Source Coding with in-Block Memory and Controllable Causal Side Information

Osvaldo Simeone
New Jersey Institute of Technology
Email: osvaldo.simeone@njit.edu

***Abstract*—The recently proposed set-up of source coding with a side information "vending machine" allows the decoder to select actions in order to control the quality of the side information. The actions can depend on the message received from the encoder and on the previously measured samples of the side information, and are cost constrained. Moreover, the final estimate of the source by the decoder is a function of the encoder's message and depends causally on the side information sequence. Previous work by Permuter and Weissman has characterized the rate-distortion-cost function in the special case in which the source and the "vending machine" are memoryless. In this work, motivated by the related channel coding model introduced by Kramer, the rate-distortion-cost function characterization is extended to a model with in-block memory. Various special cases are studied including block-feedforward and side information repeat request models.**

***Index Terms*: Source coding, block memory, side information "vending machine", feedforward, directed mutual information.**

I. Introduction and System Model

Consider the problem of source coding with controllable side information illustrated in Fig. 1. The encoder compresses a source $X^n = [X_1, ..., X_n]$ to a message W of R bits per source symbol. The decoder, based on the message W, takes actions A_i for all $i = 1, ..., n$, so as to control in a causal fashion the measured side information sequence Y^n. The action A_i is allowed to be a function of previously measured values Y^{i-1} of the side information, and the final estimate $\hat{X}_i$ is obtained by the decoder based on message W and as a causal function on the side information samples. The problem of characterizing the set of achievable tuples of rate R, average distortion D and average action cost Γ was solved in [1, Sec. II.E] under the assumptions of a memoryless source X^n and of a memoryless probabilistic model for the side information Y^n when conditioned on the source and the action sequences[1]. The distribution of the side information sequence given the source and action sequences is referred to as side information "vending machine" in [1].

In this work, we generalize the characterization of the rate-distortion-cost performance for the set-up in Fig. 1, from the memoryless scenario treated in [1], to a model in which source and side information "vending machine" have in-block memory (iBM). With iBM, the probabilistic models for source and "vending machine" have memory limited to blocks of size L samples, where L does not grow with the coding length n, as detailed below. The model under study is motivated by channel coding scenario put forth in [2] and can be considered to be the source coding counterpart of the latter.

[1]The mentioned characterization in [1, Sec. II.E] generalizes the result in [3, Sec. II] which is restricted to a model with action-independent side information.

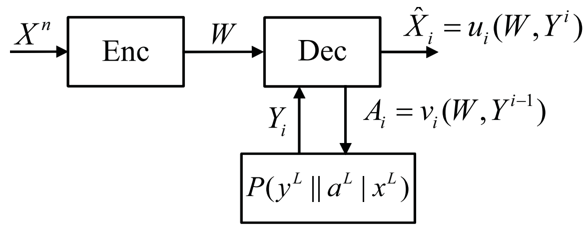

Fig. 1. Source coding with in-block memory (iBM) and causally controllable side information.

Notation: We write $[a, b] = [a, a+1, ..., b]$ for integers $b > a$; $[a, b] = a$ if $a = b$; and $[a, b]$ is empty otherwise. For a sequence of scalars $x_1, ..., x_n$, we write $x^n = [x_1, ..., x_n]$ and x^0 for the empty vector. The same notation is used for sequences of random variables $X^n = [X_1, ..., X_n]$, or sets $\mathcal{X}^n = [\mathcal{X}_1, ..., \mathcal{X}_n]$.

A. System Model

The system, illustrated in Fig. 1, is described by the following random variables.

- A **source** X^n with iBM of length L. The source X^n consists of m blocks
$$X_i^L = (X_{(i-1)L+1}, ..., X_{(i-1)L+L}) \tag{1}$$
with $i \in [1, m]$, each of L symbols, so that $n = mL$. The alphabet is possibly changing within each L-block, that is, we have $X_i \in \mathcal{X}_{t(i)+1}$, for L alphabets $\mathcal{X}_1, ..., \mathcal{X}_L$, where we have defined
$$t(i) = r(i-1, L), \tag{2}$$
with $r(x, y)$ being the remainder of x divided by y.
- A **message** $W \in [1, 2^{nR}]$ with R being the rate measured in bits per source symbol.
- An **action sequence** A^n with $A_i \in \mathcal{A}_{t(i)+1}$ for L alphabets $\mathcal{A}_1, ..., \mathcal{A}_L$.
- A **side information sequence** Y^n with $Y_i \in \mathcal{Y}_{t(i)+1}$ for L alphabets $\mathcal{Y}_1, ..., \mathcal{Y}_L$.
- A **source estimate** $\hat{X}^n$ with $\hat{X}_i \in \hat{\mathcal{X}}_{t(i)+1}$ for L alphabets $\hat{\mathcal{X}}_1, ..., \hat{\mathcal{X}}_L$.

978-1-4799-0446-4/13/$31.00 ©2013 IEEE

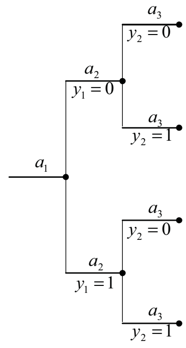

Fig. 2. An *action codetree* $\mathbf{v}^n(w, \cdot)$ for a given message $w \in [1, 2^{nR}]$ ($\mathcal{Y}_i = \{0, 1\}$, $n = 3$).

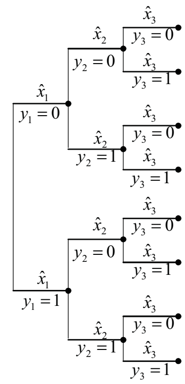

Fig. 3. A *decoder codetree* $\mathbf{u}^n(w, \cdot)$ for a given message $w \in [1, 2^{nR}]$ ($\mathcal{Y}_i = \{0, 1\}$, $n = 3$).

In order to simplify the notation, in the following, we will write $\mathcal{X}_i$ to denote $\mathcal{X}_{t(i)+1}$ also for $i > L$, and similarly for the alphabets $\mathcal{A}_i$, $\mathcal{Y}_i$ and $\hat{\mathcal{X}}_i$. The variables are related as follows.

- The *source* X^n has iBM of length L in the sense that it is characterized as
$$X_i = f_{t(i)+1}(Z_{\lceil i/L \rceil}), \qquad (3)$$
for some functions $f_i : \mathcal{Z} \to \mathcal{X}_i$, with $i \in [1, L]$, where Z_i, with $i \in [1, m]$, is a memoryless process with probability distribution $P(z)$. Note that (3) is equivalent to the condition that the distribution $P(x^n)$ factorizes as $\prod_{i=1}^m P(x_i^L)$.
- The *encoder* maps the source X^n into a message $W \in [1, 2^{nR}]$ according to some function $h : \mathcal{X}^n \to [1, 2^{nR}]$ as $W = h(X^n)$. To denote functional, rather than more general probabilistic, conditional dependence, we use the notation $1(W|X^n)$.
- The *decoder* observes the message W and takes *actions* A^n based also on the observation of the past samples of the side information sequence. Specifically, for each symbol $i \in [1, n]$ the action A_i is selected as
$$A_i = v_i(W, Y^{i-1}), \qquad (4)$$
for some functions $v_i : [1, 2^{nR}] \times \mathcal{Y}^{i-1} \to \mathcal{A}_i$. This conditional functional dependence is denoted as $1(a_i|\mathbf{v}^i, y^{i-1})$, where $\mathbf{v}^n = \mathbf{v}^n(w, \cdot)$ represents the *action codetree* (or action strategy) for a given message $w \in [1, 2^{nR}]$ in the time interval $i \in [1, n]$, that is, the collection of functions $v_i(w, \cdot)$ in (4) for all $i \in [1, n]$. A codetree $\mathbf{v}^n(w, \cdot)$ is illustrated in Fig. 2 for $\mathcal{Y}_i = \{0, 1\}$ and $n = 3$. Note that the subtrees $\mathbf{v}^i(w, \cdot)$ with any $i \in [1, n]$ can also be obtained from Fig. 2.
- The *side information* has iBM of length L in the sense that it is generated as a function of the previous actions taken *in the same block* and of the variable $Z_{\lceil i/L \rceil}$ (cf. (3)) as follows
$$Y_i = g_{t(i)+1}(A_{i-t(i)}, ..., A_i, Z_{\lceil i/L \rceil}), \qquad (5)$$
for some functions $g_i : \mathcal{A}^i \times \mathcal{Z} \to \mathcal{Y}_i$, with $i \in [1, L]$. Note that, as a special case, if the functions g_i do not depend on the actions, equations (3) and (5) imply that the sequences X^n and Y^n are L-block memoryless in the sense that their joint distribution factorizes as $\prod_{i=1}^m P(x_i^L, y_i^L)$.
- The *decoder*, based on the received message W along with the current and past samples of the side information sequence, produces the estimated sequence $\hat{X}^n$. Specifically, at each symbol $i \in [1, n]$, the estimate $\hat{X}_i$ is selected as
$$\hat{X}_i = u_i(W, Y^i) \qquad (6)$$
for some functions $u_i : [1, 2^{nR}] \times \mathcal{Y}^i \to \hat{\mathcal{X}}_i$. This conditional functional dependence is denoted as $1(\hat{x}_i|\mathbf{u}^i, y^i)$, where $\mathbf{u}^n(w, \cdot)$ represents the *decoder codetree* (or decoder strategy) for a given message $w \in [1, 2^{nR}]$ in the time interval $i \in [1, n]$, that is, the collection of functions $u_i(w, \cdot)$ in (6) for all $i \in [1, n]$. A codetree $\mathbf{u}^n(w, \cdot)$ (along with the subtrees $\mathbf{u}^i(w, \cdot)$ with $i \in [1, n]$) is illustrated in Fig. 3 for $\mathcal{Y}_i = \{0, 1\}$ and $n = 3$.

Overall, the probability distribution of the random variables $(X^n, \mathbf{V}^n, A^n, \mathbf{U}^n, Y^n, \hat{X}^n)$ factorizes as
$$\left[\prod_{i=1}^m P(x_i^L)\right] P(\mathbf{v}^n, \mathbf{u}^n|x^n) 1(a^n||\mathbf{v}^n, 0y^{n-1}) \qquad (7)$$
$$\cdot 1(\hat{x}^n||\mathbf{u}^n, y^n) \left[\prod_{i=1}^m P(y_i^L||a_i^L|x_i^L)\right],$$
where we have used the *directed conditioning* notation in [4]. Accordingly, we have defined
$$P(y^L||a^L|x^L) = \prod_{i=1}^L P(y_i|a^i, x^L) \qquad (8)$$
and similarly for the deterministic conditional relationships
$$1(a^n||\mathbf{v}^n, 0y^{n-1}) = \prod_{i=1}^n 1(a_i|\mathbf{v}^i, y^{i-1}) \qquad (9)$$

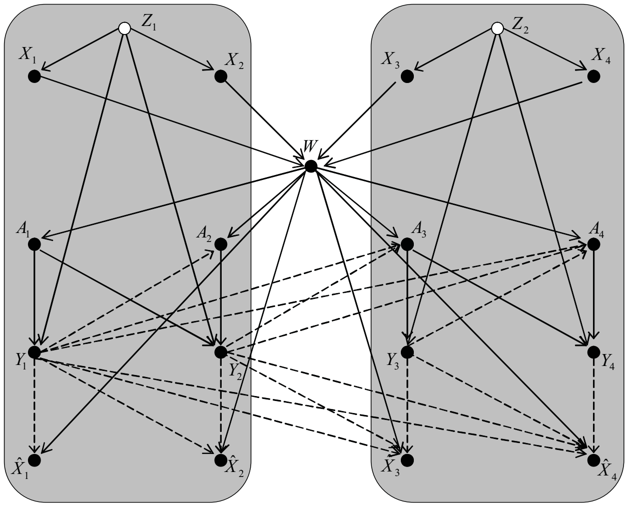

Fig. 4. FDG for a source coding problem with iBM of length $L = 2$ and $n = 4$ source symbols (and hence $m = 2$ blocks). The two blocks are shaded and the functional dependence on the side information is drawn with dashed lines.

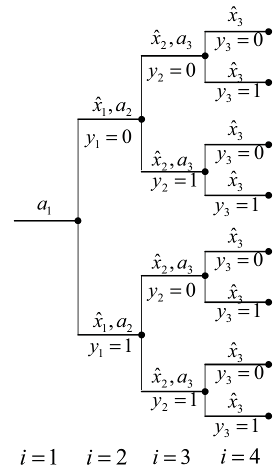

Fig. 5. A codetree $\mathbf{j}^{n+1}(w, \cdot)$ for a given message $w \in [1, 2^{nR}]$ ($\mathcal{Y}_i = \{0, 1\}$, $n = 3$).

and

$$1(\hat{x}^n||\mathbf{u}^n, y^n) = \prod_{i=1}^{n} 1(\hat{x}_i|\mathbf{u}^i, y^i). \tag{10}$$

A function dependence graph (FDG) (see, e.g., [4]) illustrating the joint distribution (7) for $L = 2$ and $n = 2$ (and thus $m = 2$) is shown in Fig. 4.

Remark 1. In (7), functions $1(a^n||\mathbf{v}^n, 0y^{n-1})$ and $1(\hat{x}^n||\mathbf{u}^n, y^n)$ are fixed as they represent the map from the branches of the codetrees $\mathbf{v}^n$ and $\mathbf{u}^n$ as indexed by the side information sequence to the action a_i and estimate $\hat{x}_i$ as illustrated in Fig. 2 and Fig. 3, respectively.

Fix a a non-negative and bounded function $d^L(x^L, \hat{x}^L)$ with domain $\mathcal{X}^L \times \hat{\mathcal{X}}^L$ to be the distortion metric and a non-negative and bounded function $\gamma^L(a^L, x^L)$ with domain $\mathcal{A}^L \times \hat{\mathcal{X}}^L$ to be the action cost metric. Under the selected metrics, a triple (R, D, Γ) is said to be achievable with distortion D and cost constraint Γ, if, for all sufficiently large m, there exist codetrees such that

$$\frac{1}{mL} \sum_{i=1}^{m} E[d^L(X_i^L, \hat{X}_i^L)] \leq D + \epsilon \tag{11}$$

and

$$\frac{1}{mL} \sum_{i=1}^{m} E[\gamma^L(A_i^L, X_i^L)] \leq \Gamma + \epsilon \tag{12}$$

for any $\epsilon > 0$. The rate-distortion-cost function $R(D, \Gamma)$ is the infimum of all achievable rates with distortion D and cost constraint Γ.

Remark 2. The system model under study reduces to that investigated in [1, Sec. II.E] for the special case with memoryless sources, i.e., with $L = 1$.

II. Main Results

In this section, the rate-distortion-cost function $R(D, \Gamma)$ is derived and some of its properties are discussed. The next section illustrates various special cases and connections to previous works.

A. Equivalent Formulation

We start by showing that the problem can be formulated in terms of a single codetree. This contrasts with the more natural definitions given in the previous section, in which two separate codetrees, namely $\mathbf{v}^n(w, \cdot)$ and $\mathbf{u}^n(w, \cdot)$, were used (see Fig. 2 and Fig. 3). Towards this end, we define a "joint" codetree $\mathbf{j}^{n+1}(w, \cdot) = (\mathbf{j}^1(w, \cdot), ..., \mathbf{j}^{n+1}(w, \cdot))$ that satisfies the functional dependencies

$$1(a_i|\mathbf{j}^i, y^{i-1}) = 1(a_i|\mathbf{v}^i, y^{i-1}), \tag{13}$$

and

$$1(\hat{x}_i|\mathbf{j}^{i+1}, y^i) = 1(\hat{x}_i|\mathbf{u}^i, y^i) \tag{14}$$

for all $i \in [1, n]$. The codetree $\mathbf{j}^{n+1}(w, \cdot)$ is illustrated in Fig. 5 for $n = 3$. Note that the subtree $\mathbf{j}^1(w, \cdot)$ only specifies the action a_1 to be taken at time $i = 1$, while the the leaves of the tree $\mathbf{j}^{n+1}(w, \cdot)$ are indexed solely by the estimated value $\hat{x}_n$.

B. Rate-Distortion-Cost Function

Using the representation in terms of a single codetree given above, we now provide a characterization of the rate-distortion-cost function.

Proposition 1. *The rate-distortion-cost function is given by*

$$R(D, \Gamma) = \frac{1}{L} \min I(X^L; \mathbf{J}^{L+1}) \tag{15}$$

where the joint distribution of the variables X^L,Y^L,A^L,$\hat{X}^L$ and of the codetree $\mathbf{J}^{L+1}$ factorizes as

$$\begin{aligned} P(x^L)P(\mathbf{j}^{L+1}|x^L)1(a^L||\mathbf{j}^L, 0y^{L-1}) \\ \cdot 1(\hat{x}^L||\mathbf{j}_2^{L+1}, y^L)P(y^L||a^L|x^L), \end{aligned} \tag{16}$$

and the minimization is performed over the conditional distribution $P(\mathbf{j}^{L+1}|x^L)$ of the codetree under the constraints

$$\frac{1}{L}E[d^L(X^L,\hat{X}^L)] \leq D \tag{17}$$

and

$$\frac{1}{L}E[\gamma^L(A^L,X^L)] \leq \Gamma. \tag{18}$$

The number of codetrees $\mathbf{J}^{L+1}$ can be limited as $|\mathcal{J}^{L+1}| \leq |\mathcal{X}^L|+3$ without loss of optimality.

Proof: The achievability of Proposition 1 follows from classical random coding arguments. Specifically, the encoder draws the codetrees $\mathbf{j}^{n+1}(w,\cdot)$ for all $w \in [1, 2^{n(R(D)+\delta)}]$ with some $\delta > 0$, as follows. First, for each $w \in [1, 2^{n(R(D)+\delta)}]$ a concatenation of m codetrees $\mathbf{j}_i^{L+1}(w,\cdot)$ of length $L+1$, with $i \in [1,m]$, is generated, such that the constituent codetrees $\mathbf{j}_i^{L+1}(w,\cdot)$ are i.i.d. and distributed with probability $P(\mathbf{j}^{L+1})$. The codetree $\mathbf{j}^{n+1}(w,\cdot)$ is then obtained by combining the leaves and the root of successive constituent codetrees: the leaves of the past codetree specify the estimates for the previous time instant, while the root of the next codetree specify the action for the current time instant. Encoding is performed by looking for a message $w \in [1, 2^{n(R(D)+\delta)}]$ such that the corresponding pair $(x^n, \mathbf{j}^{n+1}(w,\cdot))$ is (strongly) jointly typical with respect to the joint distribution $P(x^L)P(\mathbf{j}^{L+1}|x^L)$, when the sequences $(x^n, \mathbf{j}^{n+1}(w,\cdot))$ are seen as the memoryless m-sequences $(x_1^L, \mathbf{j}_1^{L+1}(w,\cdot))$,...,$(x_m^L, \mathbf{j}_m^{L+1}(w,\cdot))$. By the covering lemma [5, Lemma 3.3], rate $1/L \cdot I(X^L;\mathbf{J}^{L+1})$ suffices to guarantee the reliability of this step. Moreover, if the distribution $P(\mathbf{j}^{L+1}|x^L)$ is selected so as to satisfy (17) and (18), then, by the typical average lemma [5], the constraints (11) and (12) are also guaranteed to be met for sufficiently large n. Further details and the proof of the converse can be found in [6]. ■

Remark 3. The rate-distortion-cost function can also be expressed in terms of two separate codetrees using the definitions given in Sec. I-A. Specifically, following similar steps as in the proof of Proposition 1, the rate-distortion-cost function can be expressed as the minimization

$$R(D,\Gamma) = \frac{1}{L}\min I(X^L;\mathbf{V}^L,\mathbf{U}^L) \tag{19}$$

where the joint distribution of the variables X^L,Y^L,A^L,$\hat{X}^L$ and of the codetrees $\mathbf{V}^L$ and $\mathbf{U}^L$ factorizes as

$$\begin{aligned} &P(x^L)P(\mathbf{v}^L,\mathbf{u}^L|x^L)1(a^L||\mathbf{v}^L,0y^{L-1}) \\ &\cdot 1(\hat{x}^L||\mathbf{u}^L,y^L)P(y^L||a^L|x^L), \end{aligned} \tag{20}$$

and the minimization is performed over the conditional distribution $P(\mathbf{v}^L,\mathbf{u}^L|x^L)$ of the codetrees under the constraints (17) and (18).

Remark 4. The rate-distortion-cost function in Proposition 1 does not include auxiliary random variables, since the codetree $\mathbf{J}^{L+1}$ is part of the problem specification. This is unlike the characterization given in [1] for the memoryless case. Moreover, problem (15) is convex in the unknown $P(\mathbf{j}^{L+1}|x^L)$ and hence can be solved using standard algorithms. It is also noted that, extending [7], one may devise a Blahut-Arimoto-type algorithm for the calculation of the rate-distortion-cost function. This aspect is not further investigated here.

Remark 5. The achievable scheme used to prove Proposition 1 adapts the actions only to the side information samples corresponding to the same L-block. More precisely, the action A_i depends, through the selected codetree, only on the side information samples $Y_{i-t(i)},...,Y_{i-1}$. Since the problem definition allows, via (4), for actions that depend on all past side information samples, namely Y^{i-1}, this result demonstrates that adapting the actions across the blocks cannot improve the rate-distortion-cost function. This is consistent with the finding in [1], where it is shown that adaptive actions do not improve the rate-distortion performance for a memoryless model, i.e., with $L=1$. Similarly, one can conclude from Proposition 1 that, while adapting the estimate $\hat{X}_i$ to the side information samples within the same L-block, namely $Y_{i-t(i)},...,Y_i$, is generally advantageous, adaptation across the blocks is not. This extends the results in [3], in which it is shown that, for $L=1$, the estimate can depend only on the current value of the side information without loss of optimality.

III. Special Cases and Examples

In this section, we detail some further consequences of Proposition 1 and connections with previous work.

A. Memoryless Source ($L = 1$)

As mentioned in Remark 2, if $L = 1$, the model at hand reduces to the standard one with memoryless sources, in which the joint distribution of X^n and Y^n factorizes as $\prod_{i=1}^n P(x_i,y_i)$. This model was studied in [1], where the rate-distortion-cost function was derived. The result in [1, Sec. II-E] can be seen to be a special case of Proposition 1.

B. Action-Independent Side Information

Here we consider the case in which the side information is action independent, that is, we have $P(y^L||a^L|x^L) = P(y^L|x^L)$. Under this assumption, the action sequence does not need to be included in the model, and, from (19), the rate-distortion function is given by

$$R(D) = \frac{1}{L}\min I(X^L;\mathbf{U}^L), \tag{21}$$

where the joint distribution of the variables X^L,Y^L,$\hat{X}^L$ and of the codetree $\mathbf{U}^L$ factorizes as

$$P(x^L)P(\mathbf{u}^L|x^L)1(\hat{x}^L||\mathbf{u}^L,y^L)P(y^L|x^L), \tag{22}$$

and the minimization is performed over the conditional distribution $P(\mathbf{u}^L|x^L)$ of the codetrees under the constraint (17). Note that, given the absence of actions, we have used the formulation in terms of individual codetrees discussed in Remark 3 in order to simplify the notation. For $L=1$, the characterization (21) reduces to the one derived in [3, Sec. II].

C. Block-Feedforward Model

As a specific instance of the setting with action-independent side information, we consider here the *block-feedforward* model in which we have $Y_i = X_{i-1}$ for all i not multiple of L and Y_i equal to a fixed symbol in $\mathcal{Y}_i$ otherwise. This model is related to the feedforward set-up studied in [8] with the difference that here feedforward is limited to within the L-blocks. In other words, the side information is $Y_i = X_{i-1}$ only if X_{i-1} is in the same L-block as Y_i and is not informative otherwise. We now show that, similar to [8], the rate-distortion function with block-feedforward can be expressed in terms of directed information and does not entail an optimization over the codetrees.

Corollary 1. For the block-feedforward model, the rate-distortion function is given by

$$R(D) = \frac{1}{L} \min I(\hat{X}^L \to X^L) \tag{23}$$

where the joint distribution of the variables X^L, Y^L and $\hat{X}^L$ factorizes as

$$P(x^L)P(\hat{x}^L|x^L)P(y^L|x^L), \tag{24}$$

and the minimization is performed over the conditional distribution $P(\hat{x}^L|x^L)$ *under the constraint (17).*

Remark 6. In the feedforward model studied in [8], feedforward of the source X^n is not restricted to take place only within the L-blocks, namely we have $Y_i = X^{i-1}$ for all $i \in [1, n]$. As a result, the rate-distortion function is proved in [8] to be given by the limit of (23) over L.

Proof: The achievability is obtained by using concatenated codetrees of length L similar to Proposition 1. However, unlike Proposition 1, the codetrees are generated according to the distribution $p(\hat{x}^L||0x^{L-1})$ as done in [8]. For the converse, we refer to [6]. ■

D. Side Information Repeat Request

Consider the situation in which the decoder at any time i, upon the observation of the side information Y_i, can decide whether to take a second measurement of the side information, thus paying the associated cost, or not. To elaborate, assume a memoryless source X^n with distribution $P(x)$. At any time i, the first observation Y_{i1} of the side information is distributed according to the memoryless channel $P(y_1|x)$ when the input is $X_i = x$, while the second observation Y_{i2} depends on the action $A_i = a$ via the memoryless channel $P(y_2|x, a)$ with input $X_i = x$.

This scenario can be easily seen to be a special case of the model under study with iBM of size $L = 2$. The corresponding FDG is illustrated in Fig. 6. By comparing this FDG with the general FDG in Fig. 4, it is seen that the model under study in this section can be obtained from the one presented in Sec. I-A by appropriately setting the alphabets of given subset of variables to empty sets and by relabeling. Further discussion can be found in [6].

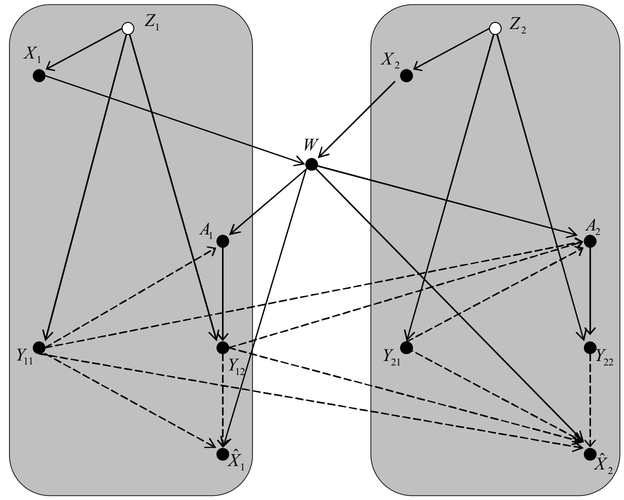

Fig. 6. FDG for the model with side information repeat request.

IV. Concluding Remarks

Models with in-block memory (iBM), first proposed in the context of channel coding problems in [2] and here for source coding, provide tractable extensions of standard memoryless models. Specifically, in this paper, we have presented results for a point-to-point system with controllable side information at the receiver and iBM. Interesting generalizations include the investigation of multi-terminal models.

Acknowledgment

The author would like to thank Gerhard Kramer for the very useful comments and suggestions.

References

[1] H. Permuter and T. Weissman, "Source coding with a side information "Vending Machine"," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4530-4544, Jul. 2011.
[2] G. Kramer, "Information networks in-block memory," in *Proc. IEEE Information Theory Workshop* (ITW 2012), Lausanne, Switzerland, Sep. 2012.
[3] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5218-5239, Dec. 2006.
[4] G. Kramer, *Directed Information for Channels with Feedback*, volume ETH Series in Information Processing. Vol. 11, Konstanz, Germany: Hartung-Gorre Verlag, 1998.
[5] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, 2012.
[6] O. Simeone, "Source coding with in-block memory and causally controllable side information," arXiv:1301.0926.
[7] F. Dupuis, W. Yu, and F. M. J. Willems, "Blahut-Arimoto algorithms for computing channel capacity and ratedistortion with side information," in *Proc. IEEE Int. Sym. Inform. Theory* (ISIT 2004), Chicago, IL, Jun./Jul. 2004.
[8] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2154-2179, Jun. 2007.