On Coding with Rate-Limited Side Information

Aviv Rosenzweig Yossef Steinberg Shlomo Shamai (Shitz) *†

Abstract

We consider an i.i.d. state-dependent channel with partial (rate limited) channel state information (CSI) at the transmitter (CSIT) and full CSI at the receiver (CSIR). The CSIT comprises a non-causal rate limited part, which is communicated by a third party (a genie) over a way-side channel, and may also include outputs of a deterministic scalar quantizer of the state.

A single letter expression for the capacity of the channel is given. In case the CSIT comprises only the rate limited part we show that the capacity suggested in [5] is too optimistic, and suggest a correction as part of our expression for the information capacity. For the general setting, an optimal coding scheme based upon multiplexing of several codebooks is presented. It is proved that the capacity of the channel is the same whether the quantizer's output is observed causally or non-causally by the encoder, and regardless of the type of input constraint considered. Using a rate distortion approach we bound the alphabet's size of the auxiliary RV of the information capacity. Finally, we turn to the AWGN channel with fading, and show that the determination of the capacity region reduces to finding the optimal genie strategies and the optimal power allocation. As a special case, we calculate the capacity region of an on-off channel and suggest sub-optimal genie strategies, which almost fully achieve the capacity region over a wide range of power and way-side channel rate constraints. In addition, we show that the optimal power allocation has a water-filling interpretation. The suggested model can be applied, for example, to an OFDM communication system.

1 Introduction

The rapid development of wireless communications systems over the last decade creates an increasing demand for a definition of corresponding realistic analytic models and for the exploration of the fundamental limits on reliable information transmission over these systems. A common model in the literature assumes a memoryless channel, whose conditional probability distribution is controlled by a time varying state. Many studies have been devoted over the years to various scenarios related to a state-dependent channel model, where full or partial channel state information (CSI) is available at the transmitter (CSIT) and/or at the receiver (CSIR). CSI at the transmitter can be observed in a causal [1], [2], or in a non-causal manner [3], [4]. In this paper we suggest a model of a channel controlled by an i.i.d. state process, where the CSIT is subject to a rate constraint. Our

*This research was supported by THE ISRAEL SCIENCE FOUNDATION.

†All authors are with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, ISRAEL. [aviv@tx, ysteinbe@ee, sshlomo@ee].technion.ac.il.

work also extends previous results by Heegard and El Gamal [5] and by Caire and Shamai [2].

The main drawback of all mentioned results is that the entropy of the CSIT and CSIR can be unbounded, thus requiring (theoretically) links with unbounded capacity for providing the CSI. However, in practice CSI is communicated over way-side channels, for which only limited resources of the system are allocated. Clearly, the use of these channels is subject to various constraints such as bandwidth, operation time and capacity constraints. Therefore, a more realistic model would assume that the CSIT and CSIR are subject at least to (possibly different) rate constraints R_e and R_d, respectively. It is noteworthy that rate limited CSI is of value even when the original CSI is of bounded entropy, since system resources, and especially way-side channel capacity, are always limited, and expensive. Complete characterization of the set of achievable rates for all pairs (R_e, R_d) is rather complicated, and to the best of our knowledge still unknown. An inner bound to the capacity region was provided by Heegard and El Gamal [5]. When the CSIR is perfect, [5] suggested an expression for the capacity of the channel. However, as we show in section 3, this expression is too optimistic.

In this paper we consider an i.i.d. state-dependent channel with partial CSIT and full CSIR. The CSIT comprises two parts. The first part is a rate limited CSI, which is communicated by a third party (a genie) over a noiseless way-side channel of capacity $R_e[bits/symbol]$. We assume that the genie has full knowledge in advance of the whole state sequence, and therefore this part of CSI is non-causal. The second part is outputs of a deterministic scalar quantizer of the channel state. The quantizer's mapping is assumed to be known to the genie and to the receiver. We consider two different cases, according to whether the quantizer's output is observed causally or non-causally by the transmitter. In the sequel we refer to these two cases as the "causal quantizer case" and the "non-causal quantizer case", respectively. In practice, these two kinds of CSI are obtained at the transmitter through a capacity-limited feedback link. Note that the rate limited part of the CSIT is non-causal by definition, even in the "causal quantizer case". Therefore, the model presented here cannot apply to real time systems. However, it has practical applications in cases where coding is not done across the time domain, e.g. FDM communication system, where coding is done across frequencies.

2 Channel Model and Problem Formulation

Consider the channel model illustrated in Fig. 1, which is characterized as follows:

A discrete memoryless channel K $(\mathcal{S}, p_S, \mathcal{X}, p_{Y|X,S}, \mathcal{Y}, f, g, \mathcal{Z}, P)$ consists of four finite alphabets $\mathcal{S}, \mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, a probability transition distribution $p_{Y|X,S}$, a probability mass function p_S on the alphabet $\mathcal{S}$, a scalar function $g : \mathcal{S} \longmapsto \mathcal{Z}$, assumed to be known to the genie and to the decoder, and an input constraint function $f : \mathcal{X} \longmapsto \mathbb{R}^+$, whose expected value is required to be less than or equal to a given value P. $p_S(s)$ is the probability that the channel is in state s. We consider only an i.i.d. state process.

Let $g^i(s^i) \stackrel{\triangle}{=} (g(s_1), \dots, g(s_i))$. Next we define a code for the non-causal quantizer case. The definition of a code for the causal quantizer case is very similar, except for the differences which are given in parentheses.

Definition 2.1 *An (n, M, M_e, ϵ, P) code for the* ***non-causal*** *quantizer (****causal*** *quantizer) case consists of the following mappings:*

$$J_e : \mathcal{S}^n \to \{1, 2, \dots, M_e\}$$

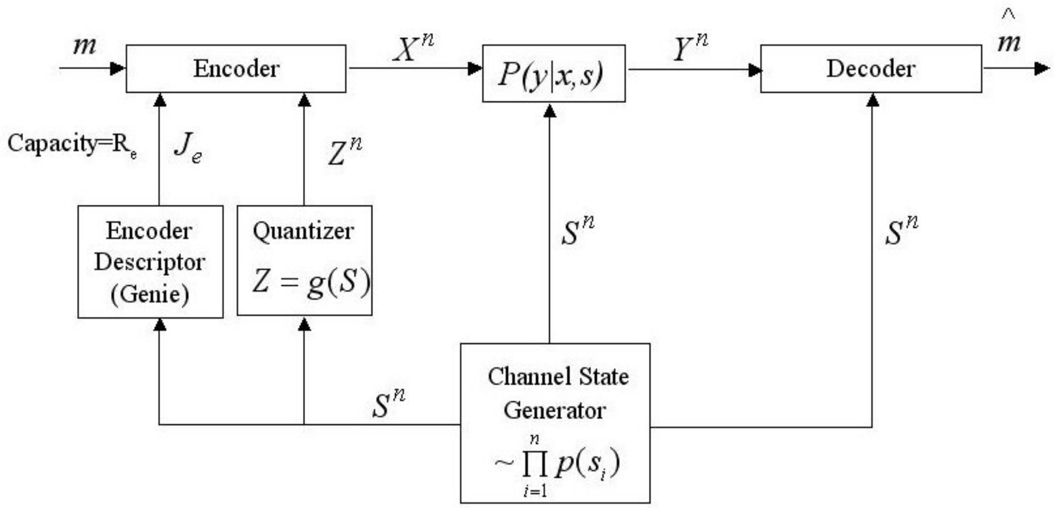

Fig. 1: Channel Model

$$f_e : \{1, 2, \ldots, M\} \times \{1, 2, \ldots, M_e\} \times \mathcal{Z}^n \to \mathcal{X}^n$$
$$\left(f_{e_i} : \{1, 2, \ldots, M\} \times \{1, 2, \ldots, M_e\} \times \mathcal{Z}^i \to \mathcal{X},\ for\ i = 1, \ldots, n\right)$$
$$f_d : \mathcal{Y}^n \times \mathcal{S}^n \to \{1, 2, \ldots, M\}$$

such that

$$P_e \triangleq \frac{1}{M} \sum_{m=1}^{M} \sum_{(s^n,y^n):(s^n,y^n)\notin f_d^{-1}(m)} p_{S^n}(s^n) p_{Y^n|X^nS^n}(y^n|x^n = f_e(m, J_e(s^n), g^n(s^n)), s^n) \leq \epsilon$$

$$\sum_{i=1}^{n} f(x_i(m, J_e, z^n)) \leq nP \ \left(\sum_{i=1}^{n} f(x_i(m, J_e, z^i)) \leq nP\right), 1 \leq m \leq M, 1 \leq J_e \leq M_e, z^n \in \mathcal{Z}^n$$

The rate pair (R, R_e) of the code is defined by: $R \triangleq \frac{1}{n}\log_2 M, R_e \triangleq \frac{1}{n}\log_2 M_e$.

Definition 2.2 1. *A triplet* (R, R_e, P) *is* achievable *for the channel* K *if for every* $\epsilon > 0$ *and sufficiently large* n *there exists an* $(n, 2^{nR}, 2^{nR_e}, \epsilon, P)$ *code.*

2. *The* capacity region *of the channel* K *at input constraint* P *is the closure of all pairs* (R, R_e) *such that* (R, R_e, P) *is achievable.*

3. *The* capacity $C(R_e, P)$ *of the channel* K *at given* (R_e, P) *is the maximal forward rate* R *such that* (R, R_e, P) *is achievable.*

For the causal quantizer case, we can consider 8 different average input constraints, corresponding to all possible combinations of averaging the most restrictive constraint:

$$\sum_{i=1}^{n} f(x_i(m, J_e, z_1^i)) \leq nP,\ 1 \leq m \leq 2^{nR}, 1 \leq J_e \leq 2^{nR_e}, z^n \in \mathcal{Z}^n \quad (1)$$

over m, and/or J_e and/or (z^n or z^i).

Similarly for the non-causal quantizer case we also can consider eight different average input constraints. The loosest input constraint for this case is given by:

$$\sum_{z^n \in \mathcal{Z}^n} \sum_{J_e=1}^{2^{nR_e}} Pr(J_e, z^n) \sum_{m=1}^{2^{nR}} Pr(m) \sum_{i=1}^{n} f(x_i(m, J_e, z^n)) \leq nP \tag{2}$$

Our task is to find the forward channel capacity given R_e and P as constraints, in the causal and non-causal quantizer cases and for all possible average input constraints.

3 Main Results

For any pair (R_e, P), such that $R_e \geq 0, P \geq 0$, define

$$\overline{C}(R_e, P) \triangleq \max_{\substack{p(s_0|s) \\ \text{s.t. } R_e \geq I(S_0;S|Z),}} \quad \max_{\substack{p(x|s_0,z) \\ E[f(x)] \leq P}} \quad I(X;Y|S,S_0,Z) \tag{3}$$

where

$$|\mathcal{S}_0| \leq \max_{z \in \mathcal{Z}} |g^{-1}(z)| + 2$$

$$E[f(x)] \triangleq \sum_{s_0,z} [\sum_{s} p(s_0|s)\delta(z = g(s))p(s)][\sum_{x} p(x|s_0,z)f(x)]$$

Theorem 3.1 *The capacity of channel K at given (R_e, P) in the causal quantizer and in the non-causal quantizer cases is the same, with any input constraint, and is given by:*

$$C(R_e, P) = \overline{C}(R_e, P) \tag{6}$$

Remarks

1. The region of all achievable triplets (R, R_e, P) is convex. In addition, there is no need to take a convex hull on the RHS of (3), i.e. $\overline{C}(R_e, P)$ is convex in both arguments.

2. The convexity property of $\overline{C}(R_e, P)$ implies that the forward channel capacity is achieved when both the genie channel rate constraint and the input constraint are satisfied with equality, in the range $0 \leq R_e \leq H(S|Z)$.

Corollary 3.2 *When Z is degenerate the capacity of channel K is given by:*

$$C(R_e, P) = \max_{\substack{p(s_0|s) \\ s.t.\ R_e \geq I(S_0;S),}} \quad \max_{\substack{p(x|s_0) \\ E[f(x)] \leq P}} \quad I(X;Y|S,S_0) \tag{7}$$

where

$$|\mathcal{S}_0| \leq |\mathcal{S}| + 2$$

In [5] a different expression for the capacity was suggested:

$$C(R_e, P) = \max_{\substack{p(s_0|s) \\ \text{s.t. } R_e \geq I(S_0;S),}} \; \max_{\substack{p(x|s_0) \\ E[f(x)] \leq P}} I(X;Y|S) \tag{9}$$

However, the following arguments show that the last expression is too optimistic, meaning the rate $I(X;Y|S)$ is unachievable in most cases.

First, observe that due to the Markov chain $S_0 \rightarrow (X,S) \rightarrow Y$ the difference between the mutual informations of (7) and (9) is non-negative:

$$I(X;Y|S) - I(X;Y|S,S_0) = H(Y|S) - H(Y|S,S_0) = I(Y;S_0|S) \geq 0 \tag{10}$$

In many cases $I(Y;S_0|S) > 0$, since in general the Markov relation $S_0 \rightarrow S \rightarrow Y$ need not hold.

Secondly, consider the following example. Let $p_{X^*|S}$ be the capacity achieving distribution when both the encoder and the decoder know S. In addition, assume S is not a deterministic function of X^*. Suggest a coding strategy: $X = S_0 = X^*$. Then we have $I(X;Y|S) = I(X^*;Y|S), I(S_0;S) = I(X^*;S)$. Meaning: according to [5], the capacity with two sided fully informed is achievable for a rate constraint $R_e = I(X^*;S) < H(S)$, whereas the encoder is only partially informed of S. However, in many cases rate-limited CSI degrades performance relatively to the fully-informed case.

3.1 Bound on the cardinality of the auxiliary RV

In order to bound the cardinality of the auxiliary RV S_0, note that we can express the information capacity (3) in an alternative way:

$$\overline{R_e}(R,P) = \min_{\substack{p(s_0|s) \\ \text{s.t. } I(X;Y|S,S_0,Z) \geq R,}} \; \min_{\substack{p(x|s_0,z) \\ E[f(x)] \leq P}} I(S_0;S|Z) \tag{11}$$

where R is a lower bound on the transmission rate in the forward channel.

Assume that $p^*_{X|S_0,Z}, p^*_{S_0|S}$ achieve the minimum in (11). We will demonstrate the existence of a probability distribution $p^{'}_{S_0|S}$ such that $p^{'}_{S_0|S}(s_0|s)$ is nonzero for at most $max\ |g^{-1}(z)| + 2$ letters of $\mathcal{S}_0$ and such that $p^*_{X|S_0,Z}, p^{'}_{S_0|S}$ also achieve the minimum in (11). Let us define:

$$d_1(s,s_0,z) \triangleq I(X;Y|s,s_0,z), d_2(s,s_0,z) \triangleq \sum_x p^*(x|s_0,z)f(x) \tag{12}$$

Then, for fixed distribution $p^*_{X|S,Z}$, (11) is identical to finding the rate distortion function of the source S with side information Z at both sides, subject to two distortion constraints:

$$E[d_1(S,S_0,Z)] = R, \; E[d_2(S,S_0,Z)] = P \tag{13}$$

Theorem 3.3 *For every $z \in \mathcal{Z}$ at most $|g^{-1}(z)| + 2$ letters of the alphabet $\mathcal{S}_0(z)$ need to be used to achieve $\overline{R_e}(R,P)$ for any given R, P. Thus,*

$$|\mathcal{S}_0| \leq \max_{z \in \mathcal{Z}} |g^{-1}(z)| + 2 \tag{14}$$

Remark When Z is a degenerate RV the bound is: $|\mathcal{S}_0| \leq |\mathcal{S}| + 2$.

Proof First, assume Z is degenerate. Then, the minimization problem we consider is equivalent to finding the rate distortion function of the memoryless source S, only that here the function is subject to two distortion constraints instead of one. Gallager ([6],corollary 9.4.3) has proved that at most $|\mathcal{S}| + 1$ letters of the reproduction alphabet need to be used to achieve any point on the rate distortion curve of an i.i.d source S. Following this result, it can be shown easily that for the two constraints case a reproduction alphabet of size $|S| + 2$ is sufficient. Thus we have: $|\mathcal{S}_0| \leq |\mathcal{S}| + 2$.

In the general case, we suggest in the direct part of theorem 3.1 an optimal coding scheme using multiplexed multiple code books over $\mathcal{Z}$. In addition, by assumption all parties of the system know the realization of Z at each moment. Therefore, we can assume the use of a different alphabet $\mathcal{S}_0(z)$ for every $z \in \mathcal{Z}$. Note that the mapping $g : \mathcal{S} \longmapsto \mathcal{Z}$ induces a partition on $\mathcal{S}$ in the sense that for fixed $z \in \mathcal{Z}$ the realization of S is taken from the subset $g^{-1}(z)$. Also, for every fixed $z \in \mathcal{Z}$ the encoder observes effectively only the rate limited part of the CSIT, and hence we can apply the bound of the degenerate Z case to get: $|\mathcal{S}_0(z)| \leq |g^{-1}(z)| + 2$.

4 Examples and Applications

4.1 AWGN Channel with Fading

In this section we consider a discrete-time AWGN channel with fading given by

$$Y_n = \sqrt{S_n} X_n + W_n \tag{15}$$

where the channel power gain S_n is an i.i.d. process, and $W_n \sim \mathcal{N}(0, 1)$ is an i.i.d. process, independent of S_n and X_n. An average input constraint $E[X^2] \leq P$ is assumed. The coding theorem of the previous section can be extended by standard arguments of to memoryless channels in discrete-time with continuous alphabets. Using the familiar bound on the entropy of a RV by the entropy of a gaussian RV with the same variance, we can upper bound $I(X; Y|S, S_0, Z)$. It turns out that this bound is achieved with equality if for every pair (s_0, z) we have $X|s_0, z \sim \mathcal{N}(0, \mathcal{E}(s_0, z))$, where $\mathcal{E}(s_0, z) \triangleq E[X^2|s_0, z]$. Hence, we have the following result:

Theorem 4.1 *The AWGN channel with fading, defined by (15), and with CSIT and CSIR as described in section 2, has the following capacity at given* (R_e, P)*:*

$$C(R_e, P) = \underset{\substack{p(s_0|s) \\ s.t.\ R_e \geq I(S_0;S|Z),}}{max} \ \underset{\substack{\mathcal{E}(s_0,z) \\ E[\mathcal{E}(s_0,z)] \leq P}}{max} \ \textstyle\sum_{s,z} p(s,z) \sum_{s_0} p(s_0|s) \frac{1}{2} \log(1 + s\mathcal{E}(s_0, z)) \tag{16}$$

where

$$|\mathcal{S}_0| \leq \underset{z \in \mathcal{Z}}{max} \ |g^{-1}(z)| + 2$$

$$E[\mathcal{E}(s_0, z)] \triangleq \sum_{s_0,z} \mathcal{E}(s_0, z) \left[\sum_s p(s_0|s) p(s, z) \right].$$

We conclude that the determination of the channel capacity reduces to finding the optimal ED strategies $\{p(s_0|s)\}$ and the optimal power allocation function $\mathcal{E}(s_0, z)$.

In the sequel of this section we focus on the contribution of the rate limited CSIT to the increase in the capacity and therefore we assume Z is degenerate.

4.1.1 On Off Channel

Assume S_n is i.i.d. distributed on $\{0, 1\}$. This model describes a communication link disturbed by a jammer.

According to (16) the capacity of this channel is given by

$$C(R_e, P) = \max_{\substack{p(s_0|s) \\ \text{s.t. } R_e \geq I(S_0;S),}} \max_{\substack{\mathcal{E}(s_0) \\ E[\mathcal{E}(s_0)] \leq P}} p_S(1) \sum_{s_0} p_{S_0|S}(s_0|1) \tfrac{1}{2} \log(1 + \mathcal{E}(s_0)) \tag{19}$$

Using Lagrange multipliers and the Kuhn-Tucker conditions we get the optimal power allocation function:

$$\mathcal{E}(s_0) = \left\{ \begin{array}{ll} \frac{p_{S|S_0}(1|s_0)}{\lambda} - 1 & , 0 < \lambda < p_{S|S_0}(1|s_0) \\ 0 & , \lambda \geq p_{S|S_0}(1|s_0) \end{array} \right\} \tag{20}$$

where $0 < \lambda < 1$ is determined by

$$k(p_{S_0|S}, \lambda) \triangleq \sum_{s_0: p_{S|S_0}(1|s_0) > \lambda} \left[\frac{p_{S|S_0}(1|s_0)}{\lambda} - 1 \right] p_{S_0}(s_0) = P \tag{21}$$

This result has an intuitive explanation: Since the channel is "on" (i.e. transmits information) only when $S = 1$, power is allocated only to those symbols s_0, for which the a-posteriori probability (at the output of the channel from S to S_0) of the forward channel being "on" for them, is greater than some threshold λ. Thus, the optimal power allocation (20) has a water-filling nature. Hence, we have the following simplified expression for the capacity of the On-Off Channel:

$$C(R_e, P) = \max_{\substack{p(s_0|s) \\ \text{s.t. } I(S_0;S) \leq R_e,}} \max_{\substack{0<\lambda<1 \\ k(p_{S_0|S}, \lambda) \leq P}} p_S(1) \sum_{s_0: p_{S|S_0}(1|s_0) > \lambda} p_{S_0|S}(s_0|1) \tfrac{1}{2} \log \left[\tfrac{p_{S|S_0}(1|s_0)}{\lambda} \right] \tag{22}$$

In case the capacity achieving distribution $p^*_{S_0|S}$ satisfies the condition $p^*_{S|S_0}(1|s_0) > \lambda$ for every $s_0 \in S_0$, (21) can be solved for λ, and we get $\lambda = \frac{p_S(1)}{1+P}$. Hence, the optimal power allocation function is:

$$\mathcal{E}(s_0) = (1 + P) \frac{p^*_{S|S_0}(1|s_0)}{p_S(1)} - 1, \forall s_0 \in S_0 \tag{23}$$

In this case (22) can also be further simplified and we get

$$C(R_e, P) = \frac{1}{2} p_S(1) \log(1 + P) + \frac{1}{2} p_S(1) D(p^*_{S_0|S}(s_0|1) \| p^*_{S_0}(s_0)) \tag{24}$$

Note that $\frac{1}{2} p_S(1) \log(1 + P)$ is the capacity of the channel without any SI at the transmitter ([2], Proposition 3, and [7]). Hence, The second term on the RHS of (24) expresses the capacity gain due to the rate limited CSIT.

Unfortunately, (22) does not seem to have a closed form solution. However, we can use this formula in order to calculate some achievable rate regions. In the sequel of this section we assume that S is i.i.d. **uniformly** distributed on $\{0,1\}$. We will discuss several sub-optimal strategies which the ED can adopt.

Strategy 1: BSC Channel

Assume a $BSC(p)$ channel from S to S_0, with crossover probability $0 \leq p = h_2^{-1}(1-R_e) \leq 0.5$. Since the a-priori distribution of S achieves the capacity of this channel, we have $I(S;S_0) = 1 - h_2(p) = R_e$. After some calculations we get that this strategy achieves the following rate region:

$$R_1(R_e,P) = \left\{ \begin{array}{cc} \frac{1}{4}[\log(1+P) + R_e] & ,0 \leq R_e \leq 1 - h_2(\frac{1}{2(1+P)}) \\ \frac{1}{4}(1-h_2^{-1}(1-R_e))\log(1+2P) & ,1 - h_2(\frac{1}{2(1+P)}) \leq R_e \leq 1 \end{array} \right\} \quad (25)$$

Strategy 2: Z channel - "On" signaling

Assume a Z channel from S to S_0 such that $p_{S_0|S}(0|0) = 1$, $p_{S_0|S}(0|1) = p$. The parameter p is chosen in order to satisfy: $I(S;S_0) = h_2(\frac{1-p}{2}) - \frac{1}{2}h_2(p) = R_e$. Power is always transmitted for $s_0 = 1$, since receiving $s_0 = 1$ signals the transmitter that the channel is surely "on". The power allocation for $s_0 = 0$ depends on the value of $\lambda = \frac{1}{2(1+P)}$ with respect to p.

It can be shown that this strategy achieves the following rate region:

$$R_2(R_e,P) = \left\{ \begin{array}{cc} \frac{1}{4}[\log(1+P) + 1 + p\log(\frac{p}{1+p})] & ,\frac{1}{1+2P} \leq p \leq 1 \\ \frac{1}{4}(1-p)\log\left(1 + \frac{2P}{1-p}\right) & ,0 \leq p \leq \frac{1}{1+2P} \end{array} \right\} \quad (26)$$

Remark The threshold λ determines if positive power is transmitted for every $s_0 \in S_0$ or only for $s_0 = 1$, through its relation to the error probability p of the BSC (in strategy 1), or to the parameter p (in strategy 2). The intuition behind this, from the encoder's point of view, is that as long as p is large enough, there is a reasonable chance that a received $s_0 = 0$ symbol was flipped over the genie channel, and therefore power is transmitted not only for $s_0 = 1$.

Strategy 3: Z channel - "Off" signaling

Assume a Z channel from S to S_0 such that $p_{S_0|S}(1|1) = 1$,$p_{S_0|S}(1|0) = p$. The parameter p is chosen as in strategy 2. Power is never transmitted for $s_0 = 0$, since receiving $s_0 = 0$ signals the transmitter that the channel is surely "off". Here we have a simpler expression for the achievable rate region:

$$R_3(R_e,P) = \frac{1}{4}\log_2\left(1 + \frac{2P}{1+p}\right), \; 0 \leq R_e \leq 1 \quad (27)$$

Finally we observe that the rate

$$\text{Convex Hull}\left\{ \begin{array}{cc} max & R_i(R_e,P) \\ i = 1,2,3 & \end{array} \right\} \quad (28)$$

is achievable by applying the best strategy of the three (or time sharing two or more strategies) for every given pair (R_e, P).

In figs. 2-3 we show the capacity region (derived numerically) in comparison to the achievable rate regions by the three strategies, and the rate region induced by (28), where $P \in \{0.01, 0.5, 1, 100\}$ is parameter.

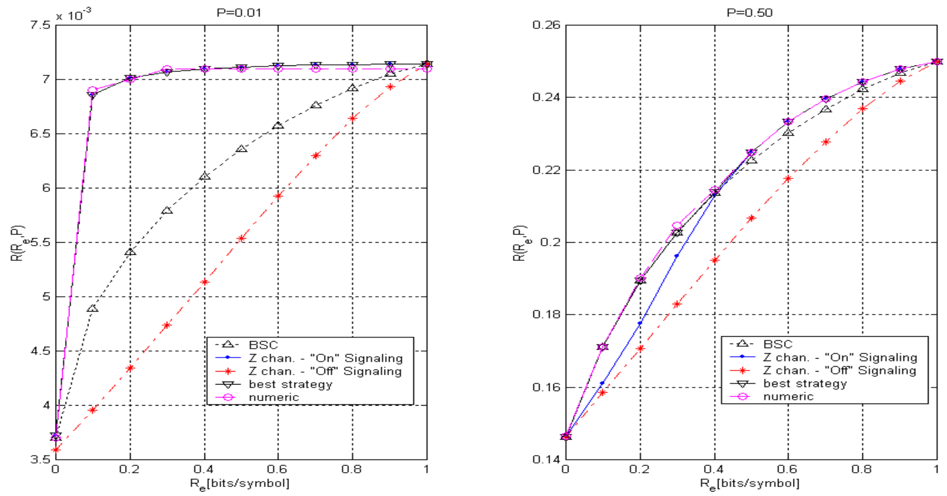

Fig. 2: Capacity region (numeric) vs. achievable rate regions by 3 strategies for $P = 0.01$ (left) and $P = 0.5$ (right)

As can be seen, the rate region (28) almost fully achieves the capacity region over a wide range of input constraints, and therefore it is a good approximated analytic expression for the capacity region of the On-Off Channel. Fig. 2 shows that for $P = 0.01$ the capacity region is almost fully achieved by strategy 2. As P is increased to the value 0.5, we see that strategy 2 characterizes the capacity region only for $R_e > 0.45$, whereas for smaller genie rates strategy 1 is almost optimal. As P is further increased to the values 1 and 100, we see in Fig. 3 that the capacity region is well approximated by strategies 1 and 3, respectively, over the whole range of R_e.

Discussion. As we show in the direct part of theorem 3.1, the optimal genie strategy is very similar to a rate distortion code. In spite of this similarity, the design of good genie strategies is quite different from the design of conventional rate-distortion codes, since we are not interested in describing the state sequence accurately, but only in a description which yields improvement in the forward channel capacity. This key difference is demonstrated in the on-off channel case, with state uniformly distributed, where the optimal rate-distortion conditional distribution of a BSC, turns out not to be optimal for the whole range of input constraints. Whereas in classical rate-distortion theory the achieving distribution of a binary symmetric source with Hamming distortion is symmetric, as is the distortion measure itself, the modified distortion measure $I(X;Y|s,s_0)$, defined for the on-off channel, is not symmetric in s and s_0, due to the asymmetry in the on and off states. Hence, as we have shown, other genie strategies, based on a Z channel, which are more coupled to the "asymmetric nature" of the on-off channel, also take part in the characterization of its capacity region.

4.2 Applications

The model suggested here cannot apply to real time systems, since the rate limited part of the CSIT is non-causal. However, it has practical applications in cases where coding is not done across the time domain, e.g. OFDM communication system, where

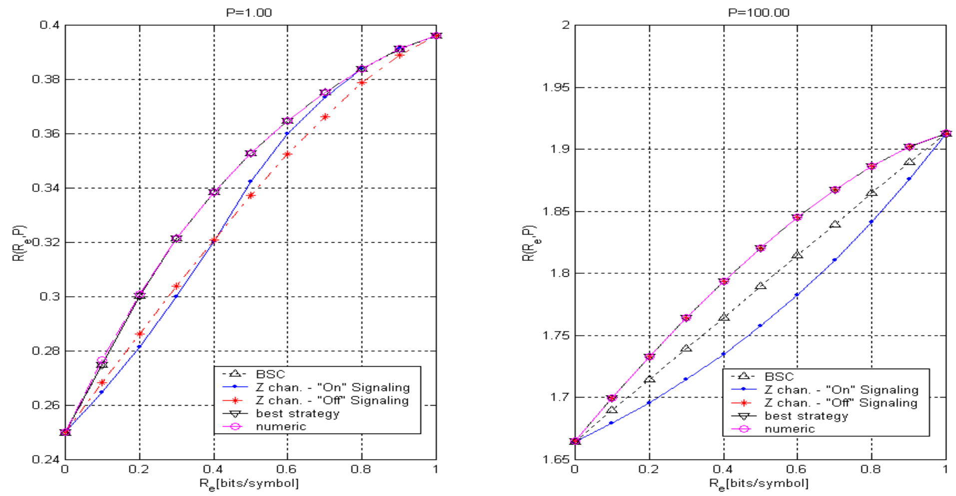

Fig. 3: Capacity region (numeric) vs. achievable rate regions by 3 strategies for $P = 1$ (left) and $P = 100$ (right)

coding is done across frequencies. In this scenario, we can assume the state sequence is, for instance, a vector of SNRs across the different frequencies. The encoder can have, a priori, some knowledge about each of the SNRs - e.g., above/below a threshold, by observing the output of a state quantizer. In this case, the third party helps the encoder by providing it with more CSI through a low rate way-side link.

References

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

[2] G. Caire and S. Shamai, "On the Capacity of Some Channels with Channel State Information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007-2019, Sept. 1999.

[3] S. Gel'fand and M. Pinsker, "Coding for channels with random parameters," *Prob. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19-31, 1980.

[4] T. M. Cover and M. Chiang, "Duality Between Channel Capacity and Rate Distortion With Two-Sided State Infonmation," *IEEE Trans. Inform. Theory*, vol. 48, No. 6, pp. 1629-1638, Jun. 2002.

[5] C. Heegard and A. A. El Gamal, "On the Capacity of Computer Memory with Defects," *IEEE Trans. Inform. Theory*, vol. 29, No. 5, pp. 731-739, Sept. 1983.

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

[7] A. Goldsmith and P. Varaiya, "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Inform. Theory*, vol. 43, No. 6, pp. 1986-1992, Nov. 1997.