R_2) \in \mathcal{C}((1-\zeta)P, \zeta P, Q, N) \tag{70}$$

for some $\zeta \in [0, 1]$.

Proof: The theorem follows trivially by recalling the proof of Theorem 6 which implies (among other things) that the users had better exploit all the allowable power levels. Therefore the border of $\mathcal{C}(P_1, P_2, Q, N)$ is, in fact, the set of rate-pairs achievable whenever the uninformed user and the informed user transmit with powers P_1 and P_2 respectively. ■

The following theorem (whose proof appears in Appendix H), gives the common message capacity under sum power constraints.

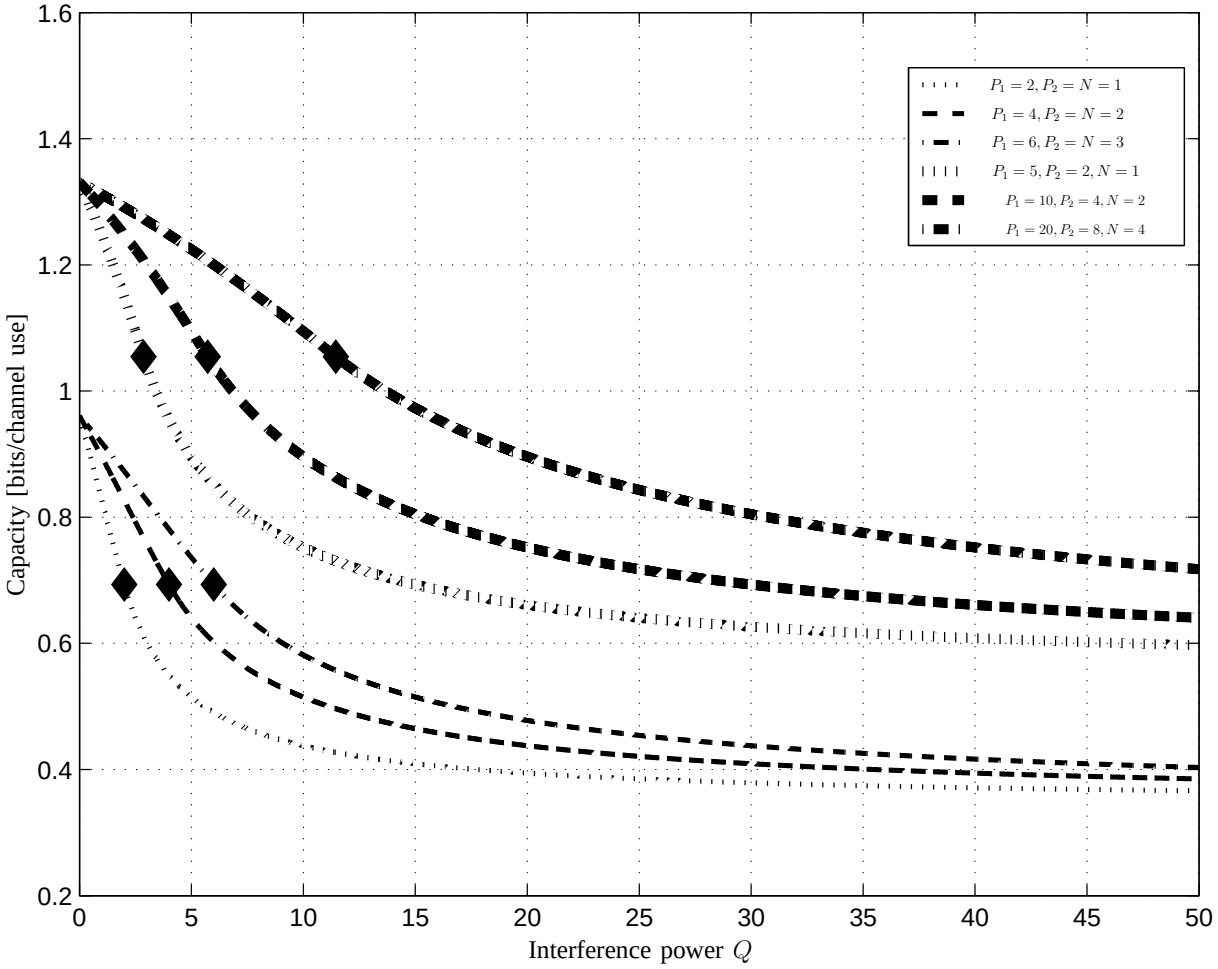

Fig. 5. Common message capacity as a function of the interference power Q.

Theorem 9 *The common message capacity of the Gaussian GGP channel under a sum power constraint, $C(P,Q,N)$, is given by the following formula*

$$C(P,Q,N) = \begin{cases} \frac{1}{2}\log\left(1+\frac{P}{N}\right) \text{ if } N+P \le Q \\ \frac{1}{2}\log\frac{(Q+P+N)^2}{4QN} \text{ if } Q_0 \le Q \le N+P \\ \max_{0\le\zeta\le 1} R(\zeta,P,Q,N) \text{ otherwise} \end{cases}, \tag{71}$$

where $Q_0 = \frac{1}{3}(N-P) + \frac{2}{3}\sqrt{N^2+NP+P^2}$ *and*

$$R(\zeta,P,Q,N) = \max_{\rho\in[-1,0]} \frac{1}{2}\log\left(1+\frac{\left(\sqrt{(1-\zeta)P}+\sqrt{\zeta P}\sqrt{1-\rho^2}\right)^2}{\left(\sqrt{Q}+\sqrt{\zeta P}\cdot\rho\right)^2+N}\right).$$

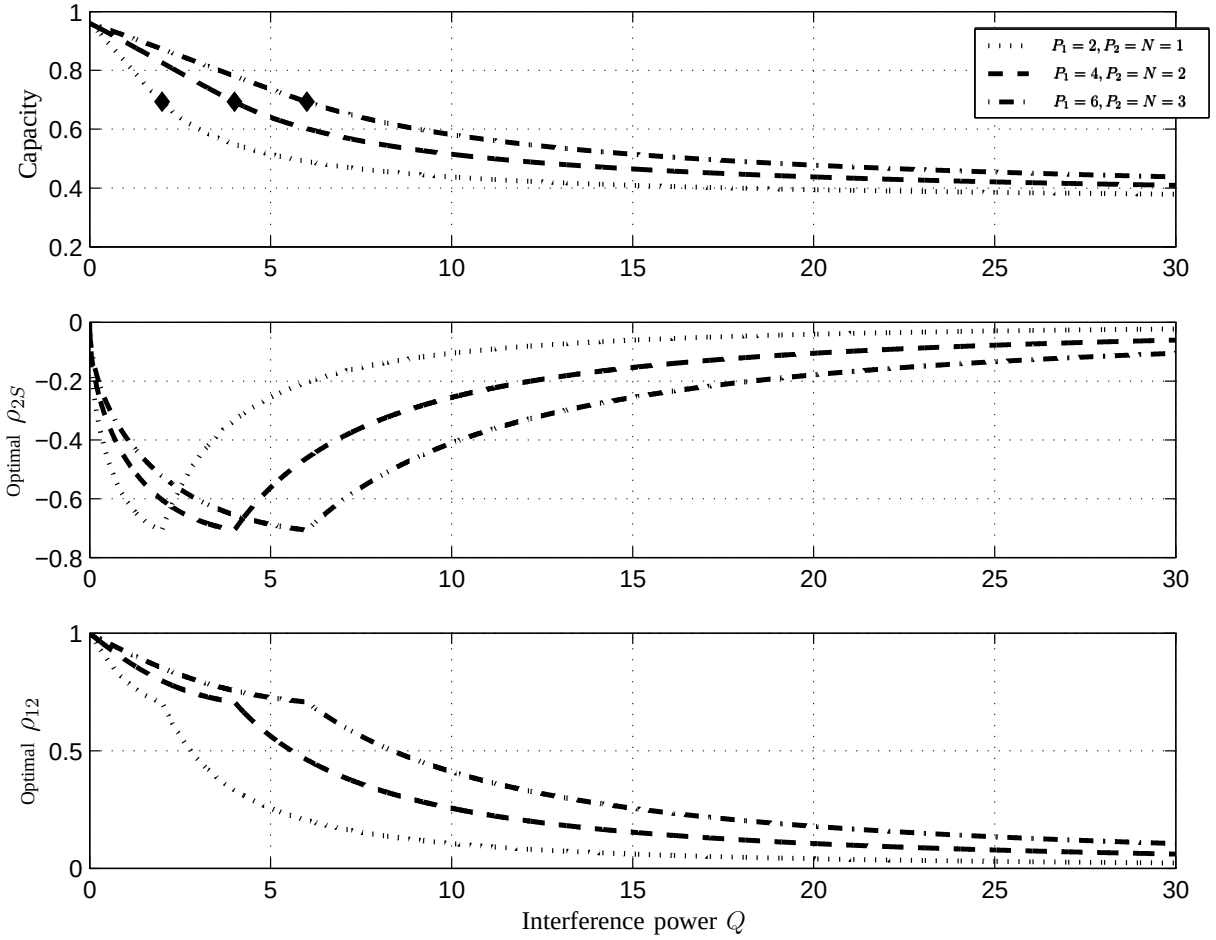

Fig. 6. Common message apacity and the optimal correlation coefficients, ρ_{12} and ρ_{2s}, as a function of the interference power Q.

The power allocation that achieves the common message capacity is

$$\zeta_{opt}(P,Q,N) = \begin{cases} 1 & \text{if } N+P \le Q \\ \frac{P+Q-N}{2P} & \text{if } Q_0 \le Q \le N+P \\ \text{argmax}_{0\le\zeta\le 1} R(\zeta,P,Q,N) & \text{otherwise} \end{cases} . \tag{72}$$

Since the line that meets the R_c axis (in the $R_c - R_2$ plane) at -45^0 at the point $R_c = C(P,Q,N)$, is an outer bound on the capacity region (being a collection of trapezoids), Theorem 9 enables to simplify the expression for $\mathcal{C}(P,Q,N)$ as follows.

- Whenever $Q \ge N+P$ the optimal ζ is 1, and $\mathcal{C}(P,Q,N)$ is the triangle whose vertices are $(R_c,R_2) = (0,0), \left(\frac{1}{2}\log\left(1+\frac{P}{N}\right),0\right)$, and $\left(0,\frac{1}{2}\log\left(1+\frac{P}{N}\right)\right)$. This means that as long as $Q \ge N+P$, the capacity region is not affected by Q, since the best strategy is to let

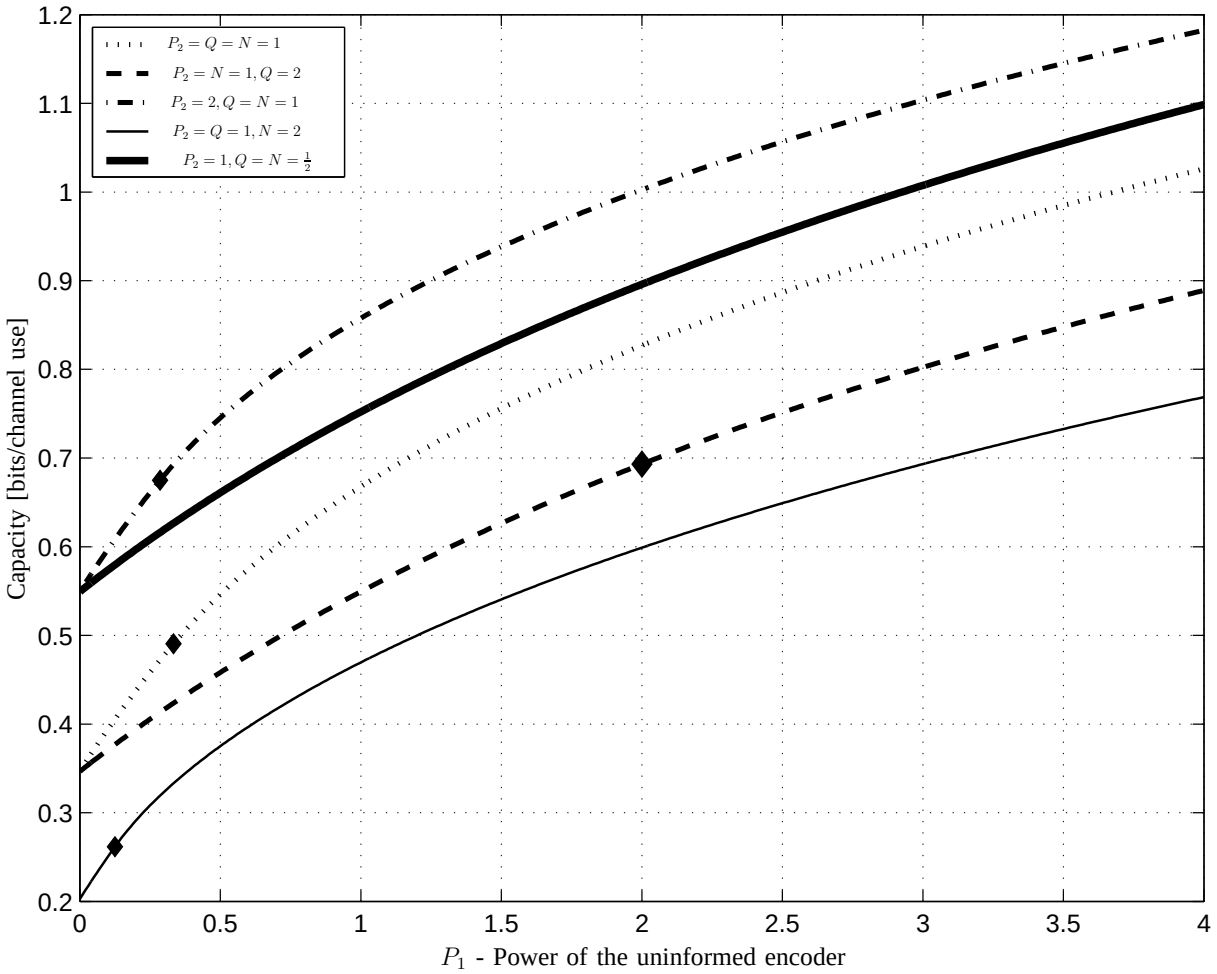

Fig. 7. Common message capacity as a function of P_1.

the informed user use all the power, and this degenerates to a single user Costa channel, where the transmitted information bits can be divided between $\mathcal{W}_2$ and $\mathcal{W}_c$.

- Whenever $\frac{1}{3}(N-P) + \frac{2}{3}\sqrt{N^2+NP+P^2} \leq Q \leq N+P$, the border of $\mathcal{C}(P,Q,N)$ contains the line segment between the points (R_c, R_2) given by

$$\left(\frac{1}{2}\log\left(\frac{P+Q+N}{3Q-P-N}\right), \frac{1}{2}\log\left(\frac{(3Q-P-N)(P+Q+N)}{4QN}\right)\right)$$

and $\left(\frac{(P+Q+N)^2}{4QN}, 0\right)$ (this follows by substituting $P_1 = (1-\zeta)P$, $P_2 = \zeta P$, and $\zeta = \frac{P+Q-N}{2P}$ in the points (c) and (d) of Proposition 1).

In Figure 9, the borders of the sum power constraint capacity regions for $P = 3, N = 1$ and three values of Q $(0.5, 2.5, 5)$ are plotted. One can see the triangle shape for $Q = 5$. In

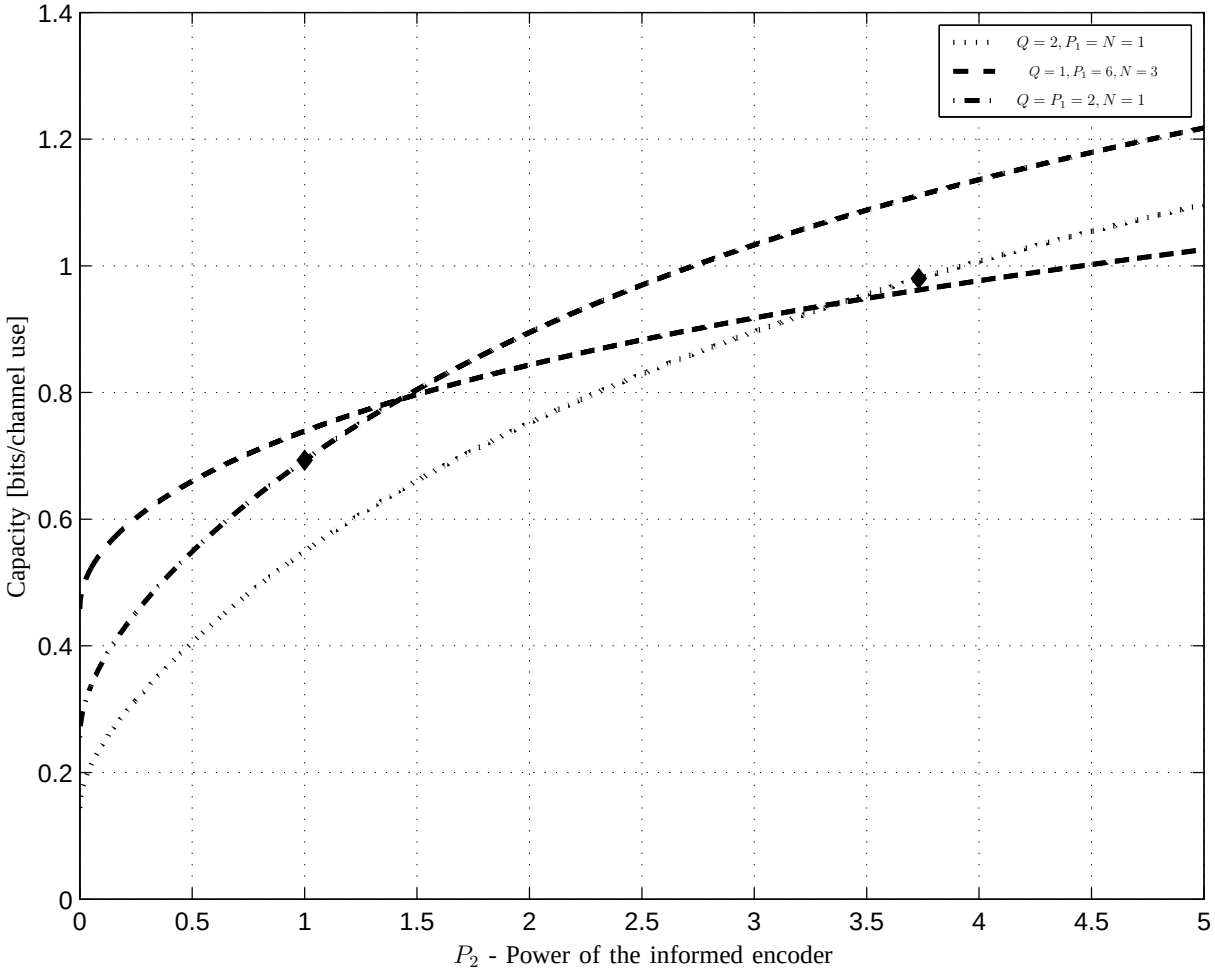

Fig. 8. Common message capacity as a function of P_2.

Figure 10, the common message capacity under sum power constraints are plotted for $N = 1$ and five values of Q $(0.2, 0.5, 1, 2, 5)$.

VI. CONCLUSIONS

In this work we analyze a setup of cooperative communication over the GP MAC, 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 that the user that is informed of the CSI transmits a private message as well. We characterize the capacity region 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 common 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

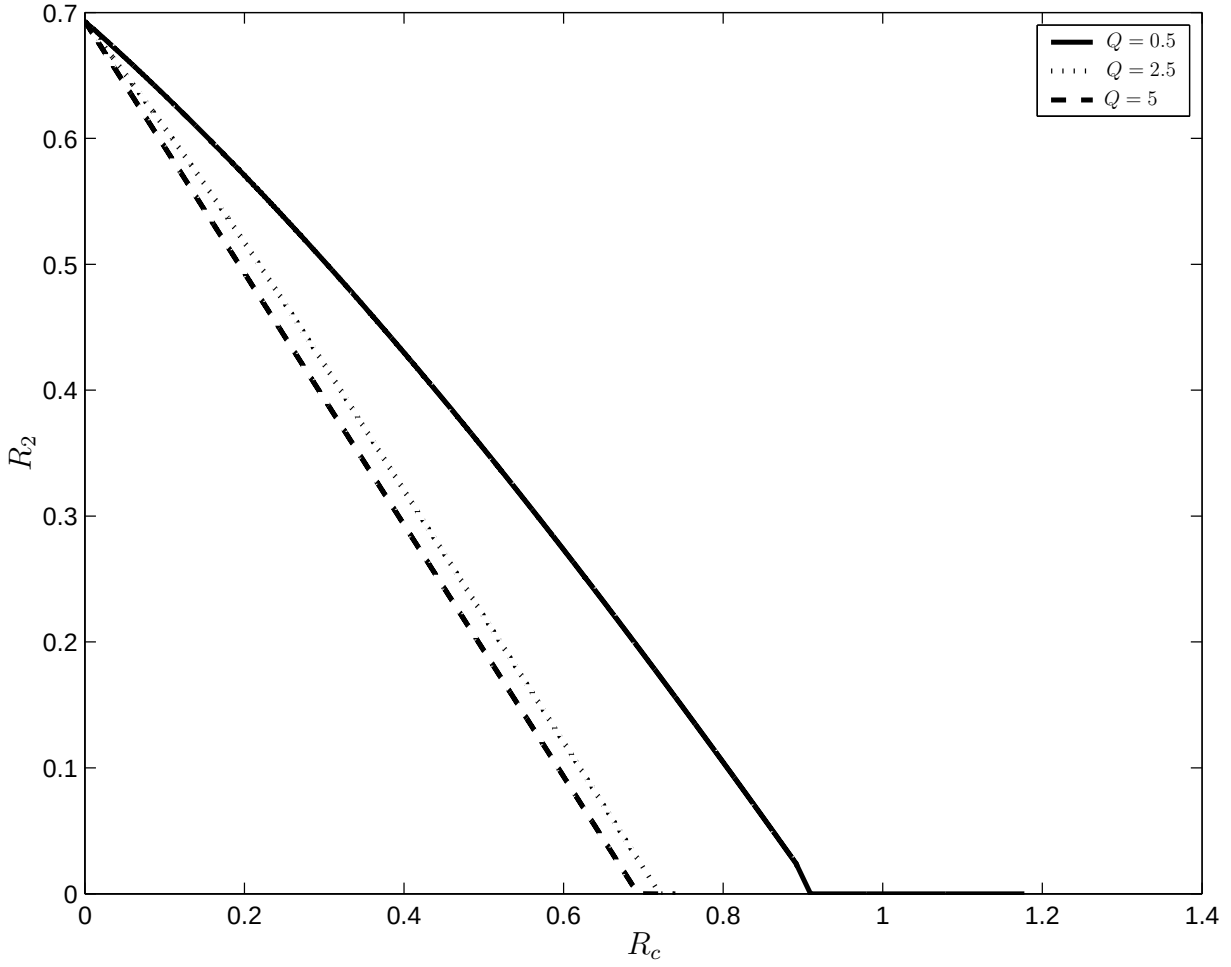

Fig. 9. Sum power constraint capacity regions for $P = 3, N = 1$.

encoder using a binning scheme, and uses it to transmit part B of the message a well as its private message. The results are then specialized to the case where it is only the common message that is being transmitted, and in this case the capacity is referred to as a common message capacity. Further, we establish two useful results for the general finite-input alphabet case. The first is a useful outer bound on the capacity region 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 region formula of this channel as a special case of the general GGP channel capacity region formula. We also characterize the capacity region of an asymmetric causal state-dependent channel which is the same channel as the GGP channel, with the exception

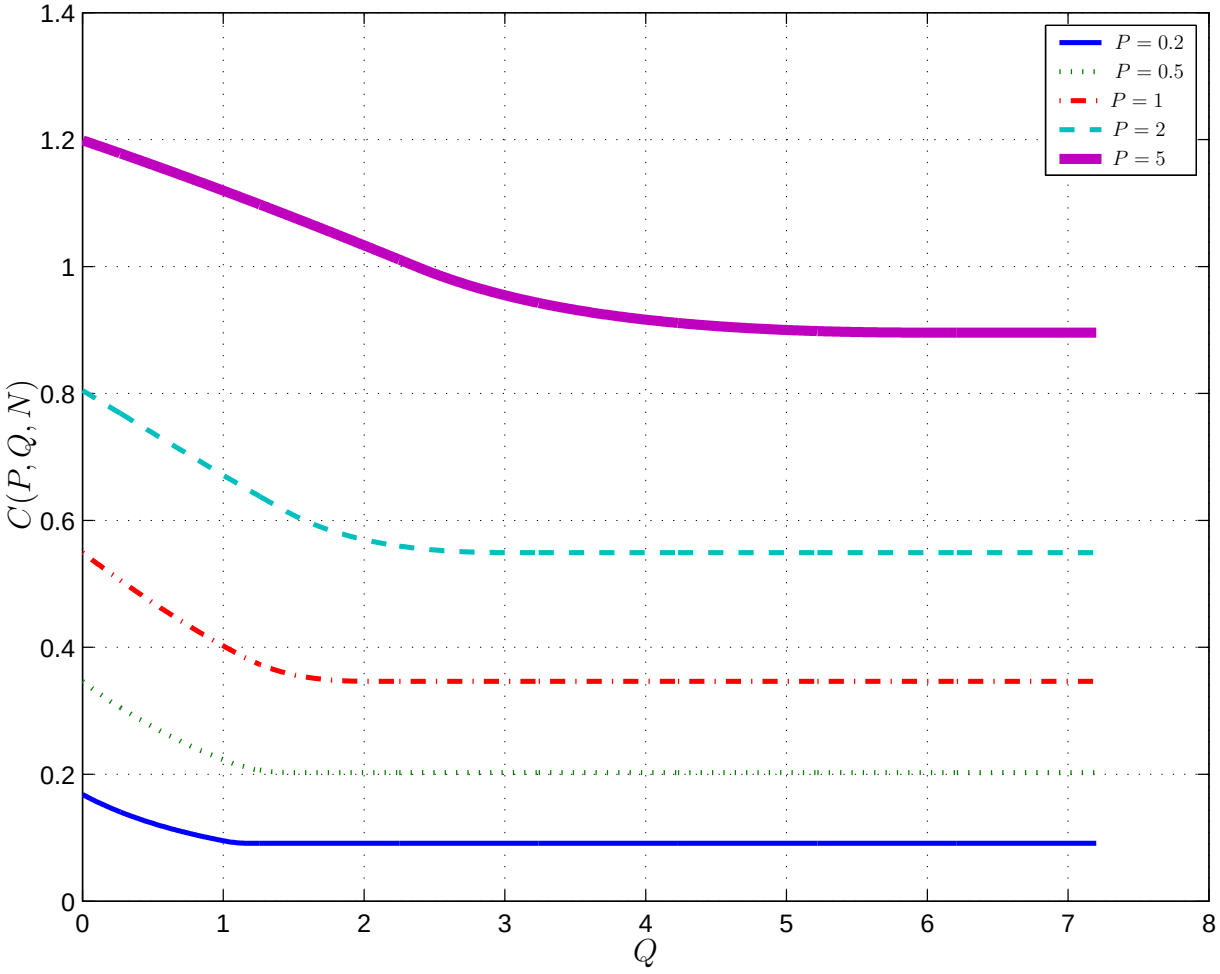

Fig. 10. Sum power constraint capacity for $N = 1$.

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. We investigate two power constraints, the first being a constraint on each of the power levels of the two encoders, and the second being a constraint on the sum of powers used by the transmitters. By proving that in the Gaussian case the operative bound is achievable, we establish a closed-form formula for the capacity region of this channel for both power constraints. Technically speaking, this outer bound enables proving that one can consider only Gaussian distributions for the single-letter expression without loss of generality. Four parameters determine the capacity region: 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 regime and an active regime. The common message capacity (which determines one of the vertices of the capacity region) formula as a function of these four parameters takes

on two different forms depending on whether the parameters lie in the active regime or the silent regime. To achieve the common message capacity, in the silent regime the informed encoder allocates a portion of its power to interference cancelation and the remaining power to coherently enhancing the uninformed user's signal. In the active regime, 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 regime, the common message 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). We also determine the optimal power allocation for the common message capacity under sum power constraints. Finally, we note that the results are extendable to a general multiuser setup under the common message regime [32].

APPENDIX

A. Proof of Theorem 1

1) Converse Part of Theorem 1: Let an $(\epsilon_n, n, R_c, R_2)$-code be given. Thus, we have using Fano's inequality

$$\begin{aligned} n(R_c, +R_2) &= H(\mathcal{W}_c, \mathcal{W}_2) \\ &\leq I(\mathcal{W}_c, \mathcal{W}_2; Y_1^n) + 1 + n(R_c + R_2)\epsilon_n. \end{aligned} \tag{73}$$

Further,

$$\begin{aligned} &I(\mathcal{W}_c, \mathcal{W}_2; Y_1^n) \\ &\leq \sum_{i=1}^{n} \left[I(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; Y_i) - I(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; S_i) \right] \\ &= \sum_{i=1}^{n} \left[I(\mathcal{W}_c, X_1(i), \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; Y_i) - I(\mathcal{W}_c, X_1(i), \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; S_i) \right], \end{aligned} \tag{74}$$

where the inequality follows exactly as in the derivation of the converse part of the proof of the capacity formula for the ordinary GP channel [2] by replacing W with $(\mathcal{W}_c, \mathcal{W}_2)$, and the last equality holds since $X_1(i)$ is a function of $\mathcal{W}_c$.

Similarly, since $\mathcal{W}_c$ and $\mathcal{W}_2$ are independent

$$\begin{aligned} &nR_2 = H(\mathcal{W}_2|\mathcal{W}_c) \\ &\leq I(\mathcal{W}_2; Y_1^n|\mathcal{W}_c) + 1 + nR_2\epsilon_n \\ &\leq \sum_{i=1}^{n} \left[I(\mathcal{W}_2, Y^{i-1}, S_{i+1}^n; Y_i|\mathcal{W}_c) - I(\mathcal{W}_2, Y^{i-1}, S_{i+1}^n; S_i|\mathcal{W}_c) \right] + 1 + nR_2\epsilon_n \\ &= \sum_{i=1}^{n} \left[I(\mathcal{W}_2, Y^{i-1}, S_{i+1}^n; Y_i|\mathcal{W}_c, X_1(i)) - I(\mathcal{W}_2, Y^{i-1}, S_{i+1}^n; S_i|\mathcal{W}_c, X_1(i)) \right] + 1 + nR_2\epsilon_n \\ &\leq \sum_{i=1}^{n} \left[I(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; Y_i|X_1(i)) - I(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n; S_i|X_1(i)) \right] + 1 + nR_2\epsilon_n, \end{aligned} \tag{75}$$

where the last inequality follows since S_i and $(\mathcal{W}_c, X_1(i))$ are independent. Therefore, defining $\bar{U}_i = \left(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n\right)$ one has

$$R_c + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i, X_1(i); Y_i) - I(\bar{U}_i, X_1(i); S_i) + \frac{1}{n} + (R_c + R_2)\epsilon_n$$
$$R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i; Y_i | X_1(i)) - I(\bar{U}_i; S_i | X_1(i)) + \frac{1}{n} + R_2\epsilon_n. \tag{76}$$

Now, we introduce a time-sharing random variable, T, distributed uniformly over $\{1, ..., n\}$, and denote the collection of random variables

$$(S, X_1, \bar{U}, X_2, Y) = (S_T, X_1(T), \bar{U}_T, X_2(T), Y_T), \tag{77}$$

to obtain

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} I(X_1(i), \bar{U}_i; Y_i) - I(X_1(i), \bar{U}_i; S_i) \\ = \;& I(X_1(T), \bar{U}; Y|T) - I(X_1(T), \bar{U}; S|T) \\ = \;& I(T, X_1, \bar{U}; Y) - I(T; Y) - I(T, X_1, \bar{U}; S) + I(T; S) \\ \leq \;& I(T, X_1, \bar{U}; Y) - I(T, X_1, \bar{U}; S), \end{aligned} \tag{78}$$

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

$$R_c + R_2 \;\leq\; I(X_1, U; Y) - I(X_1, U; S) + \frac{1}{n} + (R_c + R_2)\epsilon_n. \tag{79}$$

Similarly,

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} I(\bar{U}_i; Y_i | X_1(i)) - I(\bar{U}_i; S_i | X_1(i)) \\ = \;& I(\bar{U}; Y | X_1, T) - I(\bar{U}; S | X_1(T), T) \\ = \;& I(\bar{U}, T; Y | X_1) - I(Y; T | X_1) - I(\bar{U}, T; S | X_1) + I(T; S | X_1) \\ \leq \;& I(\bar{U}, T; Y | X_1) - I(\bar{U}, T; S | X_1), \end{aligned} \tag{80}$$

and one gets

$$R_2 \;\leq\; I(U; Y | X_1) - I(U; S | X_1) + \frac{1}{n} + R_2\epsilon_n. \tag{81}$$

The above constitutes the proof that for every (ϵ_n, n, R)-code, there exists a measure of the form (9) with essentially $R_c + R_2 \leq I(X_1, U; Y) - I(X_1, U; S)$, and $R_c \leq I(U; Y|X_1) - I(U; S|X_1)$.

It remains to show that the alphabet of the random variables U can be limited without loss of generality as stated in (10). This is done by a standard application of the support Lemma. First, fix a distribution μ of (S, X_1, U, X_2, Y) on the Borel σ-algebra of $\mathcal{P}(\mathcal{S} \times \mathcal{X}_1 \times \mathcal{U} \times \mathcal{X}_2 \times \mathcal{Y})$ that has the form (9). Note that

$$I_\mu(X_1, U; Y) - I_\mu(X_1, U; S) \;=\; I_\mu(U; Y | X_1) - I_\mu(U; S | X_1) + I_\mu(X_1; Y), \tag{82}$$

and

$$\begin{aligned} & I_\mu(X_1,U;Y) - I_\mu(X_1,U;S) \\ = \;& I_\mu(U;Y) - I_\mu(U;S) + I_\mu(X_1;Y|U) - I_\mu(X_1;S|U) \\ = \;& H_\mu(Y) - H_\mu(S) - H_\mu(Y|X_1,U) + H_\mu(S|X_1,U) \\ = \;& H_\mu(Y) - H_\mu(S) - H_\mu(X_1,Y|U) + H_\mu(X_1,S|U), \end{aligned} \tag{83}$$

Hence, it suffices to show that the following functionals of $\mu(S,X_1,U,X_2,Y)$

$$f_{s,x,\tilde{x}}(\mu) = \mu(s,x,\tilde{x}) \quad \forall (s,x,\tilde{x}) \in \mathcal{S}\times\mathcal{X}_1\times\mathcal{X}_2 \tag{84}$$

$$f_0(\mu) = \int_u d\mu(u)\,(H_\mu(X_1,S|u) - H_\mu(X_1,Y|u)) \tag{85}$$

can be preserved with another measure μ' that has the form (9). To satisfy this condition, according to the support Lemma, since there are $A = |\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|$ functionals[1], the cardinality of the alphabet of $\mathcal{U}$ can be taken to be A without loss of generality.

2) 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.

Error probability analysis: For a measure P, let $T_\epsilon(P)$ stand for the set of ϵ-typical sequences. Assume that the transmitted common message is ℓ and the informed user message is k, and that $\mathbf{s}$ is the state sequence. Let $\mathbf{s}$ and $\mathbf{x} = \mathbf{x}_\ell$ stand for the states sequence and the codeword of the uninformed encoder, respectively. One has

$$\begin{aligned} \Pr(\text{error}) = & \sum_{(\mathbf{s}',\mathbf{x}')\in T_\epsilon^c(Q_S\times P_{X_1})} \Pr(\mathbf{s}',\mathbf{x}') \\ & + \sum_{(\mathbf{s}',\mathbf{x}')\in T_\epsilon(Q_S\times P_{X_1})} \Pr(\mathbf{s}',\mathbf{x}')\Pr(\text{error}|\mathbf{s}',\mathbf{x}'). \end{aligned} \tag{86}$$

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 (86). The error event is contained in the union of the following events

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

One can easily realize as an immediate extension of [2] that $\Pr(E_1(\mathbf{s},\mathbf{x}))$ behaves essentially like $\left[1-2^{-n(I(S;U|X_1)+\epsilon)}\right]^J \leq \exp(2^{-n\epsilon})$. Given that $E_1(\mathbf{s},\mathbf{x})$ does not occur, we have that

[1]In (84), there are in fact only $|\mathcal{S}|\cdot|\mathcal{X}_1|\cdot|\mathcal{X}_2|$-1 degrees of freedom.

$(\mathbf{s},\mathbf{x})$ are jointly typical with the outputs of the encoders $\mathbf{x},\tilde{\mathbf{x}}$ and with $\mathbf{u}_{\ell,k,j(\mathbf{s},\ell,k)}$, that is, $(\mathbf{s},\mathbf{x},\mathbf{u}_{\ell,k,j(\mathbf{s},\ell,k)},\tilde{\mathbf{x}}) \in T_\epsilon(Q_S \times P_{X_1} \times P_{U,X_2|S,,X_1})$.

For $\mathbf{s},\mathbf{x},\mathbf{u}_{\ell,k,j(\mathbf{s},\ell,k)},\tilde{\mathbf{x}}$ 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,

$$\begin{aligned} \Pr(E_3) &\leq M_1 2^{-n(I(X_1;Y)+\epsilon)} + \delta \leq 2^{-2n\epsilon} + \delta \\ \Pr(E_5(\mathbf{u})) &\leq M_2 J 2^{-n(I(U;Y|X_1)+\epsilon)} + \delta \leq 2^{-2n\epsilon} + \delta. \end{aligned} \tag{87}$$

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

B. Proof of Corollary 2

The Corollary follows from (13), by substituting $Z = (U, X_1)$. The alphabet cardinality bound is slightly larger (relatively to Corollary 1) $|\mathcal{Z}| \leq |\mathcal{S}| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2| + 1$, because we have an additional functional $H(X_1|Z)$ whose zero value should be preserved. □

C. Proof of Theorem 3

Since the region described in (23) is trivially achievable by ignoring the CSI, we only need to show that this is, in fact, an outer bound on the capacity region. Since the degenerate parallel channel is a special case of a GGP channel with output (Y_1, Y_2), its capacity region can be calculated using Theorem 1 by replacing Y with (Y_1, Y_2). To establish an outer bound on the capacity region we note that

$$\begin{aligned} & I(U,X_1;Y_1,Y_2) - I(X_1,U;S) \\ \overset{(a)}{=}\ & I(U,X_1;Y_1,Y_2) - I(U;S|X_1) \\ =\ & I(X_1;Y_1,Y_2) + I(U;Y_1,Y_2|X_1) - I(U;S|X_1) \\ =\ & I(X_1;Y_1,Y_2) + I(U;Y_1|X_1) \\ & + I(U;Y_2|X_1,Y_1) - I(U;S|X_1) \\ \overset{(b)}{\leq}\ & I(X_1;Y_1,Y_2) + I(U;Y_2|X_1,Y_1) \\ =\ & I(X_1;Y_1) + I(X_1;Y_2|Y_1) + I(U;Y_2|X_1,Y_1) \\ =\ & I(X_1;Y_1) + I(X_1,U;Y_2|Y_1) \\ \leq\ & I(X_1;Y_1) + I(X_1,X_2,S,U,Y_1;Y_2) \\ \overset{(c)}{=}\ & I(X_1;Y_1) + I(X_2;Y_2). \end{aligned} \tag{88}$$

where (a) follows from $I(X_1;S)=0$, (b) holds since $I(U;Y_1|X_1) \leq I(U;S,X_1|X_1) = I(U;S|X_1)$, and (c) is because $(X_1,S,U,Y_1) \leftrightarrow X_2 \leftrightarrow Y_2$. Further,

$$\begin{aligned} & I(U;Y_1,Y_2|X_1) - I(U;S|X_1) \\ = \; & I(U;Y_1,Y_2|X_1) - I(U;S|X_1) \\ = \; & I(U;Y_1,Y_2|X_1,S) - I(U;S|X_1,Y_1,Y_2) \\ \leq \; & I(X_2;Y_1,Y_2|X_1,S) \\ \leq \; & I(X_2;Y_2), \end{aligned} \tag{89}$$

where the first inequality is because $U \leftrightarrow (X_1,X_2,S) \leftrightarrow (Y_1,Y_2)$ and the second inequality is because $(X_1,S,Y_1) \leftrightarrow X_2 \leftrightarrow Y_2$.

This concludes the converse part, and the proof of the theorem.

□

D. Proof of Theorem 4

The proof of the converse part follows similarly to the proof of Theorem 1, with the exception that in the causal case $(\mathcal{W}_c, \mathcal{W}_2, Y^{i-1}, S_{i+1}^n)$ is independent of S_i. Since the state sequence is stationary and $X_1(i)$ is deterministic given $\mathcal{W}_c$ this means that the collection of random variables defined in (77) in the causal case is such that $I(U,X_1;S)=0$. As for the cardinality of $\mathcal{U}$, in addition to the functionals (85) one should also preserve the values of $I_\mu(U,X_1;S)=0$. This can be done by increasing $|\mathcal{U}|$ by 1 (relatively to the non-causal case) since $I(U,X_1;S) = H(S) - H(X_1,S|U) + H(X_1|U)$.

As for the direct part, since (X_1,U) and S are independent, the encoding scheme of the non-causal case degenerates to one that does not include binning and hence, does not require non-causal knowledge of S^n prior to transmission.

□

E. Proof of Theorem 6

1) Proof of the Converse Part of Theorem 6: In light of the comment preceding (33) and Theorem 2, we can provide an outer bound for the capacity of the Gaussian GGP channel in terms of the closure of the set of the rate pairs (R_c, R_2) satisfying

$$R_2 \leq I(X_2;Y|S,X_1) \tag{90}$$

$$R_c + R_2 \leq I(X_2,X_1;Y|S) - I(S;X_1|Y) \tag{91}$$

for some measure $P_{S,X_1,X_2,Y} = Q_S P_{X_1} P_{X_2|S,X_1} W_{Y|S,X_1,X_2}$ such that $E(X_1^2) < P_1$, and $E(X_2^2) < P_2$. In the sequel, we shall show that in the Gaussian channel case (29), the three inequalities (20), (21) and (22) can be met with equality simultaneously.

Now,

$$\begin{aligned} & I(X_2,X_1;Y|S) - I(S;X_1|Y) \\ = \; & h(Y|S) - h(N') - h(S|Y) + h(S|X_1,Y) \\ = \; & h(Y) + h(S|X_1,Y) - h(S) - h(N'), \end{aligned} \tag{92}$$

where $h(Y|S)$ denotes the differential entropy of Y given S. Obviously for fixed second moments,

$$E(X_i^2) = \tilde{P}_i \leq P_i \quad i = 1, 2, \quad \sigma_{12} \triangleq E(X_2 X_1), \quad \sigma_{2s} \triangleq E(X_2 S), \quad E(X_1 S) = 0 \tag{93}$$

the differential entropy, $h(Y)$, is maximized if Y is Gaussian, i.e.,

$$h(Y) = \frac{1}{2}\log(2\pi e)(\tilde{P}_2 + \tilde{P}_1 + 2\sigma_{12} + 2\sigma_{2s} + Q + N). \tag{94}$$

Similarly, the conditional differential entropy, $h(S|X_1, Y)$, is maximized if (S, X_1, Y) are jointly Gaussian. Now, denote by $\hat{S}_{opt}(X_1, Y) = E(S|X_1, Y)$ the MMSE estimator of S given (X_1, Y) observe that

$$\begin{aligned}
h(S|X_1, Y) &= h(S - \hat{S}_{opt}(X_1, Y)|X_1, Y) \\
&\leq h(S - E(S|X_1, Y)) \\
&= h(S - E(S|X_1, X_2 + S + N)) \\
&\leq \frac{1}{2}\log\left[(2\pi e)E\left(S - E(S|X_1, X_2 + S + N)\right)^2\right] \\
&\leq \frac{1}{2}\log\left[(2\pi e)\min_{a,b} E\left(S - aX_1 - b(X_2 + S + N)\right)^2\right] \\
&= \frac{1}{2}\log\left[(2\pi e)E\left(S - a_{opt}X_1 - b_{opt}(X_2 + S + N)\right)^2\right],
\end{aligned} \tag{95}$$

where in fact all the inequalities are attained with equality if (S, X_1, X_2, Y) are jointly Gaussian, and

$$\begin{aligned}
a_{opt} &= -\frac{\sigma_{12}(\sigma_{2s} + Q)}{\tilde{P}_1(\tilde{P}_2 + 2\sigma_{2s} + Q + N) - \sigma_{12}^2} \\
b_{opt} &= -\frac{a_{opt}\tilde{P}_1}{\sigma_{12}}.
\end{aligned} \tag{96}$$

Plugging (96) into (95) we get,

$$h(S|X_1, Y) = \frac{1}{2}\log\left((2\pi e)\frac{Q\tilde{P}_2\tilde{P}_1 + \tilde{P}_1 NQ - \sigma_{2s}^2\tilde{P}_1 - Q\sigma_{12}^2}{\tilde{P}_2\tilde{P}_1 + 2\tilde{P}_1\sigma_{2s} + \tilde{P}_1 Q + \tilde{P}_1 N - \sigma_{12}^2}\right). \tag{97}$$

Recall the definition of the correlation coefficients (46). Eqs. (92)-(97) yield for (S, X_1, X_2) jointly Gaussian

$$\begin{aligned}
& I(X_2, X_1; Y|S) - I(S; X_1|Y) \\
&= \frac{1}{2}\log\left(\frac{(Q\tilde{P}_2\tilde{P}_1 + \tilde{P}_1 NQ - {\sigma_{2s}}^2\tilde{P}_1 - Q{\sigma_{12}}^2)(\tilde{P}_2 + 2\sigma_{12} + 2\sigma_{2s} + \tilde{P}_1 + Q + N)}{QN\left(\tilde{P}_2\tilde{P}_1 + 2\tilde{P}_1\sigma_{2s} + \tilde{P}_1 Q + \tilde{P}_1 N - {\sigma_{12}}^2\right)}\right) \qquad (98) \\
&= \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{\tilde{P}_1} + \rho_{12}\sqrt{\tilde{P}_2}\right)^2}{\tilde{P}_2(1 - \rho_{2s}^2 - \rho_{12}^2) + \left(\sqrt{Q} + \rho_{2s}\sqrt{\tilde{P}_2}\right)^2 + N}\right) + \frac{1}{2}\log\left(1 + \frac{\tilde{P}_2(1 - \rho_{2s}^2 - \rho_{12}^2)}{N}\right), \\
&\triangleq \mathbf{C}(\tilde{P}_2, \tilde{P}_1, \rho_{12}, \rho_{2s}), \qquad (99)
\end{aligned}$$

which also depends on Q and N, but for the sake of brevity this is omitted from the notation.

Combining (91)-(99), we get that for fixed second moments,

$$R_c + R_2 \;\leq\; \mathbf{C}(\tilde{P}_2, \tilde{P}_1, \rho_{12}, \rho_{2s}). \tag{100}$$

As for (90), for fixed second moments as in (93), $I(X_2; Y|S, X_1) = h(X_2+N'|S, X_1)-h(N')$ is obviously upper bounded by (S, X_1, X_2) that are jointly Gaussian, yielding

$$\begin{aligned} R_2 &\leq h(X_2 + N'|S, X_1) - h(N') \\ &= \frac{1}{2}\log E\left(\left(X_2 + N' - \frac{\sigma_{2s}}{Q}S - \frac{\sigma_{12}}{P_1}X_1\right)^2\right) - \frac{1}{2}\log N \\ &= \frac{1}{2}\log\left(1 + \frac{\tilde{P}_2(1-\rho_{2s}^2-\rho_{12}^2)}{N}\right) \\ &\triangleq \Theta(\tilde{P}_2, \rho_{2s}, \rho_{12}). \end{aligned} \tag{101}$$

The capacity region is therefore outer bounded by the closure of the convex hull of the rate pairs (R_c, R_2) satisfying

$$\begin{aligned} R_2 &\leq \Theta(\tilde{P}_2, \rho_{2s}, \rho_{12}) \\ R_c + R_2 &\leq \mathbf{C}(\tilde{P}_2, \tilde{P}_1, \rho_{12}, \rho_{2s}), \end{aligned} \tag{102}$$

for some covariance matrix $\Lambda_{X_1,X_2,S,N'}$ of (X_1, X_2, S, N'),

$$\Lambda_{X_1,X_2,S,N'} \;=\; \begin{pmatrix} \tilde{P}_1 & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \tilde{P}_2 & \sigma_{2s} & 0 \\ 0 & \sigma_{2s} & Q & 0 \\ 0 & 0 & 0 & N \end{pmatrix} \tag{103}$$

satisfying

$$\tilde{P}_1 \leq P_1\ ,\ \tilde{P}_2 \leq P_2 \tag{104}$$

and the nonnegative-definiteness condition

$$\det\left(\Lambda_{X_1,X_2,S,N'}\right) = \tilde{P}_1(\tilde{P}_2 QN - \sigma_{2s}^2 N) - \sigma_{12}^2 QN \geq 0, \tag{105}$$

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

$$\rho_{2s}^2 + \rho_{12}^2 \leq 1. \tag{106}$$

It remains to prove that one can replace $\tilde{P}_1$ and $\tilde{P}_2$ in (102) by P_1, P_2, respectively. This is equivalent to showing that the users should exploit all their allowable power.

To realize this, inspecting $\mathbf{C}(\tilde{P}_2, \tilde{P}_1, \rho_{12}, \rho_{2s})$ and $\Theta(\tilde{P}_2, \rho_{12}, \rho_{2s})$, it is evident that it suffices to consider $\rho_{12} \in [0, 1]$ and $\rho_{2s} \in [-1, 0]$. It is easy to verify that for fixed $\tilde{P}_2, \rho_{12} \in [0, 1], \rho_{2s} \in [-1, 0]$, the function $\mathbf{C}(\tilde{P}_2, \tilde{P}_1, \rho_{12}, \rho_{2s})$ increases with $\tilde{P}_1$. As $\Theta(\tilde{P}_2, \rho_{2s}, \rho_{12})$ is unaffected by $\tilde{P}_1$, this proves that one can replace $\tilde{P}_1$ by P_1.

Now, a simple argument shows that also the informed encoder should use its entire power. Let $P_1, \tilde{P}_2$ be the power used by the uninformed encoder. Assume in negation that the informed encoder does not use all its power (i.e., $\tilde{P}_2 < P_2$), and let an encoding scheme be given. Now, consider an altered encoding scheme in which the informed encoder operates as before but uses the extra power that is not exploited, $P' \triangleq P_2 - \tilde{P}_2$, adding to its original signal the uninformed encoder's signal multiplied by $\sqrt{P'}$. This new scheme is equivalent to the original scheme in which the uninformed encoder operate at power level $P_1 + P'$ rather than P_1. But, we have previously proved that the uninformed encoder had better use all its power[2], hence, this scheme improves on the original one and contradicts the assumption that the informed user does not use its entire power.

2) Proof of the Direct Part of Theorem 6: As for establishing an inner bound on the capacity, choose in (8), S, X_1, X_2, Y that are jointly Gaussian with second moments $E(X_1^2) = P_1, E(X_2^2) = P_2$, note that Y in (29) can be expanded as follows:

$$Y = (X_2 - \hat{X}_2^{lin}(X_1, S)) + \hat{X}_2^{lin}(X_1, S) + X_1 + S + N', \tag{107}$$

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 see (56). Denoting

$$\begin{aligned} X_2' &= X_2 - \hat{X}_2^{lin}(X_1, S) \\ X_1' &= \left(1 + \frac{\sigma_{12}}{P_1}\right) X_1 \\ S' &= \left(1 + \frac{\sigma_{2s}}{Q}\right) S, \end{aligned} \tag{108}$$

we get an alternative representation of Y

$$Y = X_1' + X_2' + S' + N'. \tag{109}$$

Since S, X_1, X_2, Y are jointly Gaussian, and X_2' is the error in optimal estimation of X_2 given (S, X_1), we get that S', X_2', X_1', N' are independent Gaussian random variables. Conditioning on X_1', this brings us back to Costa's model, i.e.,

$$Y' = Y - E(Y|X_1') = X_2' + S' + N' \tag{110}$$

and implies that Costa's choice of auxiliary random variable applied to our notation

$$\begin{aligned} U' &= X_2' + \alpha_{costa} S' \\ \alpha_{costa} &= \frac{E(X_2'^2)}{E(X_2'^2) + N} \end{aligned} \tag{111}$$

[2]In fact, this argument holds only after we show that (102) is not merely an outer bound on the capacity region but an achievable one too, but for the sake of brevity we present this argument here rather than after establishing the direct part in Section E.2

would be optimal in our original problem too. Substituting $E(X_2'^2) = P_2 - \frac{\sigma_{12}^2}{P_1} - \frac{\sigma_{2s}^2}{Q}$ into U' we get

$$\begin{aligned} U' &= X_2 - \frac{\sigma_{12}}{P_1} \cdot X_1 - \frac{\sigma_{2s}}{Q} \cdot S + \frac{P_2 - \frac{\sigma_{12}^2}{P_1} - \frac{\sigma_{2s}^2}{Q}}{P_2 - \frac{\sigma_{12}^2}{P_1} + \frac{\sigma_{2s}^2}{Q} + N} \cdot \left(1 + \frac{\sigma_{2s}}{Q}\right) S \\ &= X_2 - \frac{\sigma_{12}}{P_1} \cdot X_1 + \alpha_{opt} \cdot S \end{aligned} \tag{112}$$

where α_{opt} is defined in (42). Define

$$U = X_2 + \alpha_{opt} \cdot S, \tag{113}$$

for simplicity we will choose U over U' although they are both optimal choices for the auxiliary random variable.

Now,

$$\begin{aligned} & I(X_1, U; Y) - I(X_1, U; S) \\ =& I(X_1; Y) + I(U; Y|X_1) - I(U; S|X_1) \end{aligned} \tag{114}$$

because X_1 and S are independent. Further,

$$\begin{aligned} I(X_1; Y) =& h(Y) - h(Y'|X_1) \\ =& h(Y) - h(X_2' + S' + N') \\ =& h(Y') - h(X_2' + S' + N') \\ =& \frac{1}{2} \log \left(\frac{E(X_1'^2) + E(X_2'^2) + E(S'^2) + N}{E(X_2'^2) + E(S'^2) + N} \right) \\ =& \frac{1}{2} \log \left(1 + \frac{P_1 \left(1 + \frac{\sigma_{12}}{P_1}\right)^2}{P_2 - \frac{\sigma_{12}^2}{P_1} - \frac{\sigma_{2s}^2}{Q} + \left(1 + \frac{\sigma_{2s}}{Q}\right)^2 Q + N} \right) \end{aligned} \tag{115}$$

because the differential entropy of a Gaussian random variable $A \sim \mathcal{N}(0, \sigma^2)$ is $\frac{1}{2} \log(2\pi e \sigma^2)$, and by definition of Y', U', X_2'. By direct application of Costa's calculation [3],

$$\begin{aligned} & I(U; Y|X_1) - I(U; S|X_1) \\ =& \frac{1}{2} \log \left(1 + \frac{E(X_2'^2)}{N} \right) \\ =& \frac{1}{2} \log \left(1 + \frac{P_2 - \frac{\sigma_{12}^2}{P_1} - \frac{\sigma_{2s}^2}{Q}}{N} \right). \end{aligned} \tag{116}$$

Substituting (46) into (115), (116), this proves that for $U = X_2 + \alpha_{opt} S$ we get

$$\begin{aligned} & I(U;Y|X_1) + I(U;S|X_1) = \left(1 + \frac{P_2(1-\rho_{2s}^2-\rho_{12}^2)}{N}\right) \\ & I(X_1,U;Y) - I(X_1,U;S) \\ = \; & \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1}+\rho_{12}\sqrt{P_2}\right)^2}{P_2(1-\rho_{2s}^2-\rho_{12}^2)+\left(\sqrt{Q}+\rho_{2s}\sqrt{P_2}\right)^2+N}\right) \\ & + \frac{1}{2}\log\left(1 + \frac{P_2(1-\rho_{2s}^2-\rho_{12}^2)}{N}\right) \end{aligned} \tag{117}$$

which concludes the proof of the direct part of Theorem 6.

F. Proof of Corollary 5

Recall the expression for the capacity region (37), and substitute $\Delta = 1 - \rho_{12}^2 - \rho_{2s}^2$ and $\rho = \rho_{2s}$.

It remains to show that

(i) one can consider $\Delta \in [\Delta_{min}, 1]$ rather than $\Delta \in [0,1]$ without loss of generality.

(ii) if $\Delta_{min} > 0$ the maximizing ρ corresponding to Δ_{min} is $-\frac{P_1(P_2+N)}{\sqrt{QP_2}(P_1+Q)}$.

(iii) for $\Delta \in (\Delta_{min}, 1]$ the maximization over ρ can be limited to either $\rho = -\sqrt{1-\Delta}$, $\rho = 0$ or any real root of $g_\Delta(\rho)$ that satisfies $\rho \in [-\sqrt{1-\Delta}, 0]$

Let $T(\Delta, P_1, P_2, Q, N)$ be the trapezoid defined by the set of non-negative rate pairs (R_c, R_2) satisfying (39). To show (i)-(ii), assume $\Delta_{min} > 0$ and note that for $\Delta \in [0, \Delta_{min}]$, the trapezoid $T(\Delta, P_1, P_2, Q, N)$ is contained in the upper bound trapezoid (5), which in turn, can be attained by taking $\rho = -\frac{P_1(P_2+N)}{\sqrt{QP_2}(P_1+Q)}$ (which is a legitimate choice only when the condition $\rho^2 + (1-\Delta-\rho)^2 \leq 1$, as stated in (38)). This condition can be met whenever $\Delta \in [0, \Delta_{min}]$.

To establish (iii), one needs to perform the maximization in (39). Deriving w.r.t. ρ yields the equation $g_\Delta(\rho) = 0$. Hence, the maximizing ρ is either a root of $g_\Delta(\rho)$ or lies on the border, i.e, $\rho = 0$ or $\rho = -\sqrt{1-\Delta}$. □

G. Proof of Theorem 7

By specializing Theorem 6 to the case where $R_2 = 0$, we get

$$C(P_1, P_2, Q, N) = \max_{\rho_{12}, \rho_{2s}} \mathbf{C}(P_1, P_2, \rho_{12}, \rho_{2s}) \tag{118}$$

where $\mathbf{C}(P_1, P_2, \rho_{12}, \rho_{2s})$ is defined in (99), and the maximization is over $\rho_{2s} \in [-1, 0], \rho_{12} \in [0,1]$, such that, $\rho_{12}^2 + \rho_{2s}^2 \leq 1$. It remains to prove that (52) and (118) are equivalent.

Recall the definition of ρ_{12}^*, ρ_{2s}^*(see (50)).

Lemma 1 *For fixed P_1, P_2, the function $\boldsymbol{C}(P_1, P_2, \rho_{12}, \rho_{2s})$ has no more than a single extremum point at ρ_{12}^*, ρ_{2s}^* yielding*

$$\boldsymbol{C}(P_1, P_2, \rho_{12}^*, \rho_{2s}^*) \;\; = \;\; \frac{1}{2}\log\left(1 + \frac{P_1}{Q}\right) + \frac{1}{2}\log\left(1 + \frac{P_2}{N}\right). \tag{119}$$

The proof of this lemma appears in Appendix I.

Now, the r.h.s. of (119) is obviously a global maximum of our target function due to Theorem 5.

Therefore, $(\rho_{12}^*, \rho_{2s}^*)$ in (50) is the single maximal point of $\mathbf{C}(P_2, P_1, \rho_{12}, \rho_{2s})$, and the r.h.s. of (119) expresses the capacity *whenever*[3] the nonnegative-definiteness constraint (106) is not met with equality. Plugging $(\rho_{12}^*, \rho_{2s}^*)$ into (106) we get that the range of parameters for which (119) is the capacity of the GGP channel is

$$\rho_{12}^* + \rho_{2s}^* \leq 1 \Rightarrow \frac{P_1(P_2+N)^2}{P_1+Q} \leq P_2 Q. \tag{120}$$

For the complementary range of parameters, i.e., $\frac{P_1(P_2+N)^2}{P_1+Q} > P_2 Q$, for which the constraint in the optimization (106) is attained with equality, i.e., $\rho_{12}^2 + \rho_{2s}^2 = 1$, we denote $\rho = \rho_{2s}$ and substitute this relation into $\mathbf{C}(P_2, P_1, \rho_{12}, \rho_{2s})$ yielding the maximization $\max_{\rho\in[-1,0]} \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1}+\sqrt{P_2}\sqrt{1-\rho^2}\right)^2}{\left(\sqrt{Q}+\sqrt{P_2}\cdot\rho\right)^2+N}\right)$. A simple derivative of the above function w.r.t. ρ shows that either the maximizing ρ lies on the boundary $\rho = 0$ or $\rho = -1$, or the real roots of the 4^{th} order polynomial $g_0(\rho)$ in (40).

□

H. Proof of Theorem 9

Following the proof of Theorem 7, one realizes that, in fact, $C(P_1, P_2, Q, N)$ (see (52)) expresses the highest achievable rate when the uninformed and informed users transmit using power levels P_1 and P_2, respectively. Therefore, to obtain the capacity formula for the sum power constraint, (31), one can let $\zeta \in [0,1]$ stand for the portion of the power P that is devoted to the informed user, substitute $P_1 = (1-\zeta)P$ and $P_2 = \zeta P$ in $C(P_1, P_2, Q, N)$, and perform the maximization over $\zeta \in [0,1]$, that is,

$$C(P, Q, N) = \max_{\zeta\in[0,1]} C((1-\zeta)P, \zeta P, Q, N). \tag{121}$$

For fixed Q, N denote

$$\begin{aligned} f_1(P_1, P_2) &= \frac{1}{2}\log\left(1+\frac{P_1}{Q}\right) + \frac{1}{2}\log\left(1+\frac{P_2}{N}\right) \\ f_2(P_1, P_2) &= \max_{\rho\in[-1,0]} \frac{1}{2}\log\left(1 + \frac{\left(\sqrt{P_1}+\sqrt{P_2}\sqrt{1-\rho^2}\right)^2}{\left(\sqrt{Q}+\sqrt{P_2}\cdot\rho\right)^2+N}\right). \end{aligned} \tag{122}$$

Thus, by (121) and by definition of $C(P_1, P_2, Q, N)$ (52) we have

$$C(P, Q, N) = \max_{\zeta\in[0,1]} \begin{cases} f_1\left((1-\zeta)P, \zeta P\right) \text{ if } \frac{(1-\zeta)(\zeta P+N)^2}{(1-\zeta)P+Q} \leq \zeta Q \\ f_2\left((1-\zeta)P, \zeta P\right) \text{ otherwise} \end{cases}. \tag{123}$$

[3]As it was shown in Lemma 1 there is a single extremum to $\mathbf{C}(\tilde{P}_2, \tilde{P}_1, \sigma_{12}, \sigma_{2s})$.

Lemma 1 and its proceeding comment yield that for all P_1, P_2,

$$f_1(P_1, P_2) \geq f_2(P_1, P_2). \tag{124}$$

Applying this to $P_1 = (1-\zeta)P$ and $P_2 = \zeta P$, and maximizing over $\zeta \in [0,1]$, we get

$$\max_{\zeta\in[0,1]} f_1\left((1-\zeta)P, \zeta P\right) \geq \max_{\zeta\in[0,1]} f_2\left((1-\zeta)P, \zeta P\right). \tag{125}$$

Now, for fixed P, the function $f_1\left((1-\zeta)P, \zeta P\right)$ is concave w.r.t. ζ with maximum at $\zeta^* = \frac{P+Q-N}{2P}$. Hence, if $\zeta^* \in [0,1]$ and $\frac{(1-\zeta^*)(\zeta^* P+N)^2}{(1-\zeta^*)P+Q} \leq \zeta^* Q$, one has

$$C(P,Q,N) \;=\; C((1-\zeta)P, \zeta P, Q, N)|_{\zeta=\zeta^*} = \frac{1}{2}\log\frac{(Q+P+N)^2}{4QN}. \tag{126}$$

It is easy to verify that the condition that $\zeta^* \in [0,1]$ and $\frac{(1-\zeta^*)(\zeta^* P+N)^2}{(1-\zeta^*)P+Q} \leq \zeta^* Q$ is equivalent to $\frac{1}{3}(N-P) + \frac{2}{3}\sqrt{P^2+PN+N^2} \leq Q \leq P+N$.

If $\zeta^* > 1$, and $\frac{(1-\zeta^*)(\zeta^* P+N)^2}{(1-\zeta^*)P+Q} \leq \zeta^* Q$, the concavity of $f_1\left((1-\zeta)P, \zeta P\right)$ w.r.t. ζ implies that it is an increasing function of ζ for $\zeta \in [0,1]$, and (125) yields

$$C(P,Q,N) \;=\; C((1-\zeta)P, \zeta P, Q, N)|_{\zeta=1} = \frac{1}{2}\log\left(1+\frac{P}{N}\right). \tag{127}$$

It is easy to verify that the condition that $\zeta^* > 1$, and $\frac{(1-\zeta^*)(\zeta^* P+N)^2}{(1-\zeta^*)P+Q} \leq \zeta^* Q$, is equivalent to $Q \geq P+N$.

For all other cases, that is, whether $\zeta^* < 0$ or $\frac{(1-\zeta^*)(\zeta^* P+N)^2}{(1-\zeta^*)P+Q} \geq \zeta^* Q$, one gets from (123),

$$C(P,Q,N) \;=\; \max_{\zeta\in[0,1]} f_2\left((1-\zeta)P, \zeta P\right), \tag{128}$$

which concludes the proof of Theorem 9.

I. Proof of Lemma 1

A necessary condition for $(\tilde{\rho}_{12}, \tilde{\rho}_{2s})$ to be an extremum point of $\mathbf{C}(P_1, \tilde{P}_2, \rho_{12}, \rho_{2s})$ is that the following equality holds

$$\left(\frac{\partial}{\partial\sigma_{12}}\mathbf{C}(P_2, P_1, \rho_{12}, \rho_{2s}) - \frac{\partial}{\partial\rho_{2s}}\mathbf{C}(P_2, P_1, \rho_{12}, \rho_{2s})\right)\Bigg|_{(\rho_{12},\rho_{2s})=(\tilde{\rho}_{12},\tilde{\rho}_{2s})} = 0 \tag{129}$$

this yields

$$\frac{-2P_1(\rho_{2s}+Q)(P_2+Q+N+2\sigma_{12}+P_1+2\sigma_{2s})(-P_1P_2+Q\sigma_{12}+\sigma_{12}^2-P_1N+\sigma_{2s}\sigma_{12}-P_1\sigma_{2s})}{QN(P_2P_1+2P_1\sigma_{2s}+P_1Q+P_1N-\sigma_{12}^2)^2} \;=\; 0, \tag{130}$$

i.e., for $(\tilde{\rho}_{12}, \tilde{\rho}_{2s})$ to be an extremum point, at least one of the following should be met: either $\tilde{\sigma}_{2s} = -Q$ or $\tilde{\sigma}_{12} + \tilde{\sigma}_{2s} = -\frac{1}{2}(\tilde{P}_1 + \tilde{P}_2 + Q + N)$ or $\tilde{\sigma}_{2s} = \frac{(\tilde{\sigma}_{12})^2 + Q\tilde{\sigma}_{12} - \tilde{P}_1(\tilde{P}_2+N)}{\tilde{P}_1 - \tilde{\sigma}_{12}}$. We shall next show that it is only the latter that is met at an extremum point.

- $\underline{\tilde{\sigma}_{2s} = -Q}$: no extremum point satisfies this, as

$$\left.\frac{\partial^2 \mathbf{C}(\tilde{P}_1, \tilde{P}_2, \sigma_{12}, \sigma_{2s})}{\partial \sigma_{12}^2}\right|_{\sigma_{2s}=-Q} = 0. \tag{131}$$

- $\underline{\tilde{\sigma}_{12} + \tilde{\sigma}_{2s} = -\frac{1}{2}(\tilde{P}_1 + \tilde{P}_2 + Q + N)}$: substituting this relation into $\mathbf{C}(\tilde{P}_1, \tilde{P}_2, \sigma_{12}, \sigma_{2s})$ and deriving w.r.t. σ_{2s}, yields the suspected extremum $(\tilde{\sigma}_{12}, \tilde{\sigma}_{2s}) = \left(\frac{Q-\tilde{P}_1-\tilde{P}_2-N}{2}, -Q\right)$ and from (131) we have again that this cannot be a extremum.

As for the third condition $\tilde{\sigma}_{2s} = \frac{(\tilde{\sigma}_{12})^2+Q\tilde{\sigma}_{12}-\tilde{P}_1(\tilde{P}_2+N)}{\tilde{P}_1-\tilde{\sigma}_{12}}$: substituting this relation into $\mathbf{C}(\tilde{P}_1, \tilde{P}_2, \sigma_{12}, \sigma_{2s})$, we get

$$\mathbf{C}(\tilde{P}_1, \tilde{P}_2, \tilde{\sigma}_{12}, \tilde{\sigma}_{2s}) \;=\; -\frac{(Q\tilde{\sigma}_{12}^2 - \tilde{P}_1^2\tilde{P}_2 - \tilde{P}_1^2 N + \tilde{P}_1\tilde{\sigma}_{12}^2)(-\tilde{P}_2 + \tilde{P}_1 + Q - N)}{QN(\tilde{P}_1 - \tilde{\sigma}_{12})^2} \tag{132}$$

deriving w.r.t. σ_{12} yields the suspected maximum

$$\sigma_{12}^* = -\sigma_{2s}^* = \frac{\tilde{P}_1(\tilde{P}_2 + N)}{\tilde{P}_1 + Q} \tag{133}$$

substituting this point yields

$$\mathbf{C}(\tilde{P}_1, \tilde{P}_2, \sigma_{12}^*, \sigma_{2s}^*) \;=\; \frac{1}{2}\log\left(1 + \frac{\tilde{P}_1}{Q}\right) + \frac{1}{2}\log\left(1 + \frac{\tilde{P}_2}{N}\right). \tag{134}$$

□

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), 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), 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 Gel'fand-Pinsker channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT'04)*, (Chicago, IL), p. 448, July 2004.

[15] P. Moulin and J. A. OSullivan, "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, pp. 1691–1706", July 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), 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), October 2004.

[23] S. Kotagiri and J. N. Laneman, "Multiple access channels with state information known to some encoders." preprint, 2006.

[24] Y. Cemal and Y. Steinberg, "The multiple–access channel with partial state information at the encoders," *IEEE Transactions on Information Theory*, vol. 51, pp. 3392–4003, November 2005.

[25] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, "Cooperative encoding with asymmetric state information at the transmitters," in *Proc. Allerton Conf. Communications, Control, and Computing*, (Monticello, IL), September 2006.

[26] T. Weissman and N. Merhav, "Coding for the feedback Gel'fand–Pinsker channel and the feedforward Wyner–Ziv source," *IEEE Transactions on Information Theory*, vol. 52, pp. 4207–4211, September 2006.

[27] 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.

[28] 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.

[29] 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), March 2006.
[30] A. Jovičić and P. Viswanath, "Cognitive radio: An information-theoretic perspective." preprint, 2006.
[31] A. Høst-Madsen, "Capacity bounds for cooperative diversity," *IEEE Transactions on Information Theory*, vol. 52, pp. 1522–144, April 2006.
[32] A. Somekh-Baruch, S. Shamai (Shitz), and S. Verdú, "Cooperative encoding with asymmetric state information at the transmitters." preprint.
[33] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Systems Technical Journal*, vol. 52, pp. 1037–1076, September 1973.
[34] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *IEEE transactions Information Theory*, vol. 52, pp. 5468–5474, December 2006.
[35] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.