Source Coding with Composite Side Information at the Decoder

Anatoly Khina, Uri Erez
Dept. of EE-Systems, Tel Aviv University
Email: {anatolyk,uri}@eng.tau.ac.il

***Abstract*—We consider a source-coding scenario in which composite side information is available at the decoder, viz. part of it is known non-causally, whereas the other part is available only causally — by this combining the treatment of Wyner and Ziv with that of Weissman and El Gamal. We then consider the joint source–channel coding problem of transmitting a source with composite side information at the decoder over a channel with composite side information at the encoder. We show that the separation principle between source coding with composite side information and its dual channel problem holds true, thus extending the result of Merhav and Shamai to the case of composite side informations. These results provide a unified framework for treating the causal side information case, the non-causal state-information case, as well as a mixture of the two.**

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

I. INTRODUCTION

The discrete memoryless (DMS) source with side information, depicted in Figure 1, is composed of a source sequence $u_1^n = (u_1, \ldots, u_n)$ and side information sequence $q_1^n = (q_1, \ldots, q_n)$, which are drawn jointly, in an i.i.d. manner, from a joint probability mass function p:

$$p^n(u_1^n, q_1^n) = \prod_{i=1}^{n} p(u_i, q_i) \,, \quad u_i \in \mathcal{U}, q_i \in \mathcal{Q} \,, \tag{1}$$

where $\mathcal{U}$ and $\mathcal{Q}$ are finite sets denoting the source and side-information alphabets, respectively.

The encoder maps the source sequence u_1^n to an index j according to a mapping $f : \mathcal{U}^n \to \left\{1, \ldots, \lfloor 2^{nR} \rfloor\right\}$, i.e.,

$$j = f\left(u_1^n\right) \,,$$

and sends this index to the decoder. The decoder, in turn, recovers a (distorted) reconstruction $\hat{u}_1^n$ ($\hat{u}_i \in \hat{\mathcal{U}}$) of the source sequence u_1^n from the index j sent by the encoder and its side information sequence q_1^n, according to a (deterministic) mapping $\mathbf{g} : \left\{1, \ldots, \lfloor 2^{nR} \rfloor\right\} \times \mathcal{Q}^n \to \hat{\mathcal{U}}^n$:

$$\hat{u}_1^n = \mathbf{g}\left(j, q_1^n\right) = \mathbf{g}\left(f\left(u_1^n\right), q_1^n\right) \,.$$

The problem is to find the optimum rate–distortion function (RDF) $R(D)$, i.e., what is the smallest rate R needed to achieve a given average per-symbol distortion D, with respect to some distortion measure ρ.

This work was supported in part by the Israel Science Foundation under Grant 1557/12.

Different scenarios were considered for this problem, which differ from each other by the way $\hat{u}_1^n$ may depend on the side information sequence q_1^n.

In the most natural scenario $\hat{u}_1^n$ may depend on q_1^n in an arbitrary manner. The optimum trade-off between rate and distortion for this case was established in the 1976 seminal paper by Wyner and Ziv [1]. We shall refer to this scenario as the "non-causal" one, as each reconstructed source symbol $\hat{u}_i$ may depend on the whole side information sequence q_1^n. Thus, $\hat{u}_i$ is of the form

$$\hat{u}_i = g_i\left(j, q_1^n\right) \,, \tag{2}$$

where $\{g_i\}$ are the entries of $\mathbf{g}$, namely, $\mathbf{g} = (g_1, \ldots, g_n)$.

A more restrictive setting of the source-coding problem with side information at the decoder is the case where the decoder is constrained to anticipate and use only $\ell \geq 0$ future side information symbols, where ℓ is referred to as the "finite look-ahead" parameter. Thus, the reconstructed source symbol $\hat{u}_i$ may depend on the side information symbols $q_1^{i+\ell}$ but not on $q_{\ell+1}^n$:[1]

$$\hat{u}_i = g_i\left(j, q_1^{i+\ell}\right) \,. \tag{3}$$

This problem was treated by Weissman and El Gamal in [2], in which upper and lower bounds on the RDF for $0 < \ell < \infty$ were derived and the RDF for the "causal side information" was determined, where the latter corresponds to $\ell = 0$ and for which

$$\hat{u}_i = g_i\left(j, q_1^i\right) \,. \tag{4}$$

Note that for $\ell \to \infty$, this setting coincides with the original setting of Wyner and Ziv (2). Thus, the finite look-ahead setting may be regarded as a generalization of the original (unconstrained) Wyner–Ziv setting (2).

In this work, we consider yet a more general "composite side information" scenario, in which different parts of the side information are available to the decoder in a different manner, i.e., with different look-ahead lengths. Specifically, we derive the RDF for the extreme case in which part of the side information is known causally ("zero look-ahead") whereas the other — non-causally ("infinite look-ahead").

The problem of source coding with composite side information is dual to the problem of channel coding with composite

[1] In case $i + \ell$ exceeds n, q_1^n is taken rather than $q_1^{i+\ell}$.

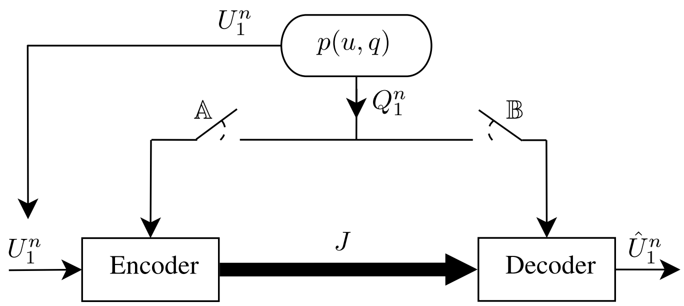

Fig. 1: Discrete memoryless source with side information. When switches $\mathbb{A}$ and/or $\mathbb{B}$ are closed, the side information is available at the encoder and/or decoder, respectively.

side information at the encoder, introduced and treated recently in [3]. In the second part of this work, we consider the joint source–channel coding problem of transmitting a source with composite side information at the decoder over a channel with composite side information at the encoder. we follow the steps of Merhav and Shamai [4], and prove that a separation between the source coding task and the channel coding task holds also for the case of composite side informations.

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 determine the rate–distortion function for the composite side information scenario where part of the side information is known causally and the other – non-causally. We then treat the problem of joint source–channel coding of a source with composite side information conveyed over a channel with composite side information in Section IV, and prove that the separation principle extends to this case, as well. Finally, we conclude the results in Section V.

II. Source-Coding with Side Information Scenarios

In this section we briefly review the different scenarios considered for the memoryless source with side information, depicted in Figure 1.

When the state q is not available at the encoder nor at the decoder ($\mathbb{A}$ and $\mathbb{B}$ are open) the problem is that of the "regular" rate–distortion problem, the RDF of which is given by

$$R(D) = \min_{p(\hat{u}|u)} I(U;\hat{U})\,, \tag{5}$$

where the minimization is over all admissible conditional probability mass functions $p(\hat{u}|u)$ satisfying the average distortion constraint

$$E\left[\rho(U,\hat{U})\right] \le D\,.$$

Consider now the case where $\mathbb{A}$ is open and $\mathbb{B}$ is closed, i.e., the case in which the side information Q is available at the decoder (but not at the encoder). When the side information Q is available in a non-causal manner (2), in the lossless case, viz. $D=0$, the solution to this problem is a corner point of the achievable region of the lossless distributed source-coding problem of Slepian and Wolf [5], and is given by

$${}^{\text{nc}}R(D=0) = H(U|Q)\,.$$

For its lossy counterpart, the solution was shown, by Wyner and Ziv [1] to be the solution of the single-letter optimization problem

$$\begin{aligned} {}^{\text{nc}}R(D) &= \min_{p(w|u),g} \left[I(U;W|Q)\right] \\ &= \min_{p(w|u),g} \left[I(U;W) - I(Q;W)\right]\,, \end{aligned} \tag{6}$$

where the minimization is over all auxiliary variables w (conditional probability mass functions $p(w|u)$) and deterministic functions $g : \mathcal{W}\times\mathcal{Q}\to\hat{\mathcal{U}}$, such that the distortion constraint

$$E[\rho(U,g(W,Q))] \le D$$

is satisfied. Moreover, the cardinality of $\mathcal{W}$ need not exceed $\mathcal{U}+1$.

Remark 1: The auxiliary variable W of (6) satisfies a Markov chain relation $W \leftrightarrow U \leftrightarrow Q$.

For the case in which the state is available only causally (4), Weissman and El Gamal [2] showed the RDF to be given by

$${}^{\text{c}}R(D) = \min_{p(w|u),g} I(U;W)\,,$$

with the minimization carried over the same set as in (6). They further reinterpreted this result as a rate–distortion problem with no side-information (as in (5)) with a new derived distortion measure

$$\tilde{\rho}\,(u,t) = E\left[\rho\,(u,t(Q))|U=u,T=t\right]\,, \tag{7}$$

where $t\in\mathcal{T}$ is a mapping $t:\mathcal{Q}\to\hat{\mathcal{U}}$, s.t.

$$\hat{u} = g\,(w,q) = t(q)\,,\quad \forall q\in\mathcal{Q} \tag{8}$$

and $\mathcal{T}$ is the set of all such mappings whose cardinality is $|\mathcal{T}| = |\hat{\mathcal{U}}|^{|\mathcal{Q}|}$. Thus, the RDF for the causal side information case is equal to

$${}^{\text{c}}R(D) = \min_{p(t|u)} I(U;T)\,. \tag{9}$$

In the case of side information of finite-lookahead (3), the RDF has no known single-letter characterization. Nevertheless, upper and lower bounds for this problem were established in [2], which become tight with the increase in computational complexity allowed.

Finally, note that the case where the state Q (or part of it) is available at the encoder ($\mathbb{B}$ is closed), is equivalent to an augmented source $\tilde{U}$, composed of the (physical) source U and the side information Q, i.e., $\tilde{U}=(U,Q)$, with the distortion measure being only w.r.t. its first part (the "original source"), i.e., $\tilde{\rho}\left(\tilde{u},\hat{\tilde{u}}\right) = \rho\,(u,\hat{u})$. Hence, the case in which the state or part of it are available at the encoder needs no special treatment being a special case of the same problem setting without state knowledge with a different "augmented" source.

III. Source Coding with Composite Side Information

The memoryless source with composite K-part side information at the decoder, is given by (1), with the side information Q being composed of K parts $\{Q_i\}_{i=1}^{K}$, with probability distribution $p(q) = p(q(1), q(2), ..., q(K))$, where "part" $q(i)$ is known to the decoder with look-ahead length $\ell_i \in \mathbb{N} \cup \{0, \infty\}$.

As even the rate–distortion function of the single finite look-ahead scenario has no (known) single-letter characterization (see [2]), we limit our focus to the extreme case where the state is composed of two parts, where one is available causally ($\ell_1 = 0$), whereas the other part is available in a non-causal manner ($\ell_2 = \infty$).

Theorem 1: The rate–distortion function with composite side information at the decoder, where the state q is composed of two parts, $q = ({}^{\mathrm{c}}q, {}^{\mathrm{nc}}q)$ with joint probability distribution

$$p(q) = p({}^{\mathrm{c}}q, {}^{\mathrm{nc}}q) \,,$$

where ${}^{\mathrm{c}}q$ is known causally (with look-ahead $\ell = 0$) to the decoder and ${}^{\mathrm{nc}}Q$ — non-causally ($\ell = \infty$), is given by

$$\begin{aligned} {}^{\text{c-nc}}R(D) &= \min I(U; W|{}^{\mathrm{nc}}Q) \\ &= \min \left[I(W;U) - I(W; {}^{\mathrm{nc}}Q)\right] \,, \end{aligned} \tag{10}$$

where the minimization is over all conditional pmfs $p(w|u)$ and functions $g : \mathcal{W} \times \mathcal{Q} \to \hat{\mathcal{U}}$ satisfying

$$E[\rho(U, g(W,Q))] \le D \,. \tag{11}$$

Proof:

Achievability: In order to achieve the desired rate (10), we use the random strategies interpretation given in (8) w.r.t. the causal side information ${}^{\mathrm{c}}q$. This allows to view the resulting problem as a rate–distortion problem without causal side information and with a different derived distortion measure $\tilde{\rho}$ given in (7), but rather with (only) non-causal side information ${}^{\mathrm{nc}}q$. For this problem, the result of (6) may readily be applied, resulting in the rate

$$R = \min_{{}^{\mathrm{nc}}g, p(w|u)} I\left(U; W|{}^{\mathrm{nc}}Q\right) \,,$$

where the minimization is carried over all deterministic functions ${}^{\mathrm{nc}}g : \mathcal{W} \times {}^{\mathrm{nc}}\mathcal{Q} \to {}^{\mathrm{c}}\mathcal{T}$ satisfying the distortion constraint

$$E\left[\tilde{\rho}\left(U, {}^{\mathrm{nc}}g\left(W, {}^{\mathrm{nc}}Q\right)\right)\right] \le D \tag{12}$$

and where ${}^{\mathrm{c}}\mathcal{T}$ is the set of all mappings from ${}^{\mathrm{c}}\mathcal{Q}$ to $\hat{\mathcal{U}}$. Note that, due to the law of total expectation, (12) is equivalent to (11), as in (9). Finally, note that optimizing over the set of all possible ${}^{\mathrm{nc}}g : \mathcal{W} \times {}^{\mathrm{nc}}\mathcal{Q} \to {}^{\mathrm{c}}\mathcal{T}$ is equivalent to optimizing over all possible ${}^{\mathrm{nc}}g : \mathcal{W} \times {}^{\mathrm{nc}}\mathcal{Q} \times {}^{\mathrm{c}}\mathcal{Q} \to \hat{\mathcal{U}}$.

Converse: The converse follows the same lines of the (only) non-causal side–information case of Wyner and Ziv, as presented in [6]:

$$\begin{aligned}
nR &\ge H(J) && (13)\\
&\ge H\left(J|{}^{\mathrm{nc}}Q_1^n\right) && (14)\\
&\ge I\left(U_1^n; J|{}^{\mathrm{nc}}Q_1^n\right) && (15)\\
&= \sum_{i=1}^{n} I\left(U_i; J\middle|U_1^{i-1}, {}^{\mathrm{nc}}Q_1^n\right) && (16)\\
&= \sum_{i=1}^{n} \Big[I\left(U_i; J, U_1^{i-1}, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_{i+1}^n \middle| {}^{\mathrm{nc}}Q_i\right) \\
&\qquad - I\left(U_i; U_1^{i-1}, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_{i+1}^n \middle| {}^{\mathrm{nc}}Q_i\right) \Big] && (17)\\
&= \sum_{i=1}^{n} I\left(U_i; J, U_1^{i-1}, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_{i+1}^n, {}^{\mathrm{c}}Q_1^{i-1} \middle| {}^{\mathrm{nc}}Q_i\right) && (18)\\
&\ge \sum_{i=1}^{n} I\left(U_i; J, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_{i+1}^n, {}^{\mathrm{c}}Q_1^{i-1} \middle| {}^{\mathrm{nc}}Q_i\right) && (19)\\
&= \sum_{i=1}^{n} I\left(U_i; W_i|{}^{\mathrm{nc}}Q_i\right) && (20)\\
&\ge \sum_{i=1}^{n} R\left(E\left[\rho\left(U_i, g^*\left(W_i, Q_i\right)\right)\right]\right) && (21)\\
&\ge nR\left(E\left[\frac{1}{n}\sum_{i=1}^{n} \rho\left(U_i, g^*\left(W_i, Q_i\right)\right)\right]\right) && (22)\\
&\ge nR\left(D\right) \,, && (23)
\end{aligned}$$

where

(13) follows from the fact that the cardinality of the alphabet of J is 2^{nR},

(14) since conditioning reduces entropy,

(15) follows from the definition of mutual information and non-negativity of entropy,

(16) from the chain-rule for mutual informations,

(17) from the chain rule for mutual informations as well,

(18) from the memoryless property (1),

(19) holds since

$$U_i \leftrightarrow \left(J, U_1^{i-1}, {}^{\mathrm{nc}}Q_i\right) \leftrightarrow \left({}^{\mathrm{nc}}Q_{i+1}^n, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{c}}Q_1^{i-1}\right)$$

forms a Markov chain,

(19) follows since conditioning reduces entropy,

(20) by defining $W_i \triangleq \left(J, {}^{\mathrm{c}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_1^{i-1}, {}^{\mathrm{nc}}Q_{i+1}^n\right)$

(21) from the definition of the information conditional rate distortion function with composite side information where g^* is the function achieving the minimum in (10),

(22) follows from Jensen's inequality and the convexity of the information conditional rate distortion function with composite side information (which can be proved like the convexity in the non-causal case; see, e.g., [7]),

(23) follows from the definition of the average distortion.

The bound on the cardinality of $\mathcal{W}$ is established via a standard application of Carathéodory's theorem, as done in [1].

Finally, the equality to the second expression in (10) is established as in (6):

$$\begin{aligned} I\left(W; U, {}^{\mathrm{nc}}Q\right) &= I\left(W; {}^{\mathrm{nc}}Q\right) + I\left(W; U|{}^{\mathrm{nc}}Q\right) && (24)\\ &= I\left(W; U\right) + I\left(W; {}^{\mathrm{nc}}Q|U\right) && (25) \end{aligned}$$

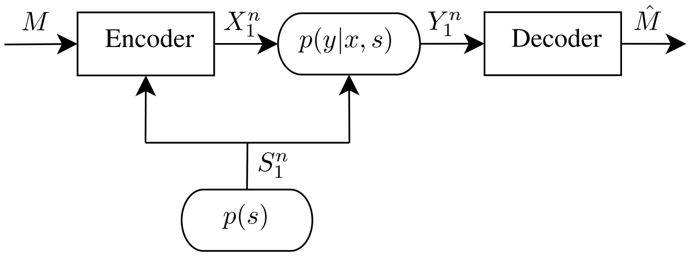

Fig. 2: State-dependent channel with state side information available at the encoder. M and $\hat{M}$ denote the transmitted and recovered message, respectively.

where (24) and (25) follow from the chain rule for mutual informations. Finally, by noting that $I\left(W; {}^{\text{nc}}Q|U\right) = 0$ due to the Markov chain of Remark 1, $W \leftrightarrow U \leftrightarrow Q$, the equality between the two expressions in (10) follows. ■

Remark 2: The result of Theorem 1 extends to infinite and continuous alphabets, under mild regularity conditions, as is done in [8].

IV. Separation between Source and Channel Coding with Composite Side Informations

In this section we consider the related joint source–channel coding problem of transmitting a memoryless source over a state-dependent memoryless channel, where *channel side information* (a.k.a. *state information*) is available at the encoder and *source side information* — at the decoder.

The state-dependent memoryless channel, depicted in Figure 2, 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

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

When the side information s is known causally at the encoder, Shannon determined the capacity of this channel to equal [9]

$${}^{\text{c}}C = \max_{p(w),h} I(W;Y)\,, \tag{26}$$

where W is an auxiliary variable independent of S and $h : \mathcal{W} \times \mathcal{S} \to \mathcal{X}$ is a deterministic mapping, such that $x = h(w,s)$.[2]

[2]Shannon represented the solution to this problem as the capacity of an equivalent memoryless channel with no side information whose inputs are all possible mappings from $\mathcal{S}$ to $\mathcal{X}$, similar to (8).

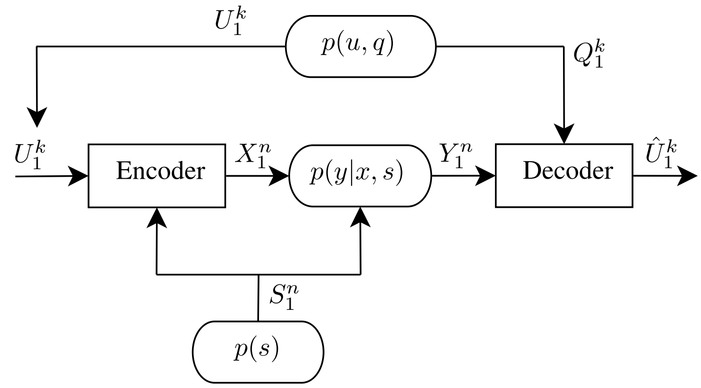

Fig. 3: Joint source–channel coding of a source with side information at the decoder and state-dependent channel with state side information at the encoder.

Gel'fand and Pinsker showed that the capacity of this channel, when the side information sequence is available non-causally at the encoder, is given by [10]

$${}^{\text{nc}}C = \max_{p(w|s),h} I(W;Y) - I(W;Y)\,, \tag{27}$$

where W is an auxiliary (which may depend on S) and $h : \mathcal{W} \times \mathcal{S} \to \mathcal{X}$ is a deterministic mapping as in (26).

A recent result combined the two, for the case of composite side information $s = ({}^{\text{c}}s, {}^{\text{nc}}s)$, where ${}^{\text{c}}s$ is known causally at the encoder and ${}^{\text{nc}}s$ — non-causally. The capacity, for this case, is equal to [3]

$${}^{\text{c-nc}}C = \max_{p(w|{}^{\text{nc}}s),h} I(W;Y) - I(W;Y)\,, \tag{28}$$

where w and h are as in (26) and (27).

Consider now the problem of joint source–channel coding, depicted in Figure 3, of conveying a block of length k of symbols generated by a memoryless source, U_1^k, with (source) side information Q_1^k available at the encoder over n channel uses of a state-dependent channel $p(y|x,s)$ with channel (state) side information S_1^n available at the decoder. The rate of the corresponding joint source–channel code denoted by r is therefore equal to $r = \lceil k/n \rceil$.

Merhav and Shamai [4], which proved that the separation principle between source and channel coding without side information at either (see, e.g., [7]) extends to the case of source coding with non-causal side information (4) and channel coding with non-causal side information, as well as to the case of source coding with non-causal side information and channel coding with causal side information. In the next theorem, we combine these two results by extending them to the case of source coding with composite side information available at the decoder and channel coding with composite side information at the encoder.

Theorem 2: Let U be a memoryless source with composite (source) side information $Q = ({}^{\text{c}}Q, {}^{\text{nc}}Q)$ at the decoder, as in Theorem 1, and state-dependent channel $p(y|x,s)$ with composite (channel state) side information at the decoder, as in (28). Then, distortion D w.r.t. a distortion measure ρ is achievable iff

$$r\ {}^{\text{c-nc}}R(D) \leq {}^{\text{c-nc}}C\,, \tag{29}$$

where ${}^{\text{c-nc}}R(D)$ and ${}^{\text{c-nc}}C$ are given in (10) and (28), respectively.

Proof:

Sufficiency: The achievable is a straightforward adaption of the achievable of the "ordinary" joint source–channel coding separation principle (see, e.g., [11, ch. 3.9]). Assume the desired distortion level D satisfies a strict inequality in condition (29). Then we can choose two constant rates R_S and R_C, representing the source and channel rates, respectively, s.t.

$$k\ {}^{\text{c-nc}}R(D) \overset{(a)}{<} kR_S = nR_C \overset{(b)}{<} n\ {}^{\text{c-nc}}C\,.$$

According to Theorem 1 and (a) one may compress the source into R_S bits per-symbol within average distortion D, for a large enough k. The resulting $kR_S = nR_C$ bits may then be conveyed reliably over the channel according to (28) and (b), for a large enough n. Thus, any distortion satisfying (29) is achievable.

Necessity: We follow the steps of [4] to establish the necessity of (29), by lower-bounding the expression $I(U_1^k; Y_1^n)$ by $k\ {}^{\text{c-nc}}R(D)$ and upper-bounding it by $n\ {}^{\text{c-nc}}C$.

Lower bound: The lower bound is proved similarly to the converse of Theorem 1. Specifically, by repeating (15)–(23) with J replaced by Y_1^n, the lower bound

$$I(U_1^k; Y_1^n) \geq k\ {}^{\text{c-nc}}R(D)\,, \tag{30}$$

is achieved.

Upper bound: The upper bound is proved in a similar manner to the converse of (28). Specifically, by repeating (b)–(g),(10) of [3], with the message W in [3] replaced by U_1^k (not to be confused with U of [3] which stands for the auxiliary variable there), the following upper bound is achieved:

$$I(U_1^k; Y_1^n) \leq n\ {}^{\text{c-nc}}C\,. \tag{31}$$

By combining the results of (30) and (31), the desired inequality holds. ■

Remark 3: Again, the result of Theorem 2 extends to infinite and continuous alphabets, as well as to the constrained channel input case, under mild regularity conditions, as is done in [8] (see also [12] and references therein).

V. Discussion

The repercussion of the present work is twofold: Establishing a general framework for the problem of sources with side information and transmission of such over state-dependent channels with state side information, where different parts of the side informations are 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 possible by observing the similarities between the converses of the "original" side information scenarios and combining them together. Finally, we note that for the general case, where the side information is composed of parts known with different finite look-ahead lengths, similar results to those of Weissman and El Gamal [2] for the single-part finite look-ahead, may be derived.

References

[1] A. D. Wyner and J. Ziv. The rate–distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory*, 22:1–10, Jan. 1976.

[2] T. Weissman and A. El Gamal. Source coding with limited-look-ahead side information at the decoder. *IEEE Trans. Info. Theory*, 52:5218–5239, Dec. 2006.

[3] A. Khina, M. Kesal, and U. Erez. State-dependent channels with composite state information at the encoder. In *Proceedings of IEEE Info. Theory Workshop (ITW), Paraty, Brazil*, pages 180–184, Oct. 2011.

[4] N. Merhav and S. Shamai (Shitz). On joint source–channel coding for the Wyner–Ziv source and the Gel'fand–Pinsker channel. *IEEE Trans. Info. Theory*, 49:2844–2855, Nov. 2003.

[5] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Info. Theory*, 19:471–480, July 1973.

[6] S. Shamai (Shitz), S. Verdú, and R. Zamir. Systematic lossy source/channel coding. *IEEE Trans. Info. Theory*, 44:564–579, March 1998.

[7] T. M. Cover and J. A. Thomas. *Elements of Information Theory, Second Edition*. Wiley, New York, 2006.

[8] A. D. Wyner. The rate–distortion function for source coding with side information at the decoder — II: General sources. *Information and Control*, 38:60–80, 1978.

[9] C. E. Shannon. Channels with side information at the transmitter. *IBM Journal of Research and Development*, 2:289–293, Oct. 1958.

[10] 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.

[11] A. El Gamal and T.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[12] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Info. Theory*, 48:1250–1276, June 2002.