A New Capacity Theorem for the Gaussian Channel with Two-sided Input and Noise Dependent State Information

Nima S. Anzabi-Nezhad*
Department of Electrical Engineering, Quchan University of Technology, Iran
nima.anzabi@qiet.ac.ir
Ghosheh Abed Hodtani
Department of Electrical Engineering, Ferdowsi University of Mashhad, Iran
ghodtani@gmail.com

Received: 10/Oct/2019 Revised: 24/Dec/2019 Accepted: 08/Jan/2020

Abstract

Gaussian interference known at the transmitter can be fully canceled in a Gaussian communication channel employing dirty paper coding, as Costa shows, when interference is independent of the channel noise and when the channel input designed independently of the interference. In this paper, a new and general version of the Gaussian channel in presence of two-sided state information correlated to the channel input and noise is considered. Determining a general achievable rate for the channel and obtaining the capacity in a non-limiting case, we try to analyze and solve the Gaussian version of the Cover-Chiang theorem mathematically and information-theoretically. Our capacity theorem, while including all previous theorems as its special cases, explains situations that can not be analyzed by them; for example, the effect of the correlation between the side information and the channel input on the capacity of the channel that can not be analyzed with Costa's "writing on dirty paper" theorem. Meanwhile, we try to exemplify the concept of "cognition" of the transmitter or the receiver on a variable (here, the channel noise) with the information-theoretic concept of "side information" correlated to that variable and known at the transmitter or at the receiver. According to our theorem, the channel capacity is an increasing function of the mutual information of the side information and the channel noise.

Keywords: Communication channel capacity; Gaussian channel capacity; correlated side information; two sided state information; interference cancellation; dirty paper coding.

1- Introduction

Side information1 channels have been extensively studied since the initiation by Shannon [1] and the subsequent study by Kusnetsov-Tsybakov [2]. The capacity of a channel with side information (CSI) known non-causally only at the transmitter and only at the receiver has been determined by Gel'fand-Pinsker(GP) [3] and Heegard-El Gamal [4] respectively.
Considering the GP theorem for the Gaussian channel, Costa [5] obtained an interesting result, i.e., the channel capacity in presence of Gaussian interference known non-causally at the transmitter is the same as the case without interference. Having extended the results of Gelfand-Pinsker, Cover-Chiang [6] established a general capacity theorem for the channel with two-sided state information. There are many other important researches in the literature, e.g.[7]-[10]. The results for the single user channel have been generalized possibly to multi user channels, at least in special cases [11]-[18].

Our Work: In this paper, we analyze the Gaussian channel with two-sided input and noise dependent state information as additive interference known at the transmitter and the receiver. The problem has three important aspects:
Information theoretic point of view: Gel'fand-Pinsker (GP) theorem [3] obtains the capacity of the channel with side information known non-causally at the transmitter. Investigating the GP theorem for channels with continuous alphabets in a special situation, Costa [5] obtained a Gaussian version of the GP theorem. As seen in (Fig. 1), the side information S_1 is considered as an additive interference and known non-causally at the transmitter. In the channel, the noise Z is independent of (X, S_1) and moreover the input X can be designed with any arbitrary correlation with the side information S_1. Costa shows that employing dirty paper coding (DPC) in which X is designed independently of S_1, the interference S_1 can be fully canceled, and so the channel capacity surprisingly is the capacity of the channel without interference.

The results of this paper has been presented, partially, in Iran Workshop on Communication and Information Theory, IWCIT 2015, as invited talk.

Cover-Chiang [6] analyze the channel with two-sided and correlated state information non-causally known at the transmitter and receiver and obtain the capacity theorem as the extended version of GP theorem.

The Gaussian version of the GP and Cover-Chiang theorems are open problems in information theory. In addition to Costa's "writing on dirty paper", there are many other important researches in the literature, e.g [10] and its references that studied the problem in special cases. In [10], the channel with *one-sided* additive interference (known at the transmitter) is analyzed in which the interference and noise have arbitrary joint distribution and the noise is dependent on the channel input and interference. The authors obtained a *lower bound* for the capacity of the channel.

In this paper, we try to analyze the Gaussian channel with two-sided state information as additive Gaussian interference (S_1, S_2) and known non-causally at the transmitter and receiver and dependent on the Gaussian channel noise Z and input X (Fig. 2). The random variables (X, S_1, S_2, Z) are *arbitrarily correlated*, and so the channel can be considered as a more general Gaussian version of the Cover-Chaing channel. We prove a general achievable rate (lower bound) for the channel (lemma 1), then, we obtain an upper bound for the capacity of the channel in the case that the channel input, the side information, and the channel noise, form the Markov chain $X \to (S_1, S_2) \to Z$ (lemma 2). We show the coincidence of the lower and upper bounds under this circumstance, and so establish our capacity theorem for the channel (Theorem 1). The theorem includes Costa's "writing on dirty paper" [5] (when Z is independent of (X, S_1, S_2) and X is independent of S_1 as Costa' channel) and the lower bound proved in [10] (with Gaussian noise and interference, ignoring S_2, and X independent of S_1) as its special cases. Our theorem shows that as in [5], the effect of known interference at the transmitter and the receiver can be fully canceled employing "dirty paper-like coding" scheme.

Practical point of view: Enormous developments of wireless communications make the spectrum into one of the most precious resources in modern communications. Costa shows that employing dirty paper coding (DPC), it is possible to fully cancel known interference at the transmitter without consuming additional power, and therefore DPC is one way to utilize the spectrum efficiently by reusing it. However, Costa's "writing on dirty paper" is not applicable for the situation there exist interference known at the receiver (S_2 in Fig. 2) or random variables (X, S_1, S_2, Z) are correlated or the channel input X can not be designed independently of S_1. One example of these situations is the cognitive interference channel in which the transmitted sequence of one transmitter is a known interference for the other transmitter and these two sequences may be dependent on each other (for example when the sources are dependent). Some other communication scenarios in which the channel input and the side information may be correlated and the related investigations can be found in [9] and [19]. In [9] the problem of optimum transmission rate under the requirement of minimum mutual information $I(S_1^n; Y^n)$ is investigated.

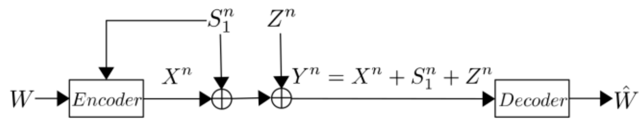

Fig. 1. Gaussian channel with additive interference known non-causally at the transmitter.

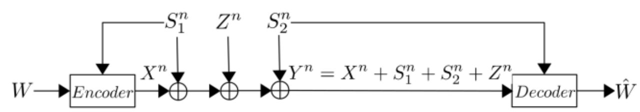

Fig. 2. Gaussian channel with side information at the transmitter and at the receiver.

Cognition on the channel noise: In this paper, we try to describe and analyze the knowledge ("cognition") of the transmitter and receiver about the channel noise Z. We consider the side information known at the transmitter S_1 (at the receiver S_2) and correlated to the channel noise Z, as the cognition of the transmitter (receiver) on the channel noise. This cognition can be perfect or imperfect. It is expected that if the side information S_1 is more correlated with Z (that means greater mutual information $I(S_1; Z)$), the transmitter acquires more knowledge about the channel noise, so employing the proper coding scheme, achieves more data rate.

Regarding the side information S_1 and S_2 as the cognition of the transmitter and the receiver on the channel noise Z, some conditions between random variables are sensible and sound. For example, it is sensible to assume that the knowledge that the transmitter got about the channel noise, gained just via the side information (S_1, S_2) . This condition can be expressed by the equation $I(X; Z|S_1, S_2) = 0$ or Markov chain $X \to (S_1, S_2) \to Z$, which is assumed in obtaining the upper bound of the capacity. Our theorem shows that the channel capacity is an increasing function of the mutual information between the side information and the channel noise $I(S_1, S_2; Z)$.

From this point of view, the subject of the knowledge is the channel noise and the transmitter and receiver acquire this knowledge via side information (S_1, S_2). Therefore, our theorem is indeed an analysis of the effect of uncertainty about the channel noise on the capacity of the channel. In [20] and [21], we analyze the problem in some different and more limited situations.

The problem of partial channel state information studied extensively, e.g. [22]-[25]. In these papers, the subject of knowledge is the state information itself. In [25] the imperfect known state information (as a channel interference) is partitioned to one perfect known and one unknown part. In [24] partially known two-sided state information is viewed as a disturbed state information by Gaussian noise and then the channel sensitivity to small perturbation is analyzed.

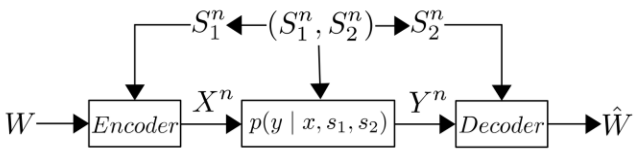

Fig. 3. Channel with side information available non-causally at the transmitter and at the receiver

In section 2, we briefly review the Cover-Chiang and the Gel'fand-Pinsker theorems and then introduce scrutiny of the Costa theorem. In section 3, we define our Gaussian channel thoroughly and present the capacity of the channel. In Section 4, we present some corollaries of the capacity theorem, some numerical comparisons and explain that how the capacity theorem can exemplify the "cognition" of the transmitter and or the receiver on the channel noise. The proofs of lower and upper bounds of the capacity are given in Section 5. Section 6 contains the conclusion. A Lemma which is used in our proofs, is given in the Appendix.

2- A Review of Previous Related Works

To clarify our approach in subsequent sections, in this section we first briefly review the Cover-Chiang capacity theorem for channels with side information available at the transmitter and at the receiver. We then review the Gel'fand-Pinsker (GP) theorem which is a special case of Cover-Chiang theorem when side information is known only at the transmitter. Finally, the Costa theorem ("writing on dirty paper" theorem), which is the Gaussian version of the GP theorem, is investigated.

Cover-Chiang Theorem

Fig.3 shows a channel with side information known at the transmitter and at the receiver where X^n and Y^n are the transmitted and the 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, $p(y|x, s_1, s_2)$, depends on the input X, the side information S_1 and S_2. It can be shown 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 [6]:

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

where the maximum is over all distributions:

$$p(y, x, u, s_1, s_2) = p(y|x, s_1, s_2)p(u, x|s_1)p(s_1, s_2) \quad (2)$$

and U is an auxiliary random variable. It is important to note that the Markov chains:

$$S_2 \longrightarrow S_1 \longrightarrow UX \quad (3)$$

$$U \rightarrow XS_1S_2 \rightarrow Y \quad (4)$$

are satisfied for all distributions in (2).

Gel'fand-Pinsker Theorem

Gel'fand-Pinsker (GP) theorem [3], can be considered as a special case of (1), when there is no side information known at the receiver ($S_2 = \phi$). The capacity of the channel is:

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

for all distributions:

$$p(y, x, u, s_1) = p(y|x, s_1)p(u, x|s_1)p(s_1). \quad (6)$$

Costa's "Writing on Dirty Paper"

Costa [5] examined the Gaussian version of the channel with side information known at the transmitter (Fig. 1). As can be seen, the side information is considered as an additive interference at the receiver. Costa showed that the channel, surprisingly, has the capacity $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$, which is the the same for channels with no interference S_1. Costa derived this capacity by using the results of Gel'fand-Pinsker theorem extended to random variables with continuous alphabets. In this subsection, we first introduce the Costa assumptions and then present a proof for this theorem in such a way that it enables us to introduce our channel and develop our theorem in subsequent sections.

The channel is defined by continuous random variables $(Y, X, U, S_1) \sim f(y, x, u, s_1)$ with following properties:

• S_1^n is a sequence of Gaussian i.i.d. random variables with distribution $S_1 \sim \mathcal{N}(0, Q_1)$.

• The output is given by $Y^n = X^n + S_1^n + Z^n$, where Z^n is the sequence of white Gaussian noise with zero mean and variance N i.e. $Z \sim \mathcal{N}(0, N)$ and independent of (X, S_1). The sequence S_1^n is non-causally known at the transmitter. The transmitted sequence X^n is assumed to have the power constraint $E\{X^2\} \leq P$.

It is readily seen that the distributions $f(y, x, u, s_1)$ having the above three properties are in the form of (6). We denote the set of all these $f(y, x, u, s_1)$'s with $\mathcal{F}_C$. Although for the Costa channel described above, no restriction has been imposed on the correlation between X and S_1, in the Costa theorem, the maximum rate corresponds to independent X and S_1, and U in form of linear combination of X and S_1. We define $\mathcal{F}'_C$ as a subset of $\mathcal{F}_C$ with elements $f'(y, x, u, s_1)$ having the following properties as well as the properties mentioned before:

• X is a zero mean Gaussian random variable with the maximum variance P and *independent* of S_1.

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

It is clear that the set $\mathcal{F}'_C$ and their marginal and conditional distributions are subsets of corresponding $\mathcal{F}_C$'s.

Achievable rate for Costa channel: From (5), when extended to memoryless channels with discrete time and continuous alphabets, we can obtain an achievable rate for the channel. The capacity of Costa channel can be written as:

$$C_{Costa} = \max_{f(u,x|s_1)} [I(U;Y) - I(U;S_1)] \quad (7)$$

where the maximum is over all $f(y,x,u,s_1)$'s in $\mathcal{F}_C$. Since $\mathcal{F}'_C \subseteq \mathcal{F}_C$ we have:

$$C_{Costa} \geq \max_{f'(u,x|s_1)} [I(U;Y) - I(U;S_1)] \quad (8)$$

$$= \max_{f'(u|x,s_1)f'(x|s_1)} [I(U;Y) - I(U;S_1)] \quad (9)$$

$$= \max_{\alpha} [I(U;Y) - I(U;S_1)] \quad (10)$$

The expression in the last bracket is calculated for distributions $f'(y,x,u,s_1)$ in $\mathcal{F}'_C$ described above. Thus, defining $R(\alpha) = I(U;Y) - I(U;S_1)$, $\max_\alpha R(\alpha)$ is an achievable rate for the channel. $R(\alpha)$ and $\max_\alpha R(\alpha)$ is calculated as:

$$R(\alpha) = \frac{1}{2}\log\left(\frac{P(P+Q_1+N)}{PQ_1(1-\alpha)^2 + N(P+\alpha^2 Q_1)}\right), \quad (11)$$

and

$$\max_{\alpha} R(\alpha) = R(\alpha^*) = \frac{1}{2}\log\left(1+\frac{P}{N}\right) \quad (12)$$

where

$$\alpha^* = \frac{P}{P+N} \quad . \quad (13)$$

Both $R(\alpha^*)$ and α^* are independent of Q_1 and then of S_1.

Converse part *of Costa theorem:* From (5) we can also obtain an upper bound for the channel capacity. We have:

$$I(U;Y) - I(U;S_1) = -H(U|Y) + H(U|S_1) \quad (14)$$

$$\leq -H(U|Y,S_1) + H(U|S_1) \quad (15)$$

$$= I(U;Y|S_1) \quad (16)$$

$$\leq I(X;Y|S_1) \quad (17)$$

where inequality (15) follows from the fact that conditioning reduces the entropy and (17) follows from Markov chain $U \to XS_1 \to Y$ which is correct for all distributions $f(y,x,u,s_1)$ in the form of (6), including the distributions in the set $\mathcal{F}_C$. Hence we can write:

$$C_{Costa} = \max_{f(u,x|s_1)} [I(U;Y) - I(U;S_1)] \quad (18)$$

$$\leq \max_{f(x|s_1)} [I(X;Y|S_1)] \quad (19)$$

$$= \max_{f(x|s_1)} [H(Y|S_1) - H(Y|X,S_1)] \quad (20)$$

$$= \max_{f(x|s_1)} [H(X+Z|S_1) - H(Z|X,S_1)] \quad (21)$$

$$\leq \max_{f(x|s_1)} [H(X+Z) - H(Z)] \quad (22)$$

$$= \frac{1}{2}\log\left(1+\frac{P}{N}\right), \quad (23)$$

where the inequality (22) is due to the fact that conditioning reduces the entropy and $H(Z|X,S_1) = H(Z)$ (because the channel noise Z is independent of (X,S_1) in the channel, as defined above). The maximum in (22) is obtained when X and Z are jointly Gaussian with $E\{X^2\} = P$ because when the variance is limited, Gaussian distribution maximizes the entropy. From (12) and (23) it is seen that the lower and the upper bounds of the capacity coincide, and therefore the channel capacity is equal to $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$. It is also concluded that for the Costa channel, the optimum condition which leads to the capacity is when $X \sim \mathcal{N}(0,P)$ and independent of S_1.

We can explain the Costa theorem more, as follows: Let consider $Y = X + S_1 + S_1' + Z$ with independent Gaussian interference S_1 with variance Q_1, S_1' with variance Q_1' and Z with variance N. If the transmitter knows nothing about this interference, then we take $U = X$ and $C = \frac{1}{2}\log\left(1+\frac{P}{N+Q_1+Q_1'}\right)$. If S_1 is known at the transmitter, then we take $U = X + \alpha S_1$ and we have $C = \frac{1}{2}\log\left(1+\frac{P}{N+Q_1'}\right)$ and if S_1 and S_1' are both known at the transmitter, then $U = X + \alpha S_1 + \beta S_1'$ and $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$.

3- Capacity of The Gaussian Channel with Two-sided Noise and Channel Input Dependent Side Information

In this section, first, we introduce a new Gaussian channel with side information known non-causally at the transmitter and side information known non-causally at the receiver both as Gaussian additive interference at the receiver [26]. Then we present a theorem that obtains the capacity of the channel. The theorem can be considered as a Gaussian version of the Cover-Chiang unifying theorem.

3-1- Definition of The channel

Consider the Gaussian channel depicted in Fig. 2. The side information at the transmitter S_1 and at the receiver S_2 is considered as additive interference at the receiver. Our channel has three differences with Costa's one as follows:

i) In our channel, a specified correlation coefficient ρ_{XS_1} between X and S_1, exists. Let the capacity of the channel be C. The channel with no restriction on ρ_{XS_1} (as is in the Costa channel) has the capacity C_1 [21]:

$$C_1 = \max_{\rho_{XS_1}} C \quad (24)$$

ii) To investigate the effect of the side information known at the receiver, we suppose that in our channel there exists Gaussian side information S_2 known non-causally at the receiver which is correlated to both X and S_1.

iii) We allow the channel input X and the side information S_1 and S_2 to be correlated to the channel noise Z.

Remark: Note that 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 distribution and it can be proved that the distribution of X is *free to choose* [21].

1) Definition of the channel: The channel is defined by continuous random variables $(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{U}, \boldsymbol{S_1}, \boldsymbol{S_2}) \sim \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{u}, \boldsymbol{s_1}, \boldsymbol{s_2})$ with following properties:

• (S_1^n, S_2^n) are i.i.d. sequences with zero mean and jointly Gaussian distributions with variance $Q_1 = \sigma_{S_1}^2$ and $Q_2 = \sigma_{S_2}^2$ respectively. The sequences S_1^n and S_2^n are non-causally known at the transmitter and at the receiver respectively.

• 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 variance N.

• Random variables (X, S_1, S_2, Z) have the covariance matrix $\boldsymbol{K}$:

$$\boldsymbol{K} = E\left\{\begin{bmatrix} X^2 & XS_1 & XS_2 & XZ \\ XS_1 & S_1^2 & S_1S_2 & S_1Z \\ XS_2 & S_1S_2 & S_2^2 & S_2Z \\ XZ & S_1Z & S_2Z & Z^2 \end{bmatrix}\right\}$$
$$= \begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_X\sigma_{S_2}\rho_{XS_2} & \sigma_X\sigma_Z\rho_{XZ} \\ \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_{S_1}^2 & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} & \sigma_{S_1}\sigma_Z\rho_{S_1Z} \\ \sigma_X\sigma_{S_2}\rho_{XS_2} & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} & \sigma_{S_2}^2 & \sigma_{S_2}\sigma_Z\rho_{S_2Z} \\ \sigma_X\sigma_Z\rho_{XZ} & \sigma_{S_1}\sigma_Z\rho_{S_1Z} & \sigma_{S_2}\sigma_Z\rho_{S_2Z} & \sigma_Z^2 \end{bmatrix} \quad (25)$$

where ρ_{XS_1} is the correlation coefficient between X and S_1, and so on.

In this channel, the Gaussian noise Z is not necessarily independent of the additive interference S_1 and S_2 and the input X. Moreover X^n is assumed to have the constraint $\sigma_X^2 \le P$. Except σ_X, all other parameters in $\boldsymbol{K}$ have *fixed values* specified for the channel and are considered as *the definition of the channel*.

• We assume that (X, U, S_1, S_2) form the Markov Chain $S_2 \to S_1 \to UX$. As mentioned earlier in (3), this Markov chain is satisfied by all distributions $f(y, x, u, s_1, s_2)$ in the form of (2) in Cover-Chiang capacity theorem. This Markov chain results in the weaker Markov chain $S_2 \to S_1 \to X$ and this implies that (for the proof see Lemma 3 in the Appendix):

$$\rho_{XS_2} = \rho_{XS_1}\rho_{S_1S_2} \quad (26)$$

We assume that the set of all these distributions $f(y, x, u, s_1, s_2)$ denoted with $\mathcal{F}$. It is readily seen that all distributions $f(y, x, u, s_1, s_2)$ in $\mathcal{F}$ are in the form of (2). Therefore we can apply the extended version of Cover-Chiang theorem for random variables with continuous alphabets to this channel.

2) The channel in optimum situation: We will show that the optimum distribution resulting in maximum transmission rate, is obtained when random variables (Y, X, U, S_1, S_2) are distributed under $f^*(y, x, u, s_1, s_2)$ with following *additional* properties:

• The random variables (X, S_1, S_2) are jointly Gaussian distributed and X has zero mean and the maximum variance P, i.e. $X \sim \mathcal{N}(0, P)$.

• As in the Costa theorem [5]:

$$U = \alpha S_1 + X. \quad (27)$$

but here X and S_1 are correlated.

We assume that the set of all these distributions $f^*(y, x, u, s_1, s_2)$ denoted with $\mathcal{F}^*$. It is clear that the set $\mathcal{F}^*$ and their marginals and conditional distributions are subsets of corresponding $\mathcal{F}$'s.

3) Some necessary definitions: Suppose $\boldsymbol{K_{opt}}$ is the covariance matrix for random variables $(\boldsymbol{X}, \boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{Z})$ in optimum situation of the channel having all properties mentioned above; defining:

$$E_{0,i} = E\{XS_i\} = \sigma_X\sigma_{S_i}\rho_{XS_i} \quad, i = 1,2 \quad (28)$$
$$E_{0,3} = E\{XZ\} = \sigma_X\sigma_Z\rho_{XZ} \quad (29)$$
$$E_{1,2} = E\{S_1S_2\} = \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} \quad (30)$$
$$E_{i,3} = E\{S_iZ\} = \sigma_{S_i}\sigma_Z\rho_{S_iZ} \quad, i = 1,2 \quad (31)$$

we can write $\boldsymbol{K_{opt}}$, and its determinant D and its minors D_1 to D_{13} as:

$$\boldsymbol{K_{opt}} = \begin{bmatrix} P & E_{0,1} & E_{0,2} & E_{0,3} \\ E_{0,1} & Q_1 & E_{1,2} & E_{1,3} \\ E_{0,2} & E_{1,2} & Q_2 & E_{2,3} \\ E_{0,3} & E_{1,3} & E_{2,3} & N \end{bmatrix}, \quad D = \det(\boldsymbol{K_{opt}}) \quad (32)$$

$$D_1 \triangleq \begin{vmatrix} Q_1 & E_{1,2} & E_{1,3} \\ E_{1,2} & Q_2 & E_{2,3} \\ E_{1,3} & E_{2,3} & N \end{vmatrix}, \quad D_2 \triangleq \begin{vmatrix} Q_1 & E_{1,2} \\ E_{1,2} & Q_2 \end{vmatrix} \quad (33\text{-}1)$$

$$D_3 \triangleq \begin{vmatrix} E_{0,1} & Q_1 & E_{1,2} \\ E_{0,2} & E_{1,2} & Q_2 \\ E_{0,3} & E_{1,3} & E_{2,3} \end{vmatrix}, \quad D_4 \triangleq \begin{vmatrix} P & E_{0,1} \\ E_{0,1} & Q_1 \end{vmatrix} \quad (33\text{-}2)$$

$$D_5 \triangleq \begin{vmatrix} P & E_{0,2} & E_{0,3} \\ E_{0,2} & Q_2 & E_{2,3} \\ E_{0,3} & E_{2,3} & N \end{vmatrix}, \quad D_6 \triangleq \begin{vmatrix} Q_2 & E_{2,3} \\ E_{2,3} & N \end{vmatrix} \quad (33\text{-}3)$$

$$D_7 \triangleq \begin{vmatrix} P & E_{0,1} & E_{0,2} \\ E_{0,2} & E_{1,2} & Q_2 \\ E_{0,3} & E_{1,3} & E_{2,3} \end{vmatrix}, \quad D_8 \triangleq \begin{vmatrix} P & E_{0,2} \\ E_{0,2} & Q_2 \end{vmatrix} \quad (33\text{-}4)$$

$$D_9 \triangleq \begin{vmatrix} P & E_{0,1} & E_{0,2} \\ E_{0,1} & Q_1 & E_{1,2} \\ E_{0,2} & E_{1,2} & Q_2 \end{vmatrix}, \quad D_{10} \triangleq \begin{vmatrix} E_{0,1} & E_{0,2} \\ E_{1,2} & Q_2 \end{vmatrix} \quad (33\text{-}5)$$

$$D_{11} \triangleq \begin{vmatrix} E_{0,1} & E_{1,2} & E_{1,3} \\ E_{0,2} & Q_2 & E_{2,3} \\ E_{0,3} & E_{2,3} & N \end{vmatrix}, \quad D_{12} \triangleq \begin{vmatrix} E_{1,2} & Q_2 \\ E_{1,3} & E_{2,3} \end{vmatrix} \quad (33\text{-}6)$$

$$D_1^{\mathcal{N}} \triangleq \begin{vmatrix} 1 & \rho_{S_1S_2} & \rho_{S_1Z} \\ \rho_{S_1S_2} & 1 & \rho_{S_2Z} \\ \rho_{S_1Z} & \rho_{S_2Z} & 1 \end{vmatrix}, \quad D_{13} \triangleq \begin{vmatrix} E_{0,2} & Q_2 \\ E_{0,3} & E_{2,3} \end{vmatrix} \quad (33\text{-}7)$$

$D_1^{\mathcal{N}}$, defined in (33), is the determinant of $\mathrm{cov}(S_1, S_2, Z)$ when the variance of random variables are normalized to 1.

3-2- The Capacity Theorem for The channel

Theorem 1: The Gaussian channel defined in 3.1.1 (Fig. 2) when the channel input X, the side information (S_1, S_2) and the channel noise Z, form the Markov chain $X \to (S_1, S_2) \to Z$, has the capacity:

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\frac{\left(1-\rho_{XS_1}^2\right)\left(1-\rho_{S_1S_2}^2\right)}{D_1^{\mathcal{N}}}\right) \quad (34)$$

$$= \frac{1}{2}\log\left(1 + \frac{P}{N}\left(1 - \rho_{XS_1}^2\right)\exp\left(2I(S_1, S_2; Z)\right)\right),$$

and the capacity (34) is achieved by employing the channel input X and the auxiliary random variable U as in optimum situation 3.1.2 and with

$$\alpha^* = \frac{(D_3+D_9)-(D_7+D_{11})}{D_1+2D_3+D_9}, \quad (35)$$

the terms defined in (33).

Proof: The proof is given in Section 5.

Remark: In the Gaussian channel defined in this section, the side information (S_1, S_2) can be dependent on the channel noise. The capacity is proved with constraint of $X \to (S_1, S_2) \to Z$. As we explain in the next section, the Markov chain states that the knowledge the transmitter has got on the channel noise Z, is acquired via the side information (S_1, S_2).

4- Interpretations of The Capacity Theorem

In this section, we examine the effect of the channel parameters on the channel capacity and explain them.

4-1-Cancellation of Interference

It is seen that with employing a DP like coding scheme, interference S_1 and S_2 can be fully canceled, as in Costa's writing on dirty paper.

4-2- The Effect of ρ_{XS_1}

Corollary 1: If there is no specific correlation between X and S_1, as Costa's dirty paper, the capacity is achieved when the channel input X is designed independently of the known S_1:

$$C = \max_{\rho_{XS_1}}\left(\frac{1}{2}\log\left(1 + \frac{P}{N}\left(1 - \rho_{XS_1}^2\right)\exp\left(2I(S_1, S_2; Z)\right)\right)\right)$$

$$= \frac{1}{2}\log\left(1 + \frac{P}{N}\exp\left(2I(S_1, S_2; Z)\right)\right). \quad (36)$$

Corollary 2: If we assume that the channel noise Z is independent of (X, S_1, S_2), from (34), the capacity of the channel is:

$$C = \frac{1}{2}\log\left(1 + \frac{P}{N}\left(1 - \rho_{XS_1}^2\right)\right) \quad (37)$$

From (24), C is reduced to the Costa capacity $\frac{1}{2}\log\left(1 + \frac{P}{N}\right)$ by maximizing it with $\rho_{XS_1} = 0$.

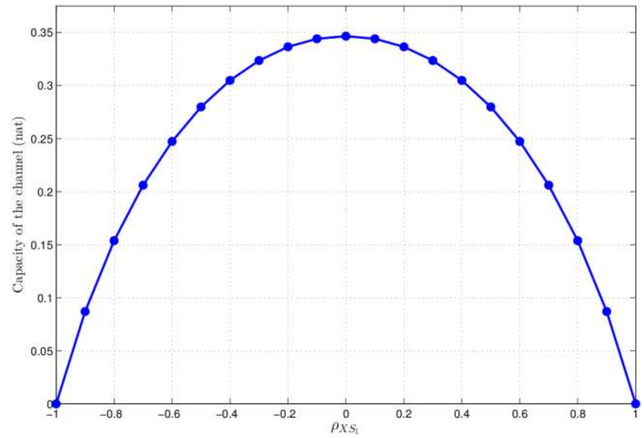

Fig. 4. Capacity of the channel with respect to ρ_{XS_1} when the channel noise Z is independent of (X, S_1, S_2) and with signal to noise ratio $\frac{P}{N} = 1$

Corollary 3: It is seen that in the case the side information S_2 is independent of the channel noise Z, the capacity of the channel is equal to the capacity when there is no interference S_2. In other words, in this case, the receiver can subtract the known S_2^n from the received Y^n without losing any worthy information.

Corollary 4: The correlation between X and S_1 decreases the capacity of the channel. It can be explained as follows: by looking at $Y = X + S_1 + Z$ in our dirty paper like coding, mitigating the input-dependent interference effect, also mitigates the input power impact on the channel capacity as this fact is seen in (37) as $\sigma_X^2\left(1 - \rho_{XS_1}^2\right)$.

As an extreme and interesting case, when $S_1 = X$ (then $\rho_{XS_1} = 1$), according to the usual Gaussian coding, the capacity seems to be $\frac{1}{2}\log\left(1 + \frac{4P}{N}\right)$, which is the capacity when $2X$ is transmitted and $Y = 2X + Z$ is received. But as our theorem shows, the capacity paradoxically is zero. It can be explained as follows: the receiver, based on his information, ought to decode according to the dirty paper like coding. In DP like coding, with given known sequence $S_{1,0}^n$, we find an auxiliary sequence U^n like U_0^n jointly typical with $S_{1,0}^n$ [5]. Jointly typicality of $(U_0^n, S_{1,0}^n)$ is equivalent to:

$$\left|\left(U_0^n - \alpha^* S_{1,0}^n\right)^T S_{1,0}^n\right| \le \delta \quad , \quad \delta \text{ small} \quad (38)$$

where $.^T$ denotes the transpose operation and α^* is computed according to (35). If $X = S_1$, there exists no such U_0^n: since $X_0^n = U_0^n - \alpha^* S_{1,0}^n = S_{1,0}^n$, we have

$$\left|\left(U_0^n - \alpha^* S_{1,0}^n\right)^T S_{1,0}^n\right| = ||S_{1,0}^n||^2 \quad (39)$$

where $||S_{1,0}^n||$ is the norm of the given known sequence $S_{1,0}^n$ and therefore (38) can not be true. In other words, in this case, encoding error occurs.

Fig. 4 shows the variation of the capacity C in (37) with respect to ρ_{XS_1} when $\frac{P}{N}=1$. It is seen that when the correlation between the channel input and the side information known at the transmitter increases, the channel capacity decreases. The maximum capacity is gained when $\rho_{XS_1}=0$, which is Costa's capacity. Fig. 5 shows the capacity C in (37) with respect to SNR for five values of ρ_{XS_1}.

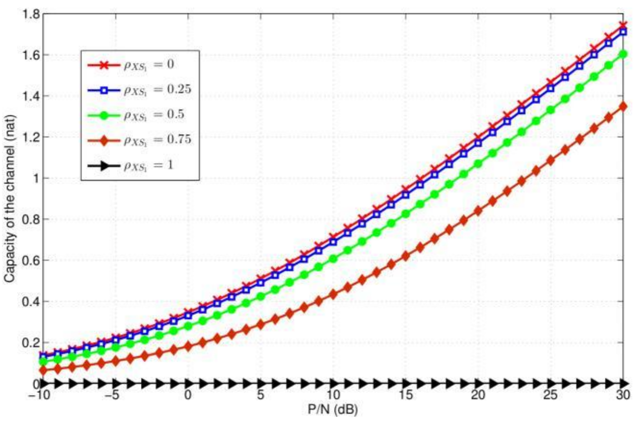

Fig. 5. Capacity of the channel with respect to SNR when the channel noise Z is independent of (X, S_1, S_2)

4-3- Cognition of Transmitter and Receiver on the Channel Noise

It is seen that the mutual information $I(S_1, S_2; Z)$ increases the channel capacity . For the sake of simplicity, we ignore the effect of ρ_{XS_1} and assume the channel capacity is given by (36).

If we suppose that $S_2=\phi$, the capacity is given by:

$$C=\frac{1}{2}\log\left(1+\left(\frac{P}{N}\right)\exp\left(2I(S_1;Z)\right)\right) \quad (40)$$

$$=\frac{1}{2}\log\left(1+\frac{P}{N}\frac{1}{\left(1-\rho_{S_1Z}^2\right)}\right). \quad (41)$$

It is seen that more correlation between the side information S_1 and the channel noise Z, results in more capacity. It can be explained by the fact that the side information known at the transmitter and correlated with the channel noise Z, carries knowledge about Z for transmitter, so enhances the transmitter's ability to cancel the channel noise. When $\rho_{S_1Z}=\pm 1$, the transmitter has got perfect knowledge about Z and the capacity reaches to infinite. Fig. 6 illustrates the capacity of the channel with respect to ρ_{S_1Z} when $\frac{P}{N}=1$. Fig. 7 shows the capacity of the channel with respect to SNR for five values of ρ_{S_1Z}.

The same situation comes about when $S_1=\phi$: the side information S_2 *known* at the receiver and correlated with the Z, carries knowledge about Z for the receiver. More correlation between S_2 and Z results in more capacity. This shows the significance of the known additive interference at the receiver and the reason why subtracting S_2^n from received Y^n is a wrong decoding strategy.

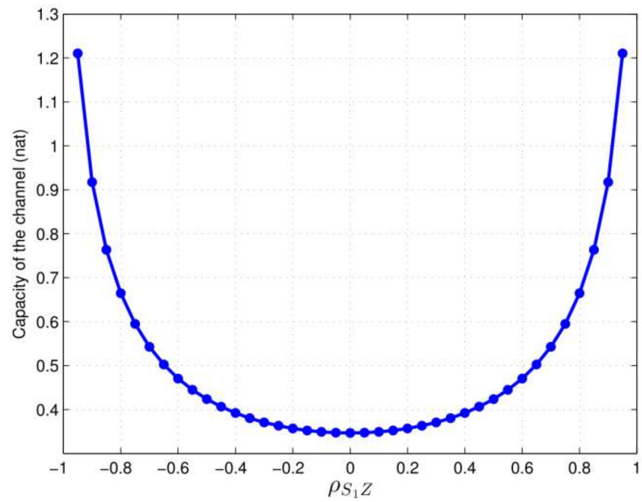

Fig. 6. Capacity of the channel with respect to ρ_{S_1Z} when $\frac{P}{N}=1$

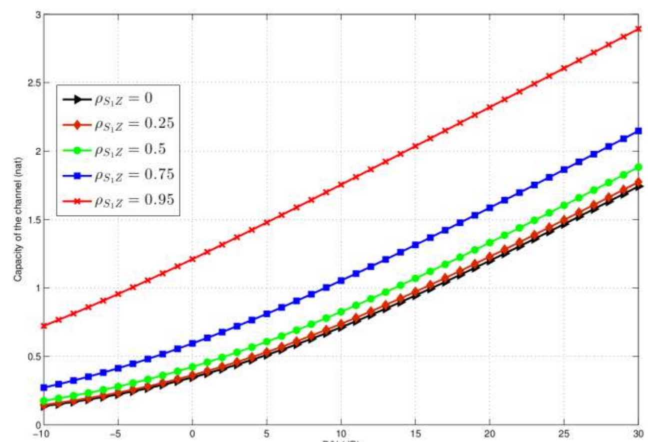

Fig. 7. Capacity of the channel with respect to SNR for five values of ρ_{S_1Z}

If the side information at the transmitter and at the receiver (S_1, S_2) exists and is correlated to the channel noise Z, the capacity increases by $I(S_1, S_2; Z)$. Fig. 8 illustrates the capacity of the channel with respects to mutual information $I(S_1, S_2; Z)$ for five values of SNR.

If the signal to noise ratio is large enough, the capacity can be written as:

$$C=\frac{1}{2}\log\left(1+\frac{P}{N}\right)+I(S_1, S_2; Z). \quad (42)$$

that shows the major effect of $I(S_1, S_2; Z)$ on the channel capacity.

Remark: The capacity of the channel and the effect of S_1 and S_2 on the capacity, reveals the cognitive role of the side information which is known at the transmitter or at the receiver and is correlated to the channel noise. As it is seen in this section, side information known at the transmitter (or receiver), carries the knowledge about the channel noise if it is correlated with the channel noise. If we regard the side information S_1 and S_2 as the cognition of the transmitter and the receiver on the channel noise Z, some conditions between random variables in our model are sensible and sound. For example, it is sensible to assume that the knowledge that the transmitter got about the channel noise, gained just via the side information (S_1, S_2). This conditions can be expressed by equation $I(X; Z|S_1, S_2) = 0$ or Markov chain $X \to (S_1, S_2) \to Z$, which is assumed in our theorem.

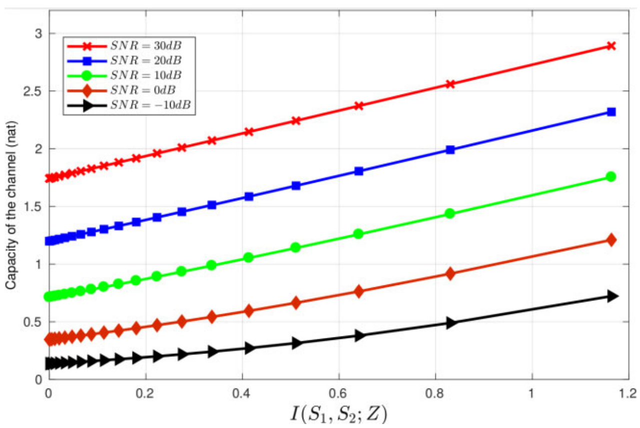

Fig. 8. Capacity of the channel with respect to $I(S_1, S_2; Z)$ for five values of SNR.

5- Proof of Theorem 1

To *prove* the theorem, first, we prove a general achievable rate for the channel. Then we obtain an upper bound for the capacity of the channel when we have the Markov chain $X \to (S_1, S_2) \to Z$. Then we show the coincidence of this upper bound with the lower bound of the capacity.

Lemma 1. Lower Bound of the Capacity: The channel defined in 3.1.1 has the lower bound R_G in (43), where $D_1^{\mathcal{N}}$ is defined in (33):

$$R_G = \frac{1}{2}\log\left(1 + \frac{\left[\sigma_X\left(1-\rho_{XS_1}^2\right) - \sigma_Z\left(\rho_{XS_1}\rho_{S_1Z} - \rho_{XZ}\right)\right]^2\left(1-\rho_{S_1S_2}^2\right)}{\sigma_Z^2\left(\left(1-\rho_{XS_1}^2\right)D_1^{\mathcal{N}} - \left(\rho_{XS_1}\rho_{S_1Z} - \rho_{XZ}\right)^2\left(1-\rho_{S_1S_2}^2\right)\right)}\right) \quad (43)$$

Corollary 5: The lower bound (43) includes the lower bound obtained in [10]. If $\rho_{XS_1} = \rho_{S_1S_2} = 0$ as in [10], then $D_1^{\mathcal{N}} = 1 - \rho_{XZ}^2$ and we have

$$R_G = \frac{1}{2}\log\left(1 + \frac{\left(1+\rho_{XZ}\frac{\sigma_Z}{\sigma_X}\right)^2}{1-\rho_{XZ}^2-\rho_{S_1Z}^2}\frac{\sigma_X^2}{\sigma_Z^2}\right), \quad (44)$$

which is the lower bound in [10].

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

$$C = \max_{f(u,x|s_1)} [I(U; Y, S_2) - I(U; S_1)] \quad (45)$$

where the maximum is over all distributions $f(y, x, u, s_1, s_2)$ in $\mathcal{F}$ defined in 3.1.1. Since $\mathcal{F}^* \subseteq \mathcal{F}$ we have:

$$C \geq \max_{f^*(u,x|s_1)} [I(U; Y, S_2) - I(U; S_1)] \quad (46)$$

$$= \max_{\alpha} [I(U; Y, S_2) - I(U; S_1)] \quad (47)$$

where the expression $I(U; Y, S_2) - I(U; S_1)$ in (47) is calculated for the distributions in $\mathcal{F}^*$ defined in 3.1.2. Thus, defining $R(\alpha) = I(U; Y, S_2) - I(U; S_1)$, we have:

$$C \geq \max_{\alpha} R(\alpha) = R(\alpha^*), \quad (48)$$

therefore $R(\alpha^*)$ is a lower bound for the channel capacity. To compute $R(\alpha^*)$, we write:

$$I(U; Y, S_2) = H(U) + H(Y, S_2) - H(U, Y, S_2) \quad (49)$$

and

$$I(U; S_1) = H(U) + H(S_1) - H(U, S_1), \quad (50)$$

For $H(Y, S_2)$ we have:

$$H(Y, S_2) = \frac{1}{2}\log\left((2\pi e)^2 \det\left(\mathrm{cov}(Y, S_2)\right)\right) \quad (51)$$

where

$$\mathrm{cov}(Y, S_2) = [e_{ij}]_{2\times 2}$$

and

$$e_{11} = P + Q_1 + Q_2 + N + 2E_{0,1} + 2E_{0,2} + 2E_{1,2} + 2E_{0,3} + 2E_{1,3} + 2E_{2,3}$$
$$e_{12} = e_{21} = E_{0,2} + E_{1,2} + Q_2 + E_{2,3}$$
$$e_{22} = Q_2.$$

After computing we have:

$$\det\left(\mathrm{cov}(Y, S_2)\right) = D_2 + D_6 + D_8 + 2D_{10} - 2D_{12} - 2D_{13}, \quad (52)$$

where the terms are defined in (28)-(33).

For $H(U, Y, S_2)$ we have:

$$H(U, Y, S_2) = \frac{1}{2}\log\left((2\pi e)^3 \det\left(\mathrm{cov}(U, Y, S_2)\right)\right) \quad (53)$$

where

$$\mathrm{cov}(U, Y, S_2) = [e_{ij}]_{3\times 3} \quad (54)$$

and

$$e_{11} = P + \alpha^2 Q_1 + 2\alpha E_{0,1}$$
$$e_{12} = e_{21} = P + (\alpha + 1)E_{0,1} + \alpha Q_1 + \alpha E_{1,2} + \alpha E_{1,3} + E_{0,2} + E_{0,3},$$
$$e_{13} = e_{31} = \alpha E_{1,2} + E_{0,2}$$
$$e_{22} = P + Q_1 + Q_2 + N + 2E_{0,1} + 2E_{0,2} + 2E_{1,2} + 2E_{0,3} + 2E_{1,3} + 2E_{2,3}$$

$$e_{23} = e_{32} = E_{0,2} + E_{1,2} + Q_2 + E_{2,3}$$
$$e_{33} = Q_2.$$

After manipulations we have:

$$\det(\text{cov}(U,Y,S_2)) = \alpha^2 D_1 + 2\alpha(\alpha-1)D_3 + 2(\alpha-1)D_7 + (\alpha-1)^2 D_9 + 2\alpha D_{11} + D_5 \quad (55)$$

For $H(S_1)$ and $H(U,S_1)$ we have:

$$H(S_1) = \frac{1}{2}\log((2\pi e)Q_1). \quad (56)$$

and

$$H(U,S_1) = \frac{1}{2}\log\left((2\pi e)^2 \det(\text{cov}(U,S_1))\right) \quad (57)$$

where

$$\text{cov}(U,S_1) = \begin{bmatrix} \alpha^2 Q_1 + P + 2\alpha E_{0,1} & \alpha Q_1 + E_{0,1} \\ \alpha Q_1 + E_{0,1} & Q_1 \end{bmatrix} \quad (58)$$

and its determinant:

$$\det(\text{cov}(U,S_1)) = D_4. \quad (59)$$

Substituting (51), (53), (56) and (57) in (49) and (50), we obtain $R(\alpha)$ as in (60):

$$R(\alpha) = \frac{1}{2}\log\left(\frac{D_4[D_8 + D_2 + D_6 + 2D_{10} - 2D_{12} - 2D_{13}]}{Q_1[(\alpha-1)^2 D_9 + \alpha^2 D_1 + 2\alpha(\alpha-1)D_3 + 2\alpha D_{11} + 2(\alpha-1)D_7 + D_5]}\right) \quad (60)$$

The optimum value of α corresponding to maximum of $R(\alpha)$ is easily obtained as:

$$\alpha^* = \frac{(D_3+D_9)-(D_7+D_{11})}{D_1+2D_3+D_9}, \quad (61)$$

Substituting α^* from (61) into (60) and using the equations (26), (28)-(31) and (33) we finally conclude that $R(\alpha^*)$ equals R_G in (43). Therefore R_G in (43) is a lower bound for the capacity of the channel defined in 3.1.1 (details of computations are omitted for the brevity).

Q.E.D

Lemma 2. Upper Bound of the Capacity: The capacity of the Gaussian channel defined in 3.1.1, when the channel input X, the side information (S_1, S_2) and the channel noise Z form the Markov chain $X \to (S_1,S_2) \to Z$, has the upper bound C in (34).

Proof of Lemma 2: First, we note that the Markov chain $X \to (S_1,S_2) \to Z$ and the Markov chain $S_2 \to S_1 \to X$ in (3), imply the weaker Markov chain $X \to S_1 \to Z$. And it can be proved that this Markov chain implies (for proof see the Appendix):

$$\rho_{XZ} = \rho_{XS_1}\rho_{S_1Z}. \quad (62)$$

For all distributions $f(y,x,u,s_1,s_2)$ in $\mathcal{F}$ defined in 3.1.1, we have:

$$\begin{aligned} I(U;Y,S_2) - I(U;S_1) &= -H(U|Y,S_2) + H(U|S_1) && (63)\\ &\le -H(U|Y,S_1,S_2) + H(U|S_1) && (64)\\ &= -H(U|Y,S_1,S_2) + H(U|S_1,S_2) && (65)\\ &= I(U;Y|S_1,S_2) && (66)\\ &\le I(X;Y|S_1,S_2) && (67)\end{aligned}$$

where (64) follows from the fact that conditioning reduces entropy, (65) follows from Markov chain $S_2 \to S_1 \to UX$ and (67) from Markov chain $U \to XS_1S_2 \to Y$ which are satisfied for any distribution in the form of (2), including the distributions in the set $\mathcal{F}$. From (1) and (67) we can write:

$$C = \max_{f(u,x|s_1)} [I(U;Y,S_2) - I(U;S_1)] \quad (68)$$
$$\le \max_{f(x|s_1)} [I(X;Y|S_1,S_2)]. \quad (69)$$

From (69) it is seen 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 predictable. To compute (69) we write:

$$\begin{aligned} I(X;Y|S_1,S_2) &= H(Y|S_1,S_2) - H(Y|X,S_1,S_2) && (70)\\ &= H(X+S_1+S_2+Z|S_1,S_2) \\ &\quad -H(X+S_1+S_2+Z|X,S_1,S_2) && (71)\\ &= H(X+Z|S_1,S_2) - H(Z|X,S_1,S_2) && (72)\\ &= H(X+Z|S_1,S_2) - H(Z|S_1,S_2) && (73)\\ &= H((X+Z),S_1,S_2) - H(S_1,S_2,Z), && (74)\end{aligned}$$

where (73) follows from the Markov chain $X \to (S_1,S_2) \to Z$. Hence, the maximum value in (69) occurs when $H((X+Z),S_1,S_2)$ is maximum. Since S_1, S_2 and Z are Gaussian, the maximum in (69) is achieved when (X,S_1,S_2) are jointly Gaussian and X has its maximum variance P; in other words, $I(X;Y|S_1,S_2)$ is computed for distribution $f^*(y,x,s_1,s_2)$ defined in 3.1.2. Let $I^*(X;Y|S_1,S_2)$ be the maximum value in (69). We have:

$$C \le I^*(X;Y|S_1,S_2) \quad (75)$$

To compute $I^*(X;Y|S_1,S_2)$, we first compute $H((X+Z),S_1,S_2)$ for distribution $f^*(y,x,s_1,s_2)$ defined in 3.1.2:

$$H((X+Z),S_1,S_2) = \frac{1}{2}\log\left((2\pi e)^3 \det\left(\text{cov}((X+Z),S_1,S_2)\right)\right) \quad (76)$$

where $\text{cov}((X+Z),S_1,S_2)$ is

$$E\left\{\begin{bmatrix} (X+Z)^2 & (X+Z)S_1 & (X+Z)S_2 \\ (X+Z)S_1 & S_1^2 & S_1S_2 \\ (X+Z)S_2 & S_1S_2 & S_1^2 \end{bmatrix}\right\} = \begin{bmatrix} P+N+2E_{0,3} & E_{0,1}+E_{1,3} & E_{0,2}+E_{2,3} \\ E_{0,1}+E_{1,3} & Q_1 & E_{1,2} \\ E_{0,2}+E_{2,3} & E_{1,2} & Q_2 \end{bmatrix} \quad (77)$$

and the determinant:

$$\det\left(\text{cov}((X+Z),S_1,S_2)\right) = D_1 + 2D_3 + D_9, \quad (78)$$

and the other term in (74):

$$H(S_1,S_2,Z) = \frac{1}{2}\log((2\pi e)^3 D_1) \quad (79)$$

Substituting (78) in (76), and from (79), we have:

$$I^*(X;Y|S_1,S_2) = \frac{1}{2}\log\left(1 + \frac{D_9+2D_3}{D_1}\right). \quad (80)$$

Rewriting (80) in terms of σ_X, σ_{S_1}, σ_{S_2}, σ_Z, ρ_{XS_1}, ρ_{S_1Z}, ρ_{S_2Z} and $\rho_{S_1S_2}$ using (28)-(31) and (33) and taking into account two Markovity results (26) and (62), we finally conclude that (details of manipulations are omitted for the brevity):

$$I^*(X;Y|S_1,S_2)=\frac{1}{2}\log\left(1+\frac{P}{N}\frac{\left(1-\rho_{XS_1}^2\right)\left(1-\rho_{S_1S_2}^2\right)}{D_1^N}\right). \quad (81)$$

Hence, C in (34) is an upper bound for the capacity of the channel when we have the Markov chain $X\to(S_1,S_2)\to Z$.

Q.E.D

For *completing* the proof of the capacity theorem, it is enough to compute the lower bound of the channel (43), when we have the Markov chain $X\to(S_1,S_2)\to Z$. Applying the equation (3) to (43), shows the coincidence of the upper and the lower bounds of the capacity of the channel in this case and considering:

$$I(S_1,S_2;Z)=\frac{1}{2}\log\left(\frac{1-\rho_{S_1S_2}^2}{D_1^N}\right) \quad (82)$$

the proof is completed.

Q.E.D

6- Conclusion

By fully *detailed* investigating the Gaussian channel in presence of two-sided input and noise dependent state information, we obtained a general achievable rate for the channel and established the capacity theorem. This capacity theorem first demonstrates the impact of the transmitter and receiver cognition on the capacity and second shows the effect of the correlation between the channel input and side information available at the transmitter and at the receiver on the channel capacity. Whereas, as expected, the cognition of the transmitter and receiver increases the capacity, the correlation between the channel input and the side information known at the transmitter decreases it.

7- Appendix

Lamma 3: Consider three zero mean random variables (X,S_1,S_2) with covariance matrix $\boldsymbol{K}$ as:

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

$$=\begin{bmatrix}\sigma_X^2 & \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_X\sigma_{S_2}\rho_{XS_2}\\ \sigma_X\sigma_{S_1}\rho_{XS_1} & \sigma_{S_1}^2 & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2}\\ \sigma_X\sigma_{S_2}\rho_{XS_2} & \sigma_{S_1}\sigma_{S_2}\rho_{S_1S_2} & \sigma_{S_2}^2\end{bmatrix} \quad (83)$$

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:

$$\rho_{XS_2}=\rho_{XS_1}\rho_{S_1S_2} \quad (84)$$

or equivalently:

$$E\{S_1^2\}E\{XS_2\}=E\{XS_1\}E\{S_1S_2\} \quad (85)$$

Proof of Lemma 3: we can write:

$$\rho_{XS_2}=\frac{E\{XS_2\}}{\sigma_X\sigma_{S_2}}=\frac{E\{E\{XS_2|S_1\}\}}{\sigma_X\sigma_{S_2}} \quad (86)$$

$$=\frac{E\{E\{X|S_1\}E\{S_2|S_1\}\}}{\sigma_X\sigma_{S_2}} \quad (87)$$

$$=\frac{\rho_{S_1S_2}}{\sigma_X\sigma_{S_1}}E\{S_1E\{X|S_1\}\} \quad (88)$$

$$=\frac{\rho_{S_1S_2}}{\sigma_X\sigma_{S_1}}E\{XS_1\} \quad (89)$$

$$=\rho_{XS_1}\rho_{S_1S_2} \quad (90)$$

where (87) follows from the Markov chain $S_2\to S_1\to X$ and (88) follows from Gaussianness of (S_1,S_2) and the fact that $E\{S_2|S_1\}=\frac{\sigma_{S_2}\rho_{S_1S_2}}{\sigma_{S_1}}S_1$ and (89) follows from the general rule that for random variables A and B we have $E\{g_1(A)g_2(B)\}=E\{g_1(A)E\{g_2(B)|A\}\}$ [27, p.234].

Q.E.D

References

[1] C. E. Shannon, "Channels with side information at the transmitter," IBM Journal of Research and Development, vol. 2, no. 4, pp. 289 –293, oct. 1958.

[2] A. V. Kosnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," Probl. Pered. Inform., vol. 10, no. 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, vol. 9, no. 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, vol. 29, no. 5, pp. 731 – 739, sep 1983.

[5] M. Costa, "Writing on dirty paper (corresp.)," Information Theory, IEEE Transactions on, vol. 29, no. 3, pp. 439 – 441, may 1983.

[6] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," Information Theory, IEEE Transactions on, vol. 48, no. 6, pp. 1629 –1638, jun 2002.

[7] S. Jafar, "Capacity with causal and noncausal side information: A unified view," Information Theory, IEEE Transactions on, vol. 52, no. 12, pp. 5468 –5474, dec. 2006.

[8] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," Found. Trends Commun. Inf. Theory, vol. 4, pp. 445–586, June 2008.

[9] N. Merhav and S. Shamai, "Information rates subject to state masking," Information Theory, IEEE Transactions on, vol. 53, no. 6, pp. 2254 –2261, june 2007.

[10] I. Bergel, D. Yellin, and S. Shamai, "A lower bound on the data rate of dirty paper coding in general noise and interference," IEEE Wireless Communications Letters, vol. 3, no. 4, pp. 417–420, 2014.

[11] Y. Steinberg, "Coding for the degraded broadcast channel with random parameters, with causal and noncausal side

information," Information Theory, IEEE Transactions on, vol. 51, no. 8, pp. 2867 –2877, aug. 2005.

[12] S. Sigurjonsson and Y.-H. Kim, "On multiple user channels with state information at the transmitters," in Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on, sept. 2005, pp. 72 –76.

[13] Y. H. Kim, A. Sutivong, and S. Sigurjonsson, "Multiple user writing on dirty paper," in Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on, june-2 july 2004, p. 534.

[14] T. Philosof and R. Zamir, "On the loss of single-letter characterization: The dirty multiple access channel," Information Theory, IEEE Transactions on, vol. 55, no. 6, pp. 2442 –2454, june 2009.

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

[16] R. Duan, Y. Liang, and S. Shamai, "State-dependent gaussian interference channels: Can state be fully canceled?" IEEE Transactions on Information Theory, vol. 62, no. 4, pp. 1957–1970, 2016.

[17] Y. Sun, Y. Liang, R. Duan, and S. S. Shitz, "State-dependent z-interference channel with correlated states," in 2017 IEEE International Symposium on Information Theory (ISIT), 2017, pp. 644–648.

[18] Y. Sun, R. Duan, Y. Liang, and S. Shamai Shitz, "State-dependent interference channel with correlated states," IEEE Transactions on Information Theory, vol. 65, no. 7, pp. 4518–4531, 2019.

[19] Y.-C. Huang and K. R. Narayanan, "Joint source-channel coding with correlated interference," in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on, 2011, pp. 1136 – 1140.

[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, vol. 17, no. 1, pp. 107–110, 2013.

[21] ——, "A more general version of the costa theorem," Journal of Communication Engineering, vol. 2, no. 4, pp. 1–19, Autumn 2013.

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

[23] A. Rosenzweig, Y. Steinberg, and S. Shamai, "On channels with partial channel state information at the transmitter," Information Theory, IEEE Transactions on, vol. 51, no. 5, pp. 1817 – 1830, may 2005.

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

[25] L. Gueguen and B. Sayrac, "Sensing in cognitive radio channels: A theoretical perspective," Wireless Communications, IEEE Transactions on, vol. 8, no. 3, pp. 1194–1198, march 2009.

[26] G. A. Hodtani, "The effect of transceiver recognition on the gaussian channel capacity," in 2015 Iran Workshop on Communication and Information Theory (IWCIT), May 2015, pp. 1–3.

[27] A. Papoulis and S. U. Pillai, Probability, Random Variables, and Stochastic Processes, 4th ed. 1em plus 0.5em minus 0.4em McGraw-Hill, 2002.

Nima Anzabi-Nezhad received the B.S. degree in electronics engineering from Ferdowsi University of Mashhad, Mashhad, Iran in 1995, and M.S. degree in philosophy of science from Sharif University of Technology, Tehran, Iran in 1999. He received the Ph.D. degree in communication engineering in 2014 from Ferdowsi University of Mashhad, Mashhad, Iran. He is an assistant professor in Department of Electrical Engineering at Quchan University of Technology, Quchan, Iran, since 2014. His research interests include information theory, communication theory and cryptography.

Ghosheh Abed Hodtani received the B.Sc. degree in electronics engineering and the M.Sc. degree in communications engineering, both from Isfahan University of Technology, Isfahan, Iran, in 1985, 1987, respectively. He joined Electrical Engineering Dept., at Ferdowsi University of Mashhad, Mashhad, Iran, in 1987. He decided to pursue his studies in 2005 and received the Ph.D. degree (with excellent grade) from Sharif University of Technology, Tehran, Iran, in 2008;and he is a full professor in electrical engineering since 2016, and has been selected as a distinguished engineering professor by Ferdowsi Academic Foundation in 2016. His research interests are in multi-user information theory, communication theory, wireless communications, information-theoretic learning and signal processing. Prof. Hodtani is the author of a textbook on electrical circuits, the winner of the best paper award at IEEE ICT -2010 and a member of technical program and steering committees of Iran workshop on Communication and Information Theory (IWCIT).