URSI AP-RASC 2019, New Delhi, India; 09 - 15 March 2019

The Cost of Probing the State in a State Dependent Additive Gaussian Channel

Viswanathan Ramachandran
Department of Electrical Engineering
Indian Institute of Technology Bombay
viswanathan@ee.iitb.ac.in

Abstract

State dependent models have found wide applications in information theory and communications. The major applications include coding for memory with defective cells, MIMO broadcast channels for future wireless networks, etc. In both these models, the state process is available only at the transmitter and not at the receiver. Acquiring the state process at the transmitter often comes with an associated cost. To model this, message dependent actions on whether to acquire the state process was recently proposed in literature. An open question here is whether the performance can be as good by only acquiring part of the state process. In this paper, we demonstrate that there is always a price to be paid if the complete state process is not acquired, settling a known open question. In the process, we also characterize the exact capacity of such *state probing models*, which was hitherto unknown.

1 Introduction

State dependent channel models have a long history in the information theory domain. They are used to model situations in which the channel transition probability is governed by an external random process, known as the state process. Various interesting scenarios arise depending upon the extent of state knowledge–for instance, the state may be known only at the encoder/decoder, at both or at none of the terminals. The case of channel state availability at the encoder can be further categorized into causal or non-causal (the entire state sequence is known before transmission). The pioneering works on state dependent models are the papers by Shannon [1] (causal case) and Gelfand & Pinsker [2] (non-causal case). The latter model is very much relevant in the context of coding for memory with defective cells, which was studied in [3]. Other practical applications of the model include wireless channels with fading, write-once memory with programmed cells, communication in the presence of jamming,etc. [4]. Costa [5] introduced the *dirty paper coding* (DPC) scheme for a state-dependent Gaussian channel with non-causal state knowledge at the encoder, wherein the result of Gelfand-Pinsker [2] was applied to prove the surprising fact that the capacity is unchanged by the presence of the state. Dirty paper coding has also found wide applications in multi-antenna (MIMO) wireless broadcast channels – see Weingarten *et al.* [6].

In the state dependent models mentioned above, the state is assumed to be generated by nature. Motivated by multi-stage encoding settings like two-stage recording on a magnetic storage device, Weissman [7] introduced *channels with action dependent states.* In this setting, an action encoder takes message dependent *actions* that influence the formation of channel states, and the channel encoding is based upon the state sequence and the message. In this setting, the actions not only control the channel states, but also facilitate message communication. Weissman [7] derived the capacity of this model in the form of a single-letter characterization.

Following Weissman [7], the action-dependent channel framework has been eextended in several works. These include Permuter *et al.* [8], who studied the source coding dual problem in which the decoder can take actions based on the observed compression index. Asnani *et al.* [9] introduced a setting in which the encoder as well as the decoder can take *probing actions* to learn the channel state, with a cost constraint associated with such actions. The main concern of the current paper is an open question posed in [9] regarding the AWGN version of their *Probing Capacity* model, which enquires whether constant probing is required. The answer turns out to be affirmative. In the process, we also settle the capacity characterization for the Gaussian probing model in [9].

Paper Organization: We briefly describe the *Probing Capacity* model of [9] in Section 2. Section 3 establishes the capacity for the same, by providing a converse argument based on revealing the actions to the receiver. Finally Section 5 concludes the paper.

2 Probing Capacity

Consider the setting shown in Figure 1. We consider a Gaussian state-dependent channel given by

$$Y = X + S + Z, \tag{1}$$

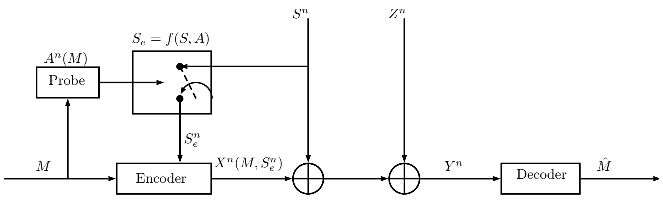

Figure 1. The original *Probing Capacity* model [9]

with $S \sim \mathcal{N}(0,Q)$ and $Z \sim \mathcal{N}(0,N)$ being independent. The input is constrained in average power to P. The encoder does not directly observe the state as in Costa [5], but an *erased* version is seen at the encoder. More specifically, based on the message M, a binary action sequence $A^n(M)$ is generated which controls the state observability at the encoder. If $A=0$, then no state information is available, which we term as an erasure. On the other hand, complete state information is revealed to the encoder when $A=1$. Thus the state information at the transmitter can be described as follows.

$$S_e = \begin{cases} S & \text{if } A=1 \\ * & \text{if } A=0 \end{cases},$$

where the random variable A is binary with $P(A=1)=\varepsilon$. Note that ε can also be interpreted as the cost constraint on the binary action sequence.

Now we define the notions of *achievable rate* and *capacity* in the current setting. The message M is assumed to be drawn uniformly from the set $[1:2^{nR}]$. The state sequence S^n is i.i.d., and independent of the message M. A $(2^{nR},n)$ code for the given setting consists of the following. (Note that the calligraphic letters below denote the alphabets of the corresponding random variables.)

- *Action Map* $A^n(M) : [1:2^{nR}] \to \mathscr{A}^n = \{0,1\}^n$, such that the action sequence obeys the cost constraint $P(A=1) \leq \varepsilon$. The state observed by the enocder is obtained as $S_e = f(A,S)$, where $f(\cdot)$ is a deterministic function.
- *Channel encoder map* $X^n(M,S_e^n) : [1:2^{nR}] \times \mathscr{S}_e^n \to \mathscr{X}^n$.
- *Decoder map* $\hat{M}(Y^n) : \mathscr{Y}^n \to [1:2^{nR}]$.

The probability of error is defined as

$$P_e = P(M \neq \hat{M}(Y^n)). \tag{2}$$

A rate R is said to be *achievable* if there exists a sequence of $(2^{nR},n)$ codes for increasing block lengths n such that $P_e^n \xrightarrow{n\to\infty} 0$. The *capacity* C is the supremum of all achievable rates.

In Asnani *et al.* [9], an achievable rate was derived for the above setting based on a power splitting scheme, though its optimality was not established therein. In fact, [9] concluded by posing the question as to whether the interference free capacity, given by $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ (Costa [5]), can be achieved even under imperfect state observation. In the sequel, we answer this question in the negative, by showing that a rate loss is incurred even in the relaxed setting where the action sequence is available at the receiver. In the process, we also settle the capacity characterization for the *Gaussian probing capacity* model in [9].

3 Rate Loss due to Imperfect Side Information

We have the following theorem.

Theorem 1. *The capacity for the Gaussian Probing channel is given by*

$$C = \max_{\substack{(P_1,P_2) \\ \varepsilon P_2 + (1-\varepsilon)P_1 \leq P}} \frac{(1-\varepsilon)}{2}\log\left(1+\frac{P_1}{Q+N}\right) + \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right). \tag{3}$$

Remark 2. *Clearly, the interpretation here is that there is a loss in capacity compared to the case of Costa's dirty paper coding setup [5], where the state is perfectly available at the transmitter.*

Proof. As in standard information theoretic proofs, there are two parts to any capacity theorem. The *achievability proof* involves showing that there exists a coding scheme to achieve the given rate. On the other hand, the *converse proof* involves showing that any achievable rate cannot exceed the given rate. The achievability for Theorem 1 is given in Subsection 3.1, while the converse is detailed in Subsection 3.2. ■

3.1 Proof of Achievability

The achievability follows by time-sharing between the strategies of a) dirty paper coding when the side information is not erased ($A=1$) and b) simply treating $S+Z$ as noise when the side information is erased ($A=0$). We assign different input powers in the two scenarios as follows.

$$X = \begin{cases} X|(A=1) = X_2 \sim \mathcal{N}(0,P_2), & \text{if } A=1 \\ X|(A=0) = X_1 \sim \mathcal{N}(0,P_1), & \text{if } A=0 \end{cases}.$$

Thus when $A=1$, which happens for ε fraction of the time, we have $S_e = S$, and so pick

$$U = X_2 + \alpha S,\ \alpha = \frac{P_2}{P_2+N},$$

where $X_2 \perp\!\!\!\perp S$ and $X_2 \sim \mathcal{N}(0,P_2)$. So by dirty paper coding [5], we achieve a rate of $\frac{1}{2}\log\left(1+\frac{P_2}{N}\right)$, for ε fraction of the time. On the other hand, when $A=0$, which happens for $(1-\varepsilon)$ fraction of the time, we have $S_e = *$ and so we simply treat $S+Z$ as the noise, use $X_1 \sim \mathcal{N}(0,P_1)$ to obtain a rate of $\frac{1}{2}\log\left(1+\frac{P_1}{Q+N}\right)$. The power constraint to be satisfied is

$$(1-\varepsilon)P_1 + \varepsilon P_2 \leq P. \tag{4}$$

Thus the average rate

$$R = \frac{(1-\varepsilon)}{2}\log\left(1+\frac{P_1}{Q+N}\right) + \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right), \tag{5}$$

is indeed achievable.

3.2 Proof of Converse

We first note that the capacity in the discrete case is given by the single-letter Gelfand-Pinsker [2] formula i.e.

$$C_{DMC} = \max[I(U;Y) - I(U;S_e)]. \tag{6}$$

This is simply a consequence of the fact that the available side information at the encoder (which appears in the covering negative mutual information term) is S_e. To obtain an

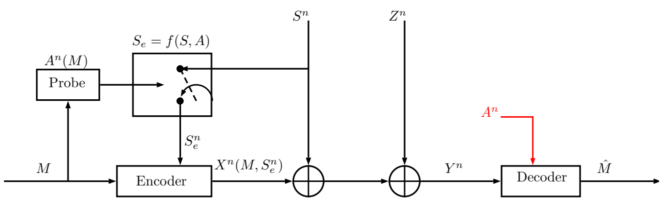

Figure 2. Relaxed Setting in which the action sequence is made available at the decoder

upper bound, we consider a relaxation of the original setting where A^n is made available to the receiver, as shown in Figure 2. The capacity of the relaxed setting would serve as an upper bound for the capacity of the original setting. On the other hand, the capacity of the relaxed setting can be determined to be

$$C_{rel} = \max[I(U;Y,A) - I(U;S_e)]. \tag{7}$$

This is because the Gelfand-Pinsker formula can be applied by considering (Y,A) to be the augmented output. Also note that via the chain rule of mutual information [10], we have

$$I(U;S_e,A) = I(U;S_e) + \cancelto{0}{I(U;A|S_e)}. \tag{8}$$

This in turn implies that (on expanding $I(U;S_e,A)$ the other way round)

$$\begin{aligned} I(U;S_e) &= I(U;A) + I(U;S_e|A) && (9)\\ &= I(U;A) + \varepsilon I(U;S|A=1). && (10) \end{aligned}$$

Hence the upper bound on the capacity of the original channel can be written as

$$\begin{aligned}
R &\leq I(U;Y,A) - I(U;S_e)\\
&\overset{(a)}{=} I(U;Y,A) - I(U;A) - \varepsilon I(U;S|A=1)\\
&= I(U;Y|A) - \varepsilon I(U;S|A=1)\\
&= \varepsilon[I(U;Y|A=1) - I(U;S|A=1)]\\
&\qquad + (1-\varepsilon)I(U;Y|A=0)\\
&= \varepsilon[h(Y|A=1) - h(Y|U,A=1)\\
&\quad - h(S|A=1) + h(S|U,A=1)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(X+S+Z|U,A=0)]\\
&\overset{(b)}{=} \varepsilon[h(Y|A=1) - h(Y|U,A=1,S)\\
&\quad - h(S) + h(S|Y,U,A=1)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(S+Z|U,A=0)]\\
&\overset{(c)}{=} \varepsilon[h(Y|A=1) - h(Y|U,A=1,S)\\
&\quad - h(S) + h(S|Y,U,A=1)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(S+Z)]\\
&\overset{(d)}{\leq} \varepsilon[h(Y|A=1) - h(Z) - h(S) + h(S|Y,A=1)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(S+Z)]\\
&= \varepsilon[h(Y|S,A=1) - h(Z)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(S+Z)]\\
&\leq \varepsilon[h(X+Z|A=1) - h(Z)]\\
&\qquad + (1-\varepsilon)[h(X+S+Z|A=0) - h(S+Z)]\\
&\overset{(e)}{\leq} \frac{\varepsilon}{2}\log\left(\frac{\mathrm{Var}[X|A=1]+N}{N}\right)\\
&\qquad + \frac{(1-\varepsilon)}{2}\log\left(\frac{\mathrm{Var}[X|A=1]+Q+N}{Q+N}\right)\\
&\overset{(f)}{\leq} \frac{\varepsilon}{2}\log\left(\frac{P_2+N}{N}\right) + \frac{(1-\varepsilon)}{2}\log\left(\frac{P_1+Q+N}{Q+N}\right)\\
&= \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right) + \frac{(1-\varepsilon)}{2}\log\left(1+\frac{P_1}{Q+N}\right)\\
&\leq \max_{\substack{(P_1,P_2)\\ \varepsilon P_2+(1-\varepsilon)P_1\leq P}} \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right) + \frac{(1-\varepsilon)}{2}\log\left(1+\frac{P_1}{Q+N}\right),
\end{aligned} \tag{11}$$

where

- (a) follows from expression (10)
- (b) follows since $h(A|B) - h(C|B) = h(A|C,B) - h(C|A,B)$ for any three random variables (A,B,C), the independence of S^n and M, and the fact that X is completely determined by U when $A=0$
- (c) follows since S and Z are independent of (A,U) when $A=0$
- (d) follows since conditioning reduces the entropy
- (e) follows since the Gaussian distribution maximizes differential entropy for a given variance, and
- (f) follows since the different variances involved are $\mathrm{Var}[X|A=1] = P_2$ and $\mathrm{Var}[X|A=0] = P_1$.

Thus we have that any achievable rate satisfies

$$R \leq \max_{\substack{(P_1,P_2) \\ \varepsilon P_2+(1-\varepsilon)P_1 \leq P}} \frac{(1-\varepsilon)}{2} \log\left(1+\frac{P_1}{Q+N}\right) + \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right). \tag{12}$$

This completes the proof of converse. Hence the capacity is indeed characterized by

$$C = \max_{\substack{(P_1,P_2) \\ \varepsilon P_2+(1-\varepsilon)P_1 \leq P}} \frac{(1-\varepsilon)}{2} \log\left(1+\frac{P_1}{Q+N}\right) + \frac{\varepsilon}{2}\log\left(1+\frac{P_2}{N}\right). \tag{13}$$

Remark 3. *Note that this settles the question left open by Asnani et al. [9], regarding the setting* Learning to Write on Dirty Paper*, where cost-constrained binary actions based on the message affect state knowledge at the encoder. We note that our converse bound involves giving the erasure pattern to the receiver, which in turn implies that any message dependent randomization or actions as in [9] will not increase the capacity. The best strategy simply amounts to timesharing between dirty paper coding and treating unknown state as noise. (Note that in Asnani et al. [9],* ε *stands for the cost constraint on the action sequence.)*

4 Numerical Computations

Now by taking $P = Q = N = 1$, we illustrate the trade-off between capacity and probing cost in Figure 3. The capacity is given by the blue curve in Figure 3. We have also depicted the line obtained by naive time-sharing between $\frac{1}{2}\log\left(1+\frac{P}{Q+N}\right)$ and $\frac{1}{2}\log\left(1+\frac{P}{N}\right)$, which is given by the black line. Thus it is evident that the capacity-achieving strategy outperforms naive time-sharing.

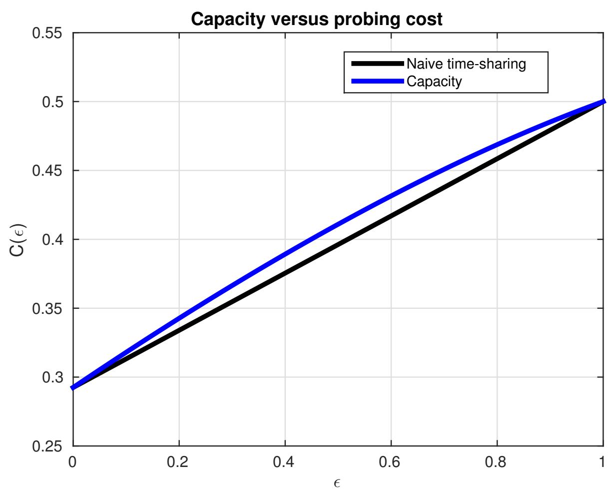

Figure 3. Illustration of the capacity versus probing cost trade-off for $P = Q = N = 1$

5 Conclusion

We revisited the setting of a Gaussian state dependent channel with state observations controlled by message dependent action sequences. It was shown that unlike dirty paper coding, complete state cancellation is not possible. The exact capacity was characterized which shows that a rate loss is incurred compared to the case of full state information known at the encoder, and the best strategy is to simply time share between dirty paper coding and treating the state as noise.

6 Acknowledgements

The work was supported in part by the Bharti Centre for Communication, IIT Bombay and a grant from the ISRO-IITB Space Technology Cell. The author also thanks Dr. Sibi Raj B Pillai, EE, IIT Bombay, for several fruitful discussions.

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, 1958.

[2] S. Gelfand and M. Pinsker, “Coding for channels with random parameters,” *Probl. Contr. Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] A. V. Kuznetsov and B. S. Tsybakov, “Coding in a memory with defective cells,” *Problemy peredachi informatsii*, vol. 10, no. 2, pp. 52–60, 1974.

[4] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

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

[6] H. Weingarten, Y. Steinberg, and S. Shamai, “The capacity region of the gaussian multiple-input multiple-output broadcast channel,” *Information Theory, IEEE Transactions on*, vol. 52, no. 9, pp. 3936–3964, 2006.

[7] T. Weissman, “Capacity of channels with action-dependent states,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5396–5411, 2010.

[8] H. Permuter and T. Weissman, “Source coding with a side information vending machine,” *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4530–4544, 2011.

[9] H. Asnani, H. Permuter, and T. Weissman, “Probing capacity,” *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7317–7332, 2011.

[10] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.