State-Dependent Channels with Composite State Information at the Encoder

Anatoly Khina
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: anatolyk@eng.tau.ac.il

Mustafa Kesal
Email: mkesal@gmail.com

Uri Erez*
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

***Abstract*— State-dependent channels have received much attention over the years, due to their relevance in many different network and multi-user communication scenarios. Nonetheless, previous treatments of this problem assumed that all of the state is available in the same manner: causally, non-causally or non-causally with a finite look-ahead. Yet, in many realistic situations, different parts of the state are known in a different manner. We consider the case where the state is composed of several parts, where each part is known with a different look-ahead. Specifically, we derive the capacity for the case where part of the state is known non-causally to the transmitter, whereas the other part is known only causally, and demonstrate that there are cases in which this capacity can be strictly larger that the capacity of the case where the state is known in a causal fashion, and strictly smaller than the capacity of the same channel, where the state is available non-causally. We note that the treatment in this work provides a unified framework for treating the causal state-information case, the non-causal state-information case, as well as a mixture of the two.**

***Index Terms*— Side information, state-dependent channels, interference, causality, finite look-ahead.**

I. Introduction

The state-dependent discrete memoryless channel (DMC), depicted in Figure 1, is described by an i.i.d. state sequence $s \in \mathcal{S}$ with a probability distribution and channel transition probability distribution

$$p(s) \quad \text{and} \quad p(y|x,s)\,,$$

respectively, where $x \in \mathcal{X}$ is the channel input and $y \in \mathcal{Y}$ is the channel output; and where $\mathcal{X}, \mathcal{Y}$ and $\mathcal{S}$ denote the channel input alphabet, channel output alphabet and state alphabet, respectively, all of which are finite sets. The memoryless property of the channel implies that

$$p(\boldsymbol{y}|\mathbf{x},\boldsymbol{s}) = \prod_{i=1}^{n} p(y_i|x_i,s_i)\,. \tag{1}$$

The first to consider this model was Shannon [1], who assumed that the state sequence is available to the transmitter as "side-information" in a causal manner, i.e., the transmitted symbol at time i, x_i, may depend on all past channel states as well as the channel state at time instance i, s_i. More specifically, the encoder maps the message $w \in \{1, 2, \ldots, 2^{nR}\}$ into $\mathcal{X}^n$ using functions:

$$x_i = f_i\left(w; s_1^i\right), \qquad 1 \le i \le n, \tag{2}$$

where $s_1^i = (s_1, \ldots, s_i)$ are the states up to time i.

The non-causal counterpart of the problem was formulated by Gel'fand and Pinsker [2], who assumed non-causal knowledge of the state sequence s at the transmitter. This allowed them to solve the problem of writing to memories with defects which was considered in [3] (see also [4]). In this scenario, the transmitted symbol at time i is a function of the entire state sequence

$$x_i = f_i\left(w; s_1^n\right), \qquad 1 \le i \le n. \tag{3}$$

In both cases, the receiver decodes the message w from the whole received vector as $\hat{w} = g(y_1^n)$.

The case where the side-information is known non-causally but only up to a finite number of time slots ahead ("finite look-ahead"), was considered in [5] (where it was called "finite anticipation") for the special case of the dirty paper channel (additive white Gaussian channels with additive interference [6]); for this case a lattice-based achievable was derived. The finite look-ahead problem for a general state-dependent DMC was treated by Weissman and El Gamal [7, Sec. VI]; however no single-letter solution for this problem is known.

Nevertheless, in certain communication scenarios, different parts of the channel state are available to the encoder in a different manner, i.e., with different look-ahead lengths. This is the case, for instance, in certain cognitive radio scenarios and dynamic ad-hoc networks, where, e.g., the message to an adjacent user is known in advance ("non-causally"), and can be regarded as a non-causal state, but the channel characteristics are known only in a causal manner or only up to a short look-ahead (depending on the memory of the channel).

In the present work, we consider the case of "composite side information", in which different parts of the channel state are known to the transmitter with different look-ahead lengths. We derive the capacity for the extreme case in which part of the state is known causally ("zero look-ahead") whereas the other – non-causally ("infinite look-ahead").

The rest of the paper is organized as follows: We start by reviewing previously known results for the different cases of availability of side information in Section II. In Section III, we provide a general framework for all of these scenarios and

* This work was supported in part by the U.S. - Israel Binational Science Foundation under grant 2008/455.

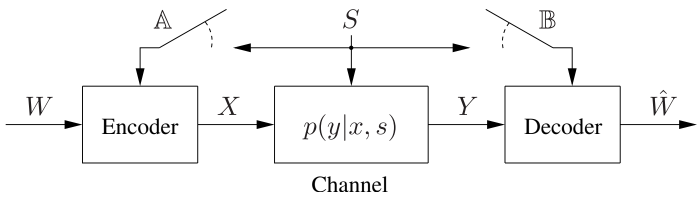

Fig. 1: The discrete memoryless state-dependent channel. When switches $\mathbb{A}$ and/or $\mathbb{B}$ are closed, the state is available at the transmitter and/or receiver, respectively.

derive the capacity for the composite state-dependent scenario where part of the state is known causally and the other – non-causally. We conclude by presenting a few examples for the composite case with causal and non-causal side-information parts, along with a discussion of the results and suggestions for further generalizations, in Section IV.

II. State-Dependent Channel Scenarios

In this section we briefly review the different scenarios considered for the state-dependent DMC, depicted in Figure 1.

When the state S is not available to the transmitter nor to the receiver ($\mathbb{A}$ and $\mathbb{B}$ are open), the channel reduces to a "regular" DMC with transition probability distribution

$$p(y|x) = \sum_{s\in\mathcal{S}} p(s)p(y|x,s). \tag{4}$$

Therefore, the capacity in this case is given by the standard capacity expression

$$C = \max_{p(x)} I(X;Y)\,. \tag{5}$$

Hence, any part of the state known to none of the transmission ends, can be absorbed into the channel transition matrix.

Consider now the case where the state S is known to the encoder but not to the receiver ($\mathbb{A}$ is closed, $\mathbb{B}$ is open) so that the transmitted symbols are governed by (2). For this scenario, Shannon showed [1] that when S is known causally, the capacity is given by the capacity of an associated DMC. The input alphabet of the associated channel, denoted by $\mathcal{T}$, is the set of all possible mappings

$$t : \mathcal{S} \to \mathcal{X}$$

which we refer to as strategies or strategy functions. The output y of the associated channel is related to the input t according to the transition probability

$$p(y|t) \triangleq \sum_{s} p(s)p(y|x = t(s), s)$$

and also

$$p(y_1^n|t_1^n) = \prod_{i=1}^{n} p(y_i|t_i)\,.$$

Thus, the capacity in this case is given by

$$C = \max_{p(t)} I(T;Y)\,, \tag{6}$$

where the maximization is taken over all possible distributions $p(t)$ of the random variable $T \in \mathcal{T}$. We also note that at most $\min\{|\mathcal{S}|(|\mathcal{X}|-1)+1, |\mathcal{Y}|\}$ of the strategies need be given positive probability in order to achieve capacity [8, Ch. 4].

For the case in which the state S is known non-causally, i.e., the encoder knows the entire state sequence in advance, Gel'fand and Pinsker showed, using random binning for the achievability part, that the capacity of this problem is given by [2]

$$C = \max_{p(u|s),\, x=x(u,s)} \{I(U;Y) - I(U;S)\}\,, \tag{7}$$

where the maximum is over all auxiliary variables U, satisfying the Markov relation $U \leftrightarrow (X,S) \leftrightarrow Y$, and the cardinality of U need not exceed $\min\{|\mathcal{S}||\mathcal{X}|, |\mathcal{Y}| + |\mathcal{S}| - 1\}$, and all deterministic functions $x = x(u,s)$.

Remark 1: The capacity of the causal-knowledge case (2) could be achieved using random binning, as in the non-causal case (3), but where the auxiliary U is independent of the state S, thus attaining a rate of

$$R = \max_{p(u)} I(U;Y)\,,$$

which is indeed equal to the capacity in (6).

The case where the encoder, at each time instance, knows the state sequence only up to a finite (constant) number of instances ahead ("finite look-ahead") was treated in [7, Sec. VI]; however no single-letter solution is known for this problem.

Finally, note that the case where the state S (or part of it) is available at the decoder ($\mathbb{B}$ is closed), is equivalent to a channel with an augmented output $\tilde{Y}$ which is composed of the channel output Y and the state S, i.e., $\tilde{Y} = (Y,S)$. Hence, the case in which the state or part of it are available at the decoder needs no special treatment being a special case of the same problem without state knowledge with a different "augmented" output.

In the next section we introduce the problem, in which different parts of the state are known with different look-aheads and derive the capacity for the case in which part of the state is known causally, whereas the other - non-causally.

III. Channel Coding with Composite State Information

The memoryless state-dependent channel with composite state information at the encoder with K parts, is given by (1), with the state S being composed of K parts $\{S_i\}_{i=1}^{K}$, with probability distribution $p(s) = p(s_1, s_2, ..., s_K)$, "part" S_i is known to the encoder with look-ahead of length $\ell_i \in \mathbb{N}$.

As even the capacity of the single finite look-ahead scenario has no (known) single-letter characterization, we limit our focus to the extreme case where the state is composed of two parts, where one is known causally ($\ell_1 = 0$), whereas the other part is known non-causally ($\ell_2 = \infty$).

Theorem 1: The capacity of the state-dependent channel (1) where the state S is composed of two parts $S = (\Gamma, \Lambda)$ with joint probability distribution

$$p(s) = p(\gamma, \lambda)\,,$$

where Γ is known causally (with look-ahead $\ell = 0$) to the encoder and Λ - non-causally ($\ell = \infty$), is given by

$$C = \max_{p(u|\lambda),\, x(u,\lambda,\gamma)} \left[I(U;Y) - I(U;\Lambda) \right] , \qquad (8)$$

where the maximization is over all auxiliary variables U, which, given Λ, are independent of Γ and satisfy $U \leftrightarrow (X, \Gamma, \Lambda) \leftrightarrow Y$, and over all *deterministic* functions $x = x(u, \lambda, \gamma)$. The cardinality of the auxiliary random variable is bounded by

$$|\mathcal{U}| \leq \min\{|\Lambda| \left[1 + |\Gamma|(|\mathcal{X}| - 1)\right], |\mathcal{Y}| + |\Lambda| - 1\} .$$

Proof:

Achievability: To achieve a rate as given (8), we use Shannon strategies t which map the causally known side-information γ to a channel input $x = t(\gamma)$. We now may view the resulting channel as a channel whose inputs are all possible strategies t, as in (6), with non-causal side-information λ available at the encoder. According to (7), the rate given by

$$R = \max_{p(u|\lambda), t_{u,\lambda}} \left[I(U;Y) - I(U;\Lambda) \right] \qquad (9)$$

is achievable for this channel, where by $t_{u,\lambda}$ we mean that t is a deterministic function of (u, λ). Since $x = t(\gamma)$, (9) can be written as

$$\begin{aligned} R &= \max_{p(u|\lambda), x = t_{u,\lambda}(\gamma)} \left[I(U;Y) - I(U;\Lambda) \right] \\ &= \max_{p(u|\lambda), x = x(u,\lambda,\gamma)} \left[I(U;Y) - I(U;\Lambda) \right] . \end{aligned}$$

Thus (8) is achievable.

Converse: The converse follows along the same lines of the converse for the non-causal side-information (only) problem (7), as presented in [9].

$$\begin{aligned} n(R - \epsilon_n) &\overset{(a)}{\leq} I(W; Y_1^n) \\ &\overset{(b)}{=} \sum_{i=1}^{n} I(W; Y_i | Y_1^{i-1}) \\ &\overset{(c)}{\leq} \sum_{i=1}^{n} I(W, Y_1^{i-1}; Y_i) \\ &\overset{(d)}{=} \sum_{i=1}^{n} I\left(W, Y_1^{i-1}, \Lambda_{i+1}^n; Y_i\right) - I\left(Y_i; \Lambda_{i+1}^n | W, Y_1^{i-1}\right) \\ &\overset{(e)}{=} \sum_{i=1}^{n} I\left(W, Y_1^{i-1}, \Lambda_{i+1}^n; Y_i\right) - I\left(Y_1^{i-1}; \Lambda_i | W, \Lambda_{i+1}^n\right) \\ &\overset{(f)}{=} \sum_{i=1}^{n} I\left(W, Y_1^{i-1}, \Lambda_{i+1}^n; Y_i\right) - I\left(W, \Lambda_{i+1}^n, Y_1^{i-1}; \Lambda_i\right) \\ &\overset{(g)}{=} \sum_{i=1}^{n} I(U_i; Y_i) - I(U_i; \Lambda_i) \\ &\leq \max_{p(u|\lambda), x(u,\gamma,\lambda)} \left\{ I(U;Y) - I(U;\Lambda) \right\} , \qquad (10) \end{aligned}$$

where (a) follows from Fano's inequality with $\epsilon_n \to 0$ for $n \to \infty$; (b), (c) and (d) follow from the chain-rule for mutual information; (e) holds true by the Csiszár-Körner identity [10, Chap. 3, §4]; in (f) we used the fact Λ_i is independent of (W, Λ_{i+1}^n); and in (g) we defined $U_i \triangleq (W, Y_1^{i-1}, \Lambda_{i+1}^n)$. Note that this choice of U satisfies $U \leftrightarrow (X, \Lambda, \Gamma) \leftrightarrow Y$ and is independent of Γ given Λ.

We shall now show that the maximum in (10) is achieved for a deterministic function, i.e., $x = x(u, \gamma, \lambda)$. Fix $p(u|s)$. Hence the second term in (10), $I(U;S)$, is constant, suggesting that maximization needs to be carried over the first term $I(U;Y)$ only. We now observe that

$$p(y|u) = \sum_{x,\lambda,\gamma} p(y|x,\gamma,\lambda) p(x|u,\gamma,\lambda) p(\lambda,\gamma|u)$$

is linear in $p(x|sC, sNC)$, since $p(y|x, \gamma, \lambda)$ is fixed, $p(\lambda, \gamma|u)$ is fixed given $p(u|s)$ and the Markov chain relation $U \leftrightarrow (X, \Gamma, \Lambda) \leftrightarrow Y$. Since $I(U;Y)$ is convex in $p(y|u)$ it is therefore convex in $p(x|u, \gamma, \lambda)$, which implies in turn that it is maximized by a deterministic mapping $x = x(u, \gamma, \lambda)$.

Cardinality Bound: The bound for the auxiliary random variable cardinality is derived in a similar manner to the cardinality bound for the Gel'fand-Pinsker problem in [9]. ■

Remark 2: For pure causal state information (at the transmitter) case, the cardinality bound reads $|\mathcal{U}| \leq \min\{|\mathcal{S}|(|\mathcal{X}| - 1) + 1, |\mathcal{Y}|\}$, whereas for the pure non-causal case, it takes the form $|\mathcal{U}| \leq \min\{|\mathcal{S}||\mathcal{X}|, |\mathcal{Y}| + |\mathcal{S}| - 1\}$. When we compare the right flanks of these expression, the non-causality reflects on the result with a contribution of $|\mathcal{S}| - 1$. In our case, this factor takes the form $|\Lambda| - 1$, since Λ is the non-causal part of the available state information. As for the left flank components, consider the random variable $V \triangleq (U|\Lambda = \lambda_0)$, i.e., the auxiliary random variable conditioned on the non-causal state information. For each instance of Λ, the cardinality of V is bounded by $|V| \leq |\Gamma|(|\mathcal{X}| - 1) + 1$. Since there are $|\Lambda|$ possible different choices, $|\mathcal{U}| \leq |\Lambda| \left[1 + |\Gamma|(|\mathcal{X}| - 1)\right]$ follows.

Remark 3: When the (entire) state S is known causally (by setting $\Lambda = \lambda_0$, i.e., it is deterministic and known by both ends of transmission), respectively non-causally (by setting $\Gamma = \gamma_0$), (8) reduces to (6), respectively (7).

Remark 4: The capacity (8) in Theorem 1 can be achieved using a single random binning scheme without the need to split the scheme into a two-step scheme as was done in the proof of Theorem 1.

Remark 5: Additional parts of the state which are known pure-causally, i.e., with negative look-ahead ($l < 0$) cannot increase capacity. This is seen from the proof of the converse of Theorem 1.

IV. Examples

In this section we further discuss the composite side-information scenario, where part is known causally and the other – non-causally, treated in Theorem 1. We start by considering three examples: In Example 1 we show a case where knowing part of the state non-causally does not increase capacity beyond the all-causal side information case; in Example 2 we present a case in which the capacity when part of

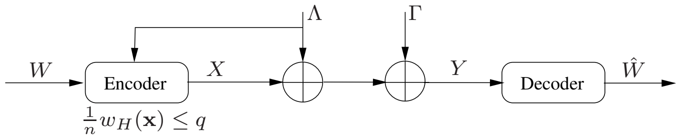

Fig. 2: Binary dirty paper channel with two interferences.

the state is known causally and the other non-causally is equal to the capacity as if both parts were available non-causally, which in turn is strictly greater than the capacity of the same channel with only causal side-information; finally, Example 3 demonstrates a case in which the capacity of the composite side-information case is strictly greater than that of the all-causal side-information case, yet strictly smaller than that of the case in which all of the state is available non-causally.

Example 1 (Binary dirty paper channel with two interferences): Consider the (noiseless) "binary dirty paper channel" with two interferences (depicted also in Figure 2):

$$Y = X \oplus \Gamma \oplus \Lambda \,,$$

where $X, Y, \Gamma, \Lambda \in \mathbb{Z}_2$ and $\oplus$ denotes addition modulo-2 (XOR). The states Γ and Λ are independent, i.i.d. with distribution Bernoulli$(1/2)$ and are known at the transmitter. The input is subject to a ("power") constraint $\frac{1}{n}w_H(\mathbf{x}) \le q$, where $0 < q < 1/2$, $w_H(\cdot)$ denotes Hamming weight, and n is the length of the codeword.

When Λ and Γ are known causally, the problem reduces to that of (only) causal side-information, the capacity of which is given in (6), and was explicitly found in [11] to equal $C_{\text{causal}} = 2q$. On the other hand, when both Λ and Γ are known non-causally, the capacity is equal to [11] $C_{\text{noncausal}} = H_b(q)$, where $H_b(\cdot)$ denotes the binary entropy function. This in turn is equal to the capacity of the interference-free case ($\Lambda = \Gamma = 0$) and is strictly larger than the capacity of the causal case, for $0 < q < 1/2$. Considering the case in which Λ is known non-causally whereas Γ – only causally, one observes that non-causal knowledge of Λ does not help beyond the causal knowledge scenario, i.e., $C_{\text{composite}} = C_{\text{causal}} = 2q$. This happens since even if Λ were known at both sides (or alternatively, equal to 0), the channel would reduce to the case of a channel with a single causally-known interference Γ, the capacity of which is $2q$. Thus the capacity of the composite side-information scenario, is equal to the causal side-information capacity in this case and does not provide further improvement:

$$C_{\text{causal}} = C_{\text{composite}} < C_{\text{noncausal}} \,.$$

Example 2 (Binary dirty paper channel with erasures): In this example we consider a state-dependent channel with binary input and ternary output, where part of the state determines, at each time instance, whether an erasure, which corresponds to one of the channel outputs, occurs or not (in the latter case only one of the two other channel output outcomes is possible). This channel, depicted in Figure 3, is described by

$$Y = \begin{cases} X \oplus \Lambda & \Gamma = 0 \\ \varepsilon & \Gamma = 1 \end{cases} , \tag{11}$$

where $X, \Gamma, \Lambda \in \mathbb{Z}_2$, $Y \in \{0, 1, \varepsilon\}$, Γ and Λ are independent with distribution Bernoulli$(1/2)$ and are known at the transmitter, and X is subject to a ("power") constraint $\frac{1}{n}w_H(\mathbf{x}) \le q$, where $0 < q < 1/2$. This channel could be thought of as a binary dirty paper channel with erasures, where the erasures are available as side information at the encoder, in addition to the additive interference.

The scenario in which both Γ and Λ are known only causally falls under the framework considered by Shannon, the capacity of which is given in (6). For our channel of interest (11), the (all-causal) capacity is equal to

$$C_{\text{causal}} = \min\{2q, 1/2\} \,;$$

note that when $q \le 1/4$ the capacity is equal to that of the same channel when no erasures are possible (i.e., when $\Gamma = 0$). This happens, since both the transmitter and the receiver are aware of erasure occurrences: The encoder – due to its causal side-information, and the decoder – since it observes a distinct outcome in case of an erasure. Thus, the encoder and decoder can ignore channel uses in which erasures occur, and in this manner "save power" that may be used during the rest of the channel uses.

The capacity of the Gel'fand-Pinsker scenario, in which Γ and Λ are known non-causally, given in (7), can be shown to equal

$$C_{\text{noncausal}} = \frac{1}{2} H_b\left(\min\left\{2q, \frac{1}{2}\right\}\right) \,,$$

for the channel of interest (11). Again note that since both transmission ends are aware of the exact places erasure occurred, they would only use the remaining channel uses to convey information. Since in the limit of large block-length n, the number of erasures is $n/2$, one is left with only $n/2$ channel uses over which information may be transmitted. Thus knowledge of the erasures in advance is superfluous, as for large n values, it is guaranteed that the number of erasures is equal (approximately) to $n/2$ and the times of their occurrences are known at both ends. Hence, in this case the capacity of the composite side-information case, in which Γ is known causally and Λ - non-causally, is equal to the case in which both are known non-causally, yet larger than the capacity of the case in which both parts are known only causally. For $q < 1/4$, these results can be summarized as follows:

$$C_{\text{causal}} < C_{\text{composite}} = C_{\text{noncausal}} \,.$$

Example 3 (Binary dirty paper channel with product interference): Consider now the binary dirty paper channel with a product interference (depicted in Figure 4):

$$Y = X \oplus \Gamma\Lambda \,,$$

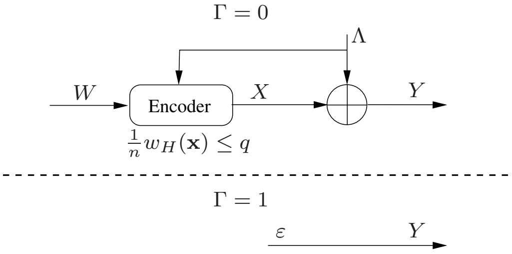

Fig. 3: Binary dirty paper channel with erasures.

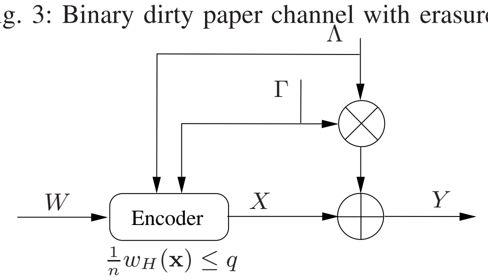

Fig. 4: Binary dirty paper channel with a product interference.

where $X, Y, \Gamma, \Lambda \in \mathbb{Z}_2$, Γ and Λ are independent with distribution Bernoulli$(1/2)$ and are known at the transmitter, and the input is subject to a ("power") constraint $\frac{1}{n}w_H(\mathbf{x}) \leq q$. Note that the additive interference is equal to the product of the interferences Γ and Λ.

When both Λ and Γ are known non-causally, this problem reduces to the non-causal side-information problem (7), the capacity of which is [11] $C_{\text{noncausal}} = H_b(q)$.

When Λ and Γ are known only causally, the problem reduces to that of (only) causal side-information, the capacity of which is given in (6) and is *strictly* smaller than that of its non-causal counterpart, as depicted in Figure 5.

Finally, consider the case in which Λ is known non-causally whereas Γ is available in a causal manner. This case falls under the framework of Theorem 1 and its capacity is given by (8) and is strictly larger than that of the causal case and strictly smaller than the capacity in the non-causal case, as depicted in Figure 5. This is true since Λ is known in advance ("non-causally") and due to the structure of the channel (the interferences are multiplied), we can anticipate approximately 2/3 of the time instances in which the interference is equal to zero. This allows us to achieve higher rates in these time instances than could be achieved in the causal case. Nonetheless, since not all the interference sequence can be anticipated in advance, as is the case in the "all non-causal" scenario, one cannot hope to achieve the same rate of the latter case. Thus, this channel demonstrates a case in which partial non-causal side-information in addition to causal one, improves performance, yet additional non-causal information beyond the part known only causally can assist further to improve performance:

$$C_{\text{causal}} < C_{\text{composite}} < C_{\text{noncausal}} \, .$$

V. Conclusions

The repercussion of the present work is twofold: Establishing a general framework for the problem of state-dependent channels with different parts of the state available with different look-ahead lengths, as well as providing a unified treatment of the different side-information scenarios, which were previously considered and treated separately. This is possibly by noting that the right flank of (c) in (10) is the same both in the converses of Shannon [1] and Gel'fand and Pinsker [2]. Thus, both the causal and the non-causal scenarios can be treated simultaneously. Finally, we note that for the general case, where the state is composed of parts known with different finite look-ahead lengths, similar results to those of Weissman and El Gamal [7] for the single-part finite look-ahead, may be derived.

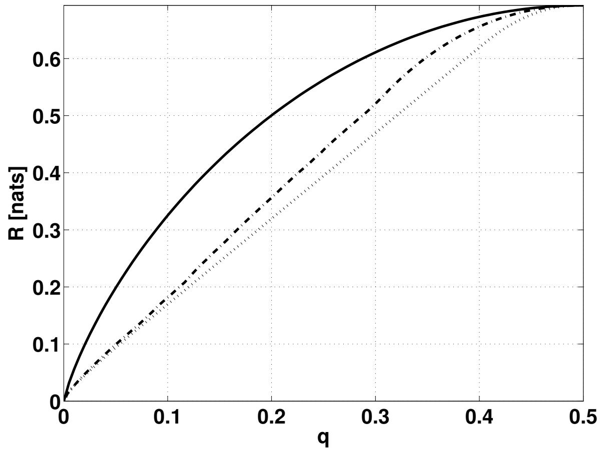

Fig. 5: Capacities in nats of the different side-information scenarios of Example 3, as a function of the Hamming input constraint q. Continuous line – Γ and Λ are known non-causally; dashed-dotted line - Γ is known causally and Λ – non-causally; dotted line – Γ and Λ are know causally.

References

[1] C. E. Shannon. Channels with side information at the transmitter. *IBM Journal of Research and Development*, 2:289–293, Oct. 1958.
[2] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problemy Pered. Inform. (Problems of Inform. Trans.)*, 9, No. 1:19–31, 1980.
[3] A. V. Kuznetsov and B. S. Tsybakov. Coding in a memory with defective cells. *translated from Prob. Peredach. Inform.*, 10:52–60, April-June, 1974.
[4] C. Heegard and A. El Gamal. On the capacity of computer memory with defects. *IEEE Trans. Information Theory*, IT-29:731–739, Sept. 1983.
[5] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice strategies for cancelling known interference. *IEEE Trans. Information Theory*, pages 3820–3833, Nov. 2005.
[6] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Information Theory*, IT-29:439–441, May 1983.
[7] T. Weissman and A. El Gamal. Source coding with limited-look-ahead side information at the decoder. *IEEE Trans. Information Theory*, 52:5218–5239, Dec. 2006.
[8] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, N.Y., 1968.
[9] A. El Gamal and T.-H. Kim. *Lecture notes on network information theory. arXiv:1001.3404v4 [cs.IT].*
[10] I. Csiszár and J. Körner. *Information Theory - Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.
[11] R. J. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Information Theory*, 49:1159–1180, May 2003.