A More General Version of the Costa Theorem

Nima S. Anzabi-Nezhad, Ghosheh Abed Hodtani, and Mohammad Molavi Kakhki
Department of Electrical Engineering, Ferdowsi University of Mashhad, Iran
nima.anzabi@gmail.com, ghodtani@gmail.com, molavi@um.ac.ir
Corresponding author: nima.anzabi@gmail.com

***Abstract*- In accordance with the Costa theorem, the interference which is independent of the channel input and known non-causally at the transmitter, does not affect the capacity of the Gaussian channel. In some applications, the known interference depends on the input and hence has some information. In this paper, we study the channel with input dependent interference and prove a capacity theorem that not only subsumes the Costa theorem but also explains interestingly interpretable impact of correlation between side information and channel input on the capacity.**

***Index Terms*—** Gaussian channel capacity, correlated side information, two sided state information.

I. INTRODUCTION

Side information channel has been actively studied since its initiation by Shannon [1]. Studying the problem of coding for computer memories with defective cells, Kusnetsov-Tsybakov [2] introduced the channel with state information known non causally at the transmitter. Gel'fand-Pinsker (GP) [3] determined the capacity of this channel (channel with channel state information (CSI) known non-causally at the transmitter). Heegard-El Gamal [4] obtained the capacity when the CSI is known only at the receiver. Cover-Chiang [5] extended these results to a general case where correlated two-sided state information are available at the transmitter and at the receiver. Costa [6] obtained a Gaussian version of the GP theorem. There are many other important researches in the literature, e.g.[7]-[9]. The results for the single user channel have been generalized possibly to multi user channels, at least in special cases [10]-[15]. The channels with partial CSI have been investigated in [16]-[18].

Our Work: Costa in his famous "Writing on dirty paper" [6] examines Gaussian channels with state information known non causally at the transmitter, where the channel state is Gaussian additive interference at the receiver- as seen in Fig.1. Costa shows that the capacity of the channel is surprisingly the same as the capacity of the channel when there is no known interference. One important feature of Costa theorem is that in the definition of the channel, there is no condition for correlation between the input and the interference known at the transmitter. But Costa argues that assuming these two random variables independent is optimum for maximizing the rate.

Manuscript received 05-August-2013 and revised 16-Oct-2013 *ISSN: 2322-3936*
Accepted on 08- Nov-2013

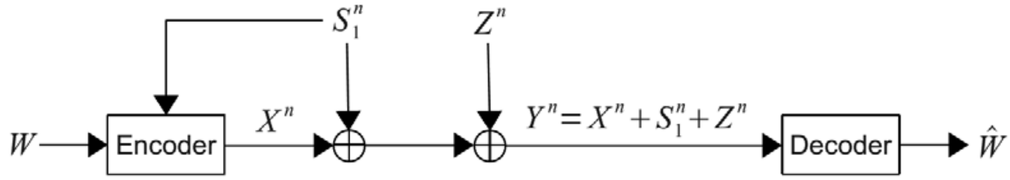

Fig. 1. Gaussian channel with additive interference known at the transmitter.

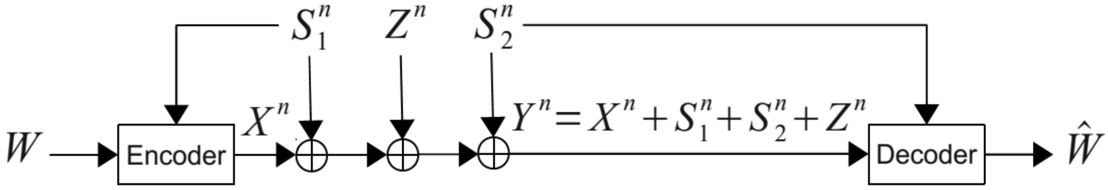

Fig. 2. Gaussian channel with correlated side information known at the transmitter and at the receiver.

Now one question arises, regarding the above assumption. If the Gaussian channel, with side information known at the transmitter, is defined with a specific correlation between the channel input and the side information, how can we find the capacity of the channel? What is the capacity of the channel if the channel input and the side information, inherently, cannot be independent of each other? And in this case how and how much does the dependency between the channel input and the side information affect the capacity? This problem not only has theoretically importance but there are practically important situations that could not be analyzed unless we answer these questions. In Section III, we point out some of these situations.

Another related question is about the side information known non-causally at the receiver (if exists as in Fig.2): How does the side information at the receiver, with correlation to the channel input and side information known at the transmitter affect the capacity?

We, in another paper, examined the Gaussian channel in presence of two-sided correlated state information in a different and more limited situation [20].

In this paper, to answer the above questions in detail, we consider the Cover-Chiang unifying theorem and show that the dependency between the channel input and the side information known at the transmitter reduces the capacity and also show that in the case of independency of side information known at the receiver and the channel noise, this side information has no effect on the channel capacity. In Section II, we review the Cover-Chiang theorem, the Gel'fand-Pinsker theorem and the Costa's dirty paper theorem. In Section III, we, first, scrutinize the dirty paper channel and then define our Gaussian channel carefully and derive a capacity theorem for it in Section IV. Then, we prove our capacity theorem in a fully detailed manner. Then in Section V, an explaining and illustrating numerical comparison is done between our capacity and the Costa capacity. Section VI includes a conclusion. In appendix we prove two lemmas that are necessary for our theorem.

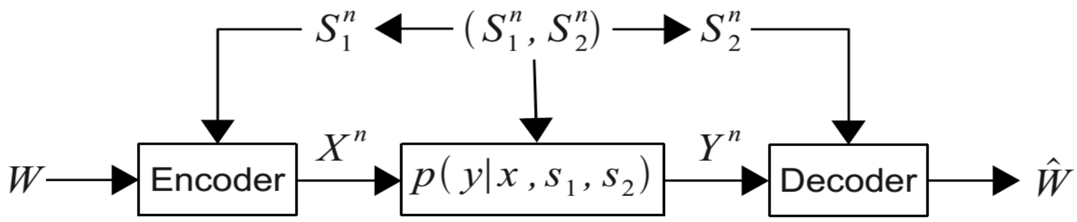

Fig. 3. Channel with side information available non-causally at the transmitter and at the receiver.

II. PRELIMINARIES

A. Cover-Chiang Theorem

Fig.3 shows a channel with side information known at the transmitter and at the receiver. X^n and Y^n are the transmitted and received sequences respectively. The sequences S_1^n and S_2^n are the side information known non-causally at the transmitter and at the receiver respectively. The transition probability of the channel $p(y \mid x, s_1, s_2)$ depends on the input X, the side information S_1 and S_2. It can be shown [5] that if the channel is memoryless and the sequences (S_1^n, S_2^n) is independent and identically distributed (i.i.d.) random variables under $p(s_1, s_2)$, then the capacity of the channel is

$$C = \max_{p(u,x|s_1)} [I(U;S_2,Y) - I(U;S_1)] \tag{1}$$

where the maximum is over all distributions:

$$p(y,x,u,s_1,s_2) = p(y \mid x,s_1,s_2)p(u,x \mid s_1)p(s_1,s_2) \tag{2}$$

and U is an auxiliary random variable for conveying the information of the known S_1^n into X^n.

It is important to note that the Markov chains:

$$S_2 \to S_1 \to UX \tag{3}$$

$$U \to XS_1S_2 \to Y \tag{4}$$

are satisfied for all above distributions (2).

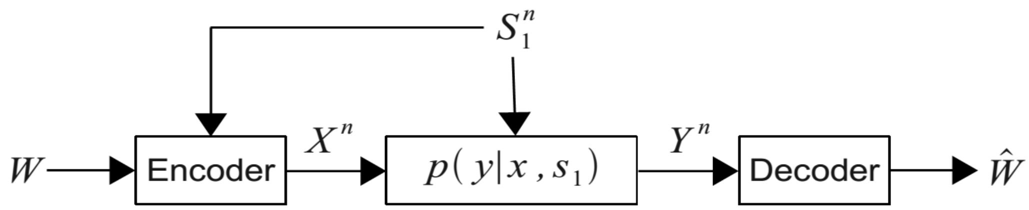

Fig. 4. Channel with side information known at the transmitter.

B. Gel'fand-Pinsker Theorem

Cover-Chiang theorem may be considered in special cases in which S_1 and/or S_2 is empty. The situation $S_2 = \phi$ (no side information at the receiver) leads to the Gel'fand-Pinsker theorem [3].

The memoryless channel with transition probability $p(y \mid x, s_1)$ and the side information sequence S_1^n (which are i.i.d. : $p(s_1)$) known non-causally at the transmitter (Fig.4) has the capacity

$$C = \max_{p(u,x|s_1)} [I(U;Y) - I(U;S_1)] \tag{5}$$

for all distributions;

$$p(y,x,u,s_1) = p(y \mid x,s_1)p(u,x \mid s_1)p(s_1) \tag{6}$$

where U is an auxiliary random variable

C. Costa's Writing on Dirty Paper

Costa [6] examined the Gaussian version of the channel with side information known at the transmitter (Fig.1). It is seen that the side information is considered as an additive interference at the receiver. Costa derived the capacity by using the result of Gel'fand-Pinsker theorem extended to random variables with continuous alphabets. The proof briefly is as follows: In Costa channel S_1^n is a sequence of Gaussian i.i.d. random variables with power Q_1. The transmitted sequence X^n is assumed to have the power constraint $E\{X^2\} \le P$. The output $Y^n = X^n + S_1^n + Z^n$ where Z^n is the sequence of white Gaussian noise with zero mean and power N, i.e., $Z : \mathrm{N}(0, N)$ and independent of (X, S_1). Costa established the capacity by obtaining a lower bound and an upper bound and proving the equality of these two bounds. Although for the Costa channel, no restriction has been imposed on the correlation between X and S_1, the achievable rate of $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ is obtained by taking S_1 and X independent and the auxiliary random variable U in (6) as $U = \alpha S_1 + X$. On the other hand, it can be shown that:

$$C \leq \max_{p(x|s_1)} \left[I(X,Y \mid S_1) \right] = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \tag{7}$$

so $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ is an upper bound for the capacity of channel and then the capacity of channel.

III. The Gaussian Channel in the Presence of Two-sided Input Dependent State Information

In this section we, first, examine the Costa's "dirty paper" channel and show that the "dirty paper" capacity is unable to describe some important theoretical and practical situations. Then, we introduce a Gaussian channel in the presence of two-sided input dependent state information. In the next section we state and prove a theorem about the capacity of this channel.

A. Costa's Dirty Paper Channel:

We can scrutinize the Costa's "dirty paper" channel (Fig.1) with properties C.1-C.3 below:

C.1: S_1^n is a sequence of Gaussian independent and identically distributed random variables with distribution $S_1 : \mathrm{N}(0, Q_1)$.

C.2: The transmitted sequence X^n is assumed to have the power constraint $E\{X^2\} \leq P$.

C.3: The output $Y^n = X^n + S_1^n + Z^n$ where Z^n is the sequence of white Gaussian noise with zero mean and power N ($Z : \mathrm{N}(0, N)$) and independent of both X and S_1.

It is readily seen that the distributions $p(y, x, u, s_1)$ having the above three properties are in the form of (6). We denote the set of all this $p(y, x, u, s_1)$ with Π_C.

It is notable in the definition of the channel that there is no condition for the correlation between X and S_1. But Costa shows that the situation that S_1 is independent of X, and the auxiliary random variable U is designed as linear combination of X and S_1, is optimum and maximizes the transmitting rate. So we consider the set Π'_C of all distributions $p'(y, x, u, s_1)$ that moreover have the properties C.4 and C.5 below:

C.4: X is a zero mean Gaussian random variable with the maximum average power of P and *independent* of S_1.

C.5: The auxiliary random variable U takes the linear form $U = \alpha S_1 + X$.

As mentioned before, in Gaussian channel defined with the set Π_C of distributions $p(y, x, u, s_1)$ with properties C.1-C.3, there is no condition for the correlation between the channel input X and the

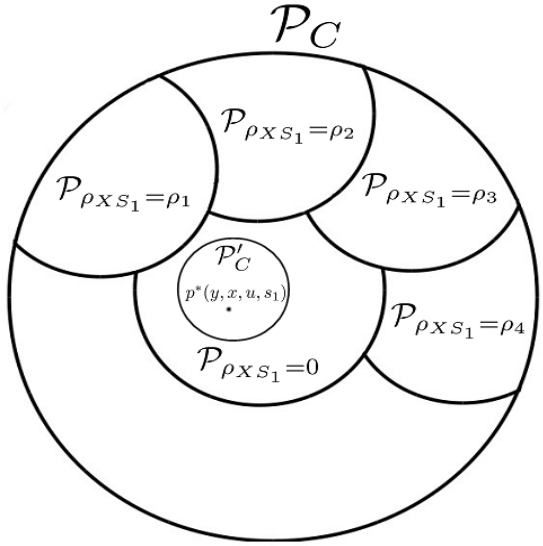

Fig. 5. Partitioning Π_C into $\Pi_{\rho_{XS_1}}$'s. $p^*(y,x,u,s_1)$ is the optimum distribution for the Costa channel.

side information S_1. So $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$ is the capacity of a channel in which the side information S_1 can be freely correlated to the channel input X; Specially the maximum rate is achieved when S_1 and X are independent of each other. Now suppose that we partition the set Π_C (of all distributions $p(y,x,u,s_1)$ with properties C.1-C.3) into subsets $\Pi_{\rho_{XS_1}}$ of distributions $p(y,x,u,s_1)$ in Π_C with a specific correlation coefficient ρ_{XS_1} (Fig.5). It is obvious that Π'_C (the set of distributions with properties C.1-C.4) is a subset of $\Pi_{\rho_{XS_1}=0}$ and so the optimum distribution that results in the capacity of the Costa channel does not belong to any other partitions. So it is clear that if a channel is defined with random variables (Y,X,U,S_1): $p(y,x,u,s_1)$ in $\Pi_{\rho_{XS_1}}$ and $\rho_{XS_1}\neq 0$, the Costa theorem can not be used for the capacity of the channel.

While in most of models studied in the literature, the channel state and channel input are assumed independent, there are practically important situations, in which the state and input must be considered correlated to each other. In these situations, the channel state (interference, here) is, inherently, a signal with information, for example, the information-bearing sequences of some other transmitter. For example, we may consider a cognitive interference channel, in which the transmitted sequence of one transmitter is known interference for another transmitter and may be dependent to the transmitted sequence of that transmitter. For another example we can consider the MIMO broadcast channel in which an interference for a user is transmitted sequence for the other user and all sequences are produced at one transmitter and therefore can not be independent of each other. In an other application of channels with known channel state at the transmitter, the transmitter wishes to send not only the message, but also the channel state information to the receiver (in [19],[21]-[23]). In this scenario, the correlation between the channel input and channel state is used to trade off between the achievable rate and the error of state estimation. For another example we can consider a measurement system in which the observer sends a measuring signal which itself can interfere or affects the interference of the system and it is obvious that this interference can not be independent of the input signal. Other communication scenarios in which the channel input and the state information may be dependent to each other can be found in the literature such as [9].

B. Definition of the Channel:

Here we define a Gaussian channel with following major modifications:

1) The correlation coefficient between X and S_1 (ρ_{XS_1}) is specified.

2) We suppose that the Gaussian side information S_2 known at the receiver, exists and is correlated to both X and S_1.

It is important to note that as we prove in Lemma 1 in Appendix, assuming the input random variable X correlated to S_1 and S_2 with specific correlation coefficients, does not impose any restriction on X's own distribution and the distribution of X is still free to choose.

Consider the Gaussian channel depicted in Fig.2. The side information at the transmitter and at the receiver is considered as additive interferences at the receiver. Our channel is defined with properties MC.1-MC.4 below:

MC.1: (S_1^n, S_2^n) are i.i.d. sequences with zero mean and jointly Gaussian distributions with power $\sigma_{S_1}^2 = Q_1$ and $\sigma_{S_2}^2 = Q_2$ respectively (so we have $S_1 \sim \mathrm{N}(0,Q_1)$ and $S_2 \sim \mathrm{N}(0,Q_2)$).

MC.2: Random variables (X, S_1, S_2) have the covariance matrix $\mathbf{K}$:

$$\mathbf{K} = E\left\{\begin{bmatrix} X^2 & XS_1 & XS_2 \\ XS_1 & S_1^2 & S_1S_2 \\ XS_2 & S_1S_2 & S_2^2 \end{bmatrix}\right\} \tag{8}$$

$$= \begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_X\sigma_{S_2}\rho_{XS_2} \\ \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_{S_1}^2 & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} \\ \sigma_X\sigma_{S_2}\rho_{XS_2} & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} & \sigma_{S_2}^2 \end{bmatrix} \tag{9}$$

and moreover X^n is assumed to have the constraint $\sigma_X^2 \le P$. All values in $\mathbf{K}$ except σ_X, are fixed and must be considered as parts of *the definition of the channel*. Moreover we note that since $\mathbf{K}$ is positive semidefinite, $\det(\mathbf{K})$ (10) is nonnegative and thus is an increasing function of σ_X ; therefore when reaches to its maximum, we have $\sigma_X^2 = P$.

$$\det(\mathbf{K}) = \sigma_X^2 \sigma_{S_1}^2 \sigma_{S_2}^2 (1 - \rho_{XS_1}^2 - \rho_{XS_2}^2 - \rho_{S_1S_2}^2 + 2\rho_{XS_1}\rho_{XS_2}\rho_{S_1S_2}) \tag{10}$$

MC.3: The output sequence $Y^n = X^n + S_1^n + S_2^n + Z^n$, where Z^n is the sequence of white Gaussian noise with zero mean and power N ($Z : \mathrm{N}(0, N)$) and independent of X, S_1 and S_2.

MC.4: (X, U, S_1, S_2) form the Markov Chain $S_2 \to S_1 \to UX$. (We note that as mentioned earlier, this Markov chain (3) is satisfied by all distributions $p(y, x, u, s_1, s_2)$ in the form of (2) in Cover-Chiang capacity theorem and is physically acceptable).

Since this Markov chain results in the weaker Markov chain $S_2 \to S_1 \to X$, as we prove in Lemma 2 in Appendix, this property implies that in covariance matrix $\mathbf{K}$ we have:

$$\rho_{XS_2} = \rho_{XS_1}\rho_{S_1S_2} \tag{11}$$

It is readily seen that all distributions $p(y, x, u, s_1, s_2)$ specified with MC.1-MC.4 are in the form of (2), so we can use the extended version of Cover-Chiang theorem to random variables with continuous alphabets about the capacity of this channel. We denote the set of all this distributions $p(y, x, u, s_1, s_2)$ with $\Pi_{\rho_{XS_1}}$ (again).

Comparing our channel (distributions in $\Pi_{\rho_{XS_1}}$ defined with MC.1-MC.4) with Costa channel (distributions in Π_C defined with C.1-C.3), a question may arise: (if we ignore S_2,) what is the relationship between capacities of these channels? To answer this question we consider that (Fig.5):

$$\Pi_C = \bigcup_{\rho_{XS_1}} \Pi_{\rho_{XS_1}} \tag{12}$$

So if C_M be the capacity of the channel defined with MC.1-MC.4, we can write:

$$C_{Costa} = \max_{S_2=0, \rho_{XS_1}} C_M. \tag{13}$$

We will show that the situation that (X, S_1, S_2) are jointly Gaussian and the auxiliary random variable U is designed as linear combination of X and S_1, is optimum and maximizes the transmitting rate. So we consider the set $\Pi^*_{\rho_{XS_1}}$, a subset of $\Pi_{\rho_{XS_1}}$, as the set of all $p^*(y, x, u, s_1, s_2)$ that have the properties MC.5 and MC.6 below, in addition to MC.1-MC.4 (although the channel is defined only with MC.1-MC.4):

MC.5: Random variables (X, S_1, S_2) are jointly Gaussian distributed. Moreover X is with zero mean and has the maximum power of P (so we have $X \sim N(0, P)$).

Naming the covariance matrix in this special case as $\mathbf{K}^*$, for simplicity, by defining $A_i \cong E\{XS_i\} = \sigma_X \sigma_{S_i} \rho_{XS_i}, i = 1,2$ and $B \cong E\{S_1 S_2\} = \sigma_{S_1} \sigma_{S_2} \rho_{S_1 S_2}$, we rewrite:

$$\mathbf{K}^* = \begin{bmatrix} P & A_1 & A_2 \\ A_1 & Q_1 & B \\ A_2 & B & Q_2 \end{bmatrix}. \tag{14}$$

MC.6: Following Costa, we consider U in the form of linear combination of X and S_1 (*but here X and S_1 are correlated to each other)*:

$$U = \alpha S_1 + X. \tag{15}$$

It is clear that the set of distributions $\Pi^*_{\rho_{XS_1}}$ (defined in MC.1-MC.6) and their marginal and conditional distributions are subsets of corresponding sets of distributions $\Pi_{\rho_{XS_1}}$ defined in MC.1-MC.5).

In the next section, for summarizing our expressions, we use the minors of $\mathbf{K}^*$; so we define:

$$d_P \cong \begin{vmatrix} Q_1 & B \\ B & Q_2 \end{vmatrix} = \sigma_{S_1}^2 \sigma_{S_2}^2 \left(1 - \rho_{S_1 S_2}^2\right) \tag{16}$$

$$d_{A_1} \cong \begin{vmatrix} A_1 & B \\ A_2 & Q_2 \end{vmatrix} = \sigma_X \sigma_{S_1} \sigma_{S_2}^2 \left(\rho_{XS_1} - \rho_{XS_2} \rho_{S_1 S_2}\right) \tag{17}$$

$$d_{A_2} \cong \begin{vmatrix} A_1 & Q_1 \\ A_2 & B \end{vmatrix} = \sigma_X \sigma_{S_1}^2 \sigma_{S_2} \left(\rho_{XS_1} \rho_{S_1 S_2} - \rho_{XS_2}\right) \tag{18}$$

$$d_{Q_1} \cong \begin{vmatrix} P & A_2 \\ A_2 & Q_2 \end{vmatrix} = \sigma_X^2 \sigma_{S_2}^2 \left(1 - \rho_{XS_2}^2\right) \tag{19}$$

$$d_{Q_2} \cong \begin{vmatrix} P & A_1 \\ A_1 & Q_1 \end{vmatrix} = \sigma_X^2 \sigma_{S_1}^2 \left(1 - \rho_{XS_1}^2\right) \tag{20}$$

$$d_B \cong \begin{vmatrix} P & A_1 \\ A_2 & B \end{vmatrix} = \sigma_X^2 \sigma_{S_1} \sigma_{S_2} \left(\rho_{S_1 S_2} - \rho_{XS_1} \rho_{XS_2}\right) \tag{21}$$

$$\begin{aligned} D &\cong P d_P - A_1 d_{A_1} + A_2 d_{A_2} \\ &= \left(P Q_1 Q_2 + 2 A_1 A_2 B\right) - \left(P B^2 + Q_1 A_2^2 + Q_2 A_1^2\right). \end{aligned} \tag{22}$$

And it is seen from (11) that $d_{A_2} = 0$.

IV. OUR CAPACITY THEOREM

In this section we state and prove the capacity theorem for the Gaussian channel defined in section III.

Theorem 1: The Gaussian channel defined with properties MC.1-MC.4 has the capacity:

$$C_M = \frac{1}{2}\log\left(1+\frac{P}{N}\left(1-\rho_{XS_1}^2\right)\right) \tag{23}$$

Corollary 1: It is seen that for the Gaussian channel with side information known non-causally at the transmitter, the dependency between the channel input and the side information, decreases the capacity of the channel.

Corollary 2: As mentioned in the previous section (13), we can obtain Costa capacity by maximizing C_M with $\rho_{XS_1} = 0$.

Corollary 3: It is seen that for our channel, in which S_2 is independent of Z , S_2 has no effect on the capacity of the channel.

Proof of Theorem 1: To prove the theorem, we first use extended version of Cover-Chiang capacity (1) for random variables with continuous alphabets to show that C_M (23) is a lower bound for the capacity of the channel (theorem 2), then we show that C_M is an upper bound for the capacity too (theorem 3), so C_M is the capacity of the channel.

First of all, we note that the GP theorem in Subsect. 2.2 (and similarly Cover-Chiang theorem in Subsect. 2.1), is correct for distributions (6) (or (2)) which are restricted to specific correlation between X and S_1. For proving this, it is sufficient to notice that in encoder, the sequence X^n can be produced strongly jointly typical with sequences (U^n, S^n) using the given distribution $p(x \mid u, s_1)$.

The First Part of the Proof (Achievability):

Theorem 2 (A Lower Bound for the Capacity of the Channel): The capacity of the Gaussian channel defined with properties MC.1-MC.4 has the lower bound:

$$R(\alpha^*) = \max_{\alpha} R(\alpha) = \frac{1}{2}\log\left(1+\frac{P}{N}\left(1-\rho_{XS_1}^2\right)\right) \tag{24}$$

where

$$R(\alpha)=\frac{1}{2}\log\left(\frac{d_{Q_2}\left(Q_2N+d_{Q_1}+d_P+2d_{A_1}\right)}{Q_1\left((\alpha-1)^2D+N\left(\alpha^2d_P+2\alpha d_{A_1}+d_{Q_1}\right)\right)}\right) \tag{25}$$

and

$$\alpha^*=\frac{D-Nd_{A_1}}{D+Nd_P}=1-\frac{1+\sqrt{\frac{P}{Q_1}}\rho_{XS_1}}{1+\frac{P}{N}\left(1-\rho_{XS_1}^2\right)}. \tag{26}$$

Proof of Theorem 2: Using the extension of Cover-Chiang capacity theorem (1) for random variables with continuous alphabets, the capacity of our channel can be written as:

$$C_M=\max_{p(u,x|s_1)}\left[I(U;Y,S_2)-I(U;S_1)\right] \tag{27}$$

where the maximum is over all distributions $p(y,x,u,s_1,s_2)$ in $\Pi_{\rho_{XS_1}}$ (with properties MC.1-MC.4). And since $\Pi^*_{\rho_{XS_1}}\subseteq\Pi_{\rho_{XS_1}}$ we have:

$$C_M\geq\max_{p^*(u,x|s_1)}\left[I(U;Y,S_2)-I(U;S_1)\right] \tag{28}$$

$$=\max_{p^*(u|x,s_1)p^*(x|s_1)}\left[I(U;Y,S_2)-I(U;S_1)\right] \tag{29}$$

$$=\max_{\alpha}\left[I(U;Y,S_2)-I(U;S_1)\right] \tag{30}$$

where the expression $I(U;Y,S_2)-I(U;S_1)$ in (30) must be computed for the distributions in $\Pi^*_{\rho_{XS_1}}$ (defined with properties MC.1-MC.6). Naming this expression as $R(\alpha)$, we have:

$$C_M\geq\max_{\alpha}R(\alpha)=R(\alpha^*), \tag{31}$$

then $R(\alpha^*)$ is a lower bound for the capacity of the channel. For computing $R(\alpha^*)$, we write:

$$I(U;Y,S_2)=H(U)+H(Y,S_2)-H(U,Y,S_2) \tag{32}$$

and

$$I(U;S_1)=H(U)+H(S_1)-H(U,S_1). \tag{33}$$

For computing $H(Y,S_2)$:

$$H(Y,S_2)=\frac{1}{2}\log\left((2\pi e)^2\det(cov(Y,S_2))\right) \tag{34}$$

where

$$\operatorname{cov}[Y,S_2]=[e_{ij}]_{2\times2} \tag{35}$$

with

$$\left.\begin{aligned} e_{11} &= P+Q_1+Q_2+N+2A_1+2A_2+2B, \\ e_{12} &= e_{21} = A_2+B+Q_2 \quad \text{and} \quad e_{22}=Q_2 \end{aligned}\right\} \tag{36}$$

then

$$\det\left(cov\left(Y,S_2\right)\right) = Q_2 N + d_{Q_1} + d_P + 2d_{A_1}. \tag{37}$$

And for $H\left(U,Y,S_2\right)$:

$$H\left(U,Y,S_2\right) = \frac{1}{2}\log\left(\left(2\pi e\right)^3 \det\left(cov\left(U,Y,S_2\right)\right)\right) \tag{38}$$

where

$$\text{cov}[U,Y,S_2] = [e_{ij}]_{3\times 3} \tag{39}$$

$$\left.\begin{aligned} & e_{11} = \alpha^2 Q_1 + P + 2\alpha A_1, \\ & e_{12} = e_{21} = \left(\alpha+1\right)A_1 + \alpha Q_1 + \alpha B + P + A_2, \\ & e_{13} = e_{31} = \alpha B + A_2, \\ & e_{22} = P+Q_1+Q_2+N+2A_1+2A_2+2B, \\ & e_{23} = e_{32} = A_2+B+Q_2 \quad and \quad e_{33}=Q_2 \end{aligned}\right\} \tag{40}$$

and after the manipulations we have:

$$\det(cov(U,Y,S_2)) = N(\alpha^2 d_P + 2\alpha d_{A_1} + d_{Q_1}) + (\alpha-1)^2 D. \tag{41}$$

And for $H\left(S_1\right)$:

$$H\left(S_1\right) = \frac{1}{2}\log\left(\left(2\pi e\right)Q_1\right). \tag{42}$$

And $H\left(U,S_1\right)$:

$$H\left(U,S_1\right) = \frac{1}{2}\log\left(\left(2\pi e\right)^2 \det\left(cov\left(U,S_1\right)\right)\right) \tag{43}$$

where

$$cov\left(U,S_1\right) = \begin{bmatrix} \alpha^2 Q_1 + P + 2\alpha A_1 & \alpha Q_1 + A_1 \\ \alpha Q_1 + A_1 & Q_1 \end{bmatrix} \tag{44}$$

and its determinant:

$$\det\left(cov\left(U,S_1\right)\right) = d_{Q_2}. \tag{45}$$

Substituting (34), (38),(42) and (43) in (32) and (33), we obtain (25) and after maximizing it over α we conclude

$$\alpha^* = \frac{D - Nd_{A_1}}{D + Nd_P}. \tag{46}$$

Now, if we compute $R(\alpha^*)$ by putting (46) into (25) and then rewrite the resulted expression in terms of σ_X , σ_{S_1} , σ_{S_2} , ρ_{XS_1} , ρ_{XS_2} and $\rho_{S_1S_2}$ by (16)-(22) and take into account the equality (11), we finally conclude (24).

The Second Part of the Proof (Converse): In Theorem 3 we derive an upper bound for the capacity of the channel:

Theorem 3 (An Upper Bound for the Capacity of the Channel): C_M (23) is an upper bound for the capacity of our Gaussian channel defined with properties MC.1-MC.4.

Proof of the Theorem 3: for all distributions $p(y, x, u, s_1, s_2)$ in $\Pi_{\rho_{XS_1}}$ defined with properties MC.1-MC.4, we have:

$$I(U;Y,S_2) - I(U;S_1) = -H(U \mid Y,S_2) + H(U \mid S_1) \tag{47}$$

$$\leq -H(U \mid Y,S_1,S_2) + H(U \mid S_1) \tag{48}$$

$$= -H(U \mid Y,S_1,S_2) + H(U \mid S_1,S_2) \tag{49}$$

$$= I(U;Y \mid S_1,S_2) \tag{50}$$

$$\leq I(X;Y \mid S_1,S_2) \tag{51}$$

where (48) follows from the fact that conditioning reduces entropy and (49) follows from Markov chain $S_2 \to S_1 \to UX$ and (51) from Markov chain $U \to XS_1S_2 \to Y$ which are satisfied by every distribution in the form of (2), including the distributions in the set $\Pi_{\rho_{XS_1}}$. Now from (1) and (51) we can write:

$$C_M = \max_{p(u,x|s_1)} \left[I(U;Y,S_2) - I(U;S_1) \right] \tag{52}$$

$$\leq \max_{p(x|s_1)} \left[I(X;Y \mid S_1,S_2) \right]. \tag{53}$$

(53) says that the capacity of the channel cannot be greater than the capacity when both S_1 and S_2 are available at both the transmitter and the receiver, which is physically acceptable and predictable. Then for computing (53) we write:

$$I(X;Y \mid S_1,S_2) = H(Y \mid S_1,S_2) - H(Y \mid X,S_1,S_2) \tag{54}$$

$$= H(X + Z \mid S_1,S_2) - H(Z) \tag{55}$$

$$= H((X + Z),S_1,S_2) - H(S_1,S_2) - H(Z). \tag{56}$$

So maximum in (53) occurs when $H((X+Z), S_1, S_2)$ has its maximum value. Thus, according to Gaussianness of S_1, S_2 and Z, when (53) reaches to its maximum, (X, S_1, S_2) are jointly Gaussian and X has its maximum power of P and it means that $I(X;Y \mid S_1, S_2)$ must be computed for distribution $p^*(y, x, s_1, s_2)$ defined with properties MC.1-MC.6. Naming this maximum value as $I^*(X;Y \mid S_1, S_2)$, we have:

$$C_M \leq I^*(X;Y \mid S_1, S_2) \tag{57}$$

For computing $I^*(X;Y \mid S_1, S_2)$ we write:

$$H((X+Z), S_1, S_2) = \frac{1}{2}\log\left((2\pi e)^3 \det(cov((X+Z), S_1, S_2))\right) \tag{58}$$

where

$$cov((X+Z), S_1, S_2) = E\left\{\begin{bmatrix} (X+Z)^2 & (X+Z)S_1 & (X+Z)S_2 \\ (X+Z)S_1 & S_1^2 & S_1S_2 \\ (X+Z)S_2 & S_1S_2 & S_1^2 \end{bmatrix}\right\} \tag{59}$$

$$= \begin{bmatrix} P+N & A_1 & A_2 \\ A_1 & Q_1 & B \\ A_2 & B & Q_2 \end{bmatrix} \tag{60}$$

and its determinant:

$$\det(cov((X+Z), S_1, S_2)) = D + Nd_P, \tag{61}$$

and

$$H(Z) = \frac{1}{2}\log(2\pi e N) \tag{62}$$

and

$$H(S_1, S_2) = \frac{1}{2}\log\left((2\pi e)^2 d_P\right). \tag{63}$$

Now after substituting (61) in (58), and from (62) and (63) we have:

$$I^*(X;Y \mid S_1, S_2) = \frac{1}{2}\log\left(\frac{D + Nd_P}{Nd_P}\right). \tag{64}$$

Now if we rewrite (64) in terms of σ_X, σ_{S_1}, σ_{S_2}, ρ_{XS_1}, ρ_{XS_2} and $\rho_{S_1S_2}$ by (16)-(22) and take into account the equality (11), we finally conclude that:

$$I^*(X;Y \mid S_1, S_2) = \frac{1}{2}\log\left(1 + \frac{P}{N}\left(1 - \rho_{XS_1}^2\right)\right). \tag{65}$$

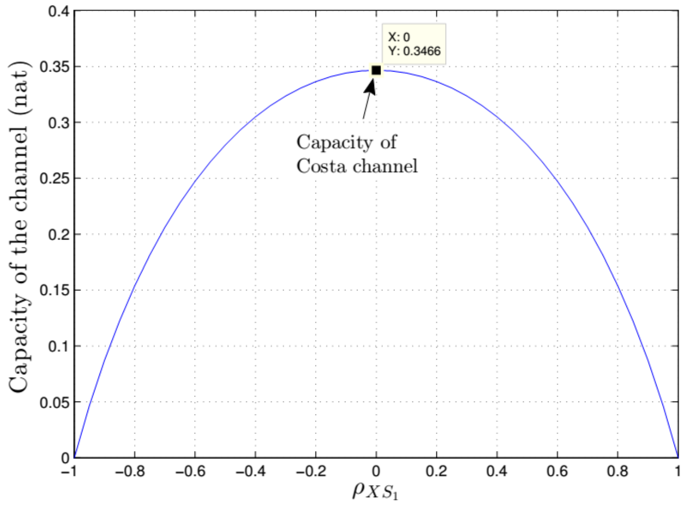

Fig. 6. Capacity of the channel when $P = N = 1$.

V. Numerical Result

In this section we present a few illustrating figures about the results obtained in previous sections.

Fig.6 shows the variation of the capacity with respect to the correlation between X and S_1, ρ_{XS_1}. The figure is plotted for $P = N = 1$ and $-1 \le \rho_{XS_1} \le 1$. It is seen that when X and S_1 are independent of each other, the capacity reaches to its maximum value. This point is the capacity of the Costa channel $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right) = 0.3466$ and for all other cases the capacity is less than it. When X and S_1 are fully dependent, that is $\rho_{XS_1} = \pm 1$, the capacity of channel is zero.

Fig.7 shows the variation of optimum α in linear function (31), with respect to ρ_{XS_1} .The figure is plotted for $P = N = Q_1 = 1$. It is seen, again, that in the case of the independency of X and S_1, we have $\alpha^* = \frac{P}{P+N} = 0.5$ that is the same value of the Costa channel.

VI. Conclusion

By fully detailed investigation of the Gaussian channel in the presence of two-sided state information with dependency on the input and establishing a capacity theorem for the channel, we illustrated the impact of the correlation between the channel input and state information on the capacity, quantitatively. We proved that for our channel, while the state information known at the receiver has no effect on the capacity, the correlation between the state information known at the transmitter and the channel input, reduces the capacity of the channel.

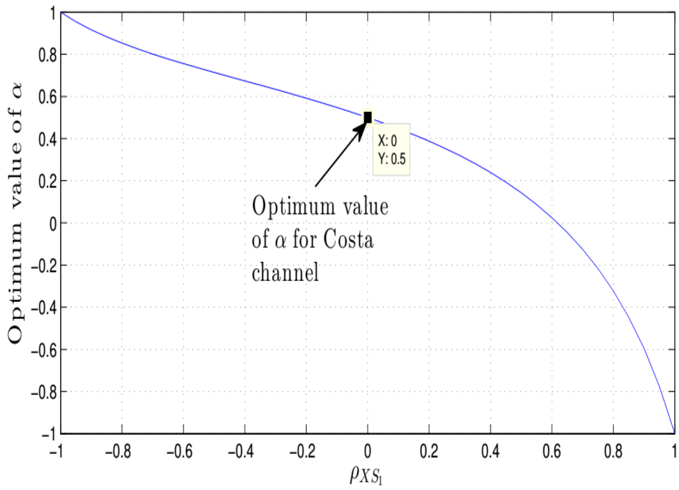

Fig. 7. Optimum value of α when $P = N = Q_1 = 1$.

APPENDIX

Lemma 1: Two continuous random variables X and S with arbitrary probability density functions $f_X(x)$ and $f_S(s)$ can be correlated to each other with specific correlation coefficient ρ_{XS}.

Proof: It is sufficient to show that for arbitrary $f_X(x)$ and $f_S(s)$, there exists a joint density function $f_{X,S}(x,s)$ such that $E\{XS\} = \sigma_X \sigma_S \rho_{XS} + E\{X\}E\{S\}$. Suppose that $F_X(x)$ and $F_S(s)$ are the distribution functions of $f_X(x)$ and $f_S(s)$ respectively. Consider the following function:

$$f_{X,S}(x,s) = f_X(x) f_S(s)\left[1 + \rho\left(2F_X(x) - 1\right)\left(2F_S(s) - 1\right)\right] \tag{66}$$

It can be easily shown that (66) is a joint density function with marginal densities $f_X(x)$ and $f_S(s)$ [24, p.176]. Moreover defining

$$a_X = \int_{-\infty}^{+\infty} x f_X(x)\left(2F_X(x) - 1\right)dx \tag{67}$$

and

$$a_S = \int_{-\infty}^{+\infty} s f_S(s)\left(2F_S(s) - 1\right)ds \tag{68}$$

we consider that:

$$\begin{aligned} E\{XS\} &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xs f_{X,S}(x,s)\,dxds \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xs f_X(x) f_S(s)\left[1 + \rho\left(2F_X(x) - 1\right)\left(2F_S(s) - 1\right)\right]dxds \\ &= E\{X\}E\{S\} + \rho a_X a_S \end{aligned}$$

or

$$\rho = \frac{\sigma_X \sigma_S}{a_X a_S} \rho_{XS}. \tag{69}$$

So the joint density function (66) with ρ in (69) is the answer of the problem if we can prove that a_X and a_S in (67) and (68) exist and are positive. To prove that, first we notice that

$$\int_{-\infty}^{+\infty} F_X(x)\big(1 - F_X(x)\big)dx = a_X + [xF_X(x)(1 - F_X(x))]_{-\infty}^{+\infty}. \tag{70}$$

The last expression is equal to zero because $F_X(\pm\infty)\big(1 - F_X(\pm\infty)\big)$ *is exactly equal* to zero by definition. And considering that $F_X(x)\big(1 - F_X(x)\big)$ is a positive and continuous function for all values of x, we can conclude that the integral in (70) exist and has positive value. So a_X (and similarly a_S) exists and is non-zero.

Lemma 2: Suppose that three zero mean random variables (X, S_1, S_2) have the covariance matrix $\mathbf{K}$:

$$\mathbf{K} = E\left\{ \begin{bmatrix} X^2 & XS_1 & XS_2 \\ XS_1 & S_1^2 & S_1S_2 \\ XS_2 & S_1S_2 & S_2^2 \end{bmatrix} \right\} \tag{71}$$

$$= \begin{bmatrix} \sigma_X^2 & \sigma_X \sigma_{S_1} \rho_{XS_1} & \sigma_X \sigma_{S_2} \rho_{XS_2} \\ \sigma_X \sigma_{S_1} \rho_{XS_1} & \sigma_{S_1}^2 & \sigma_{S_1} \sigma_{S_2} \rho_{S_1 S_2} \\ \sigma_X \sigma_{S_2} \rho_{XS_2} & \sigma_{S_1} \sigma_{S_2} \rho_{S_1 S_2} & \sigma_{S_2}^2 \end{bmatrix} \tag{72}$$

and moreover suppose (S_1, S_2) are *jointly Gaussian* random variables. Then if (X, S_1, S_2) form Markov chain $S_2 \to S_1 \to X$, *(even if X is not Gaussian)* we have:

$$\rho_{XS_2} = \rho_{XS_1} \rho_{S_1 S_2} \tag{73}$$

or equivalently:

$$E\{S_1^2\}E\{XS_2\} = E\{XS_1\}E\{S_1S_2\} \tag{74}$$

Proof

$$\rho_{XS_2} = \frac{E\{XS_2\}}{\sigma_X \sigma_{S_2}} = \frac{E\{E\{XS_2 \mid S_1\}\}}{\sigma_X \sigma_{S_2}} \tag{75}$$

$$= \frac{E\{E\{X \mid S_1\}E\{S_2 \mid S_1\}\}}{\sigma_X \sigma_{S_2}} \tag{76}$$

$$= \frac{\rho_{S_1S_2}}{\sigma_X \sigma_{S_1}} E\{S_1 E\{X \mid S_1\}\} \tag{77}$$

$$= \frac{\rho_{S_1S_2}}{\sigma_X \sigma_{S_1}} E\{XS_1\} \tag{78}$$

$$= \rho_{XS_1} \rho_{S_1S_2} \tag{79}$$

where (77) follows from Gaussianness of (S_1, S_2) and the fact that $E\{S_2 \mid S_1\} = \frac{\sigma_{S_2}\rho_{S_1S_2}}{\sigma_{S_1}} S_1$ and (78) follows from this general rule that: for random variables A and B we have $E\{g_1(A)g_2(B)\} = E\{g_1(A)E\{g_2(B) \mid A\}\}$ [24, p.234].

REFERENCES

[1] C. E. Shannon., Channels with side information at the transmitter, *IBM Journal of Research and Development*, 2(4), pp. 289 –293, Oct. 1958.

[2] A. V. Kosnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, 10(2), pp. 52–60, Apr./Jun. 1974.Translated from Russian.

[3] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, 9(1), pp.19–31, 1980.

[4] C. Heegard and A. El Gamal, "On the capacity of computer memory with defects, " *Information Theory, IEEE Transactions on*, 29(5), pp. 731 – 739, Sept. 1983.

[5] T. M. Cover and Mung Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *Information Theory, IEEE Transactions on*, 48(6), pp.1629 –1638, June 2002.

[6] M. Costa, "Writing on dirty paper (corresp.)," *Information Theory, IEEE Transactions on*, 29(3), pp.439 – 441, May 1983.

[7] S. Jafar, "Capacity with causal and noncausal side information: A unified view," *Information Theory, IEEE Transactions on*, 52(12), pp. 5468 –5474, Dec. 2006.

[8] Guy Keshet, Yossef Steinberg, and Neri Merhav, "Channel coding in the presence of side information," *Found. Trends Commun. Inf. Theory*, 4, pp. 445–586, June 2008.

[9] N. Merhav and S. Shamai, "Information rates subject to state masking," *Information Theory, IEEE Transactions on*, 53(6), pp. 2254 –2261, June 2007.

[10] Y. Steinberg and S. Shamai, "Achievable rates for the broadcast channel with states known at the transmitter," *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pp. 2184 –2188, Sept. 2005.

[11] S. Sigurjonsson and Young-Han Kim, "On multiple user channels with state information at the transmitters," *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pp. 72 –76, Sept. 2005.

[12] Y. H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, p. 534, July 2004.

[13] Reza Khosravi-Farsani and Farokh Marvasti, "Multiple access channels with cooperative encoders and channel state information," *Submitted to European Transactions on Telecommunications*, Sep. 2010, Available at: http://arxiv.org/abs/1009.6008.

[14] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," *Information Theory, IEEE Transactions on*, 55(6), pp. 2442 –2454, June 2009.

[15] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information," *Information Theory, IEEE Transactions on*, 51(8), pp. 2867–2877, Aug. 2005.

[16] A. Rosenzweig, Y. Steinberg, and S. Shamai, "On channels with partial channel state information at the transmitter," *Information Theory, IEEE Transactions on*, 51(5), pp.1817–1830, May 2005.

[17] B. Chen, S. C. Draper, and G. Wornell, "Information embedding and related problems: Recent results and applications," *Allerton Conference*, USA, 2001.

[18] A. Zaidi and P. Duhamel, "On channel sensitivity to partially known two-sided state information," *Communications, 2006. ICC '06. IEEE International Conference on*, volume 4, pp. 1520 –1525, June 2006.

[19] A Sutivong, Mung Chiang, T.M. Cover, and Young-Han Kim, "Channel capacity and state estimation for state-dependent gaussian channels," *Information Theory, IEEE Transactions on*, 51(4), pp.1486–1495, April 2005.

[20] N. S. Anzabi-Nezhad, G. A. Hodtani, and M. Molavi Kakhki, "Information theoretic exemplification of the receiver re-cognition and a more general version for the Costa theorem," *IEEE Communication Letters*, 17(1), pp.107–110, 2013.

[21] A Sutivong, T.M. Cover, and Mung Chiang, "Tradeoff between message and state information rates," *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, p. 303, 2001.

[22] A Sutivong, T.M. Cover, Mung Chiang, and Young-Han Kim, "Rate vs. distortion trade-off for channels with state information," In *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, p. 226, 2002.

[23] Young-Han Kim, A Sutivong, and T.M. Cover. State amplification, "*Information Theory, IEEE Transactions on*, 54(5), pp. 1850–1859, May 2008.

[24] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, 4th edition, 2002.