Information Embedding on Actions

Behzad Ahmadi
CWCSPR, ECE Dept.
NJIT

Himanshu Asnani
ISL, EE Dept.
Stanford University

Osvaldo Simeone
CWCSPR, ECE Dept.
NJIT

Haim H. Permuter
ECE Dept.
Ben-Gurion University of the Negev

***Abstract*—The problem of optimal actuation for channel and source coding was recently formulated and solved in a number of relevant scenarios. In this class of models, actions are taken either to acquire side information in an efficient way for source coding, or to control or probe effectively the channel state for channel coding. In this paper, the problem of embedding information on the actions is introduced by considering the presence of an additional decoder that observes only a function of the actions. For the source coding model, this decoder wishes to reconstruct a lossy version of the source being transmitted over the point-to-point link, while, for the channel coding problem, the decoder wishes to retrieve a portion of the message conveyed over the link. In both cases, single-letter performance characterizations are provided for various special cases, along with specific examples.**

***Index Terms*—Action-dependent channel coding, cribbing, side information vending machine.**

I. Introduction

The recent works [1], [2] study the problem of *optimal actuation for source and channel coding* in resource-constrained systems. Specifically, in [1], an extension of the Wyner-Ziv source coding problem is considered in which the decoder or the encoder can take actions that affect the quality of the side information available at the decoder's side. In [2][3], a related channel coding problem is studied in which the encoder in a point-to-point channel can take actions to affect, or probe, the state of a channel. Generalizations of these works include models with additional design constraints [4], [5], with adaptive actions [6], with memory [7], [5] and with multiple terminals [8][9][10].

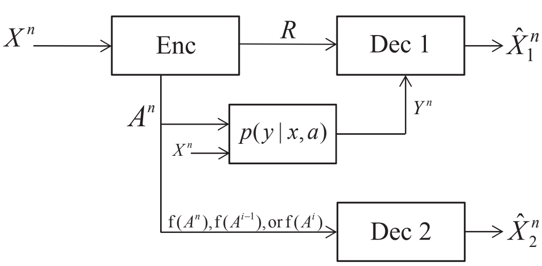

Figure 1. Source coding with encoder-side actions for information acquisition and with information embedding on actions.

To illustrate the problem of interest in this paper, consider the example of a link in which the encoder takes actions to probe the channel by sending a training packet. Assume that the encoder is interested in piggybacking some information in the training packet. The main question addressed here is: *How much information can be embedded in the actions without affecting the performance of the link*? Or, to turn the question around, *what is the performance loss for the link as a function of the amount of information that is encoded in the actions*?

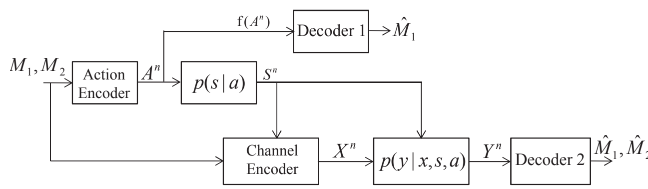

Figure 2. Channel coding with actions for channel state control and with information embedding on actions.

The idea of embedding information in the actions is related to the classical problem of information hiding (see, e.g., [11] and references therein). In information hiding, a message is embedded in a host data under distortion constraints. The message is then retrieved by a decoder that observes the host signal through a noisy channel. Note that the (host) signal onto which the message is embedded is a given process. Instead, in the set-up of information embedding on actions considered here, the (action) signal on which information is embedded is designed to optimize the given communication task.

A. Contributions and Paper Organization

The main contributions of this paper are as follows.

- **Encoder-side actions for side information acquisition:** We first consider the source coding set-up in Fig. 1, in which a decoder (Decoder 2), observing a function of the action sequence A^n, is added to the problem of source coding with actions taken at the encoder formulated in [1]. Note that the latter includes only Encoder and Decoder 1. Sec. II derives the achievable rate-distortion-cost region in the special case in which the channel $p_{Y|X,A}(y|x,a)$ with source and action (X, A) as inputs and side information Y as output is such that Y is a deterministic function of A;
- **Actions for channel control and probing:** Then, we study the set-up in Fig. 2, which generalizes the model in [2] by adding a decoder (Decoder 1), observing a function of the action sequence A^n taken by the encoder. A single-letter characterization of the achievable capacity-cost region is obtained in Sec. III. Finally, the special

978-1-4799-0446-4/13/$31.00

case of actions for channel probing [3] is elaborated on with an example in Sec. III-C.

II. Encoder-Side Actions for Side Information Acquisition

In this section, following [1, Sec. III], we consider scenarios in which the encoder takes the actions affecting the side information of Decoder 1, as shown in Fig. 1.

A. System Model

The model is defined by the probability mass functions (pmfs) $p_X(x)$ and $p_{Y|AX}(y|a,x)$, by the function f: $\mathcal{A} \to \mathcal{B}$, and by discrete alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2$, as follows. The source sequence X^n is such that $X_i \in \mathcal{X}$ for $i \in [1,n]$, where $A_i \in \mathcal{A}$ for $i \in [1,n]$, is independent and identically distributed (i.i.d.) with pmf $p_X(x)$. The Encoder measures sequence X^n and encodes it in a message M of nR bits, which is delivered to Decoder 1. Moreover, based on the measured sequence X^n, the encoder also selects an action sequence A^n.

The action sequence affects the quality of the measurement Y^n of sequence X^n obtained at the Decoder 1. Specifically, given $A^n = a^n$ and $X^n = x^n$, the sequence Y^n is distributed as $p(y^n|a^n,x^n) = \prod_{i=1}^n p_{Y|A,X}(y_i|a_i,x_i)$. The cost of the action sequence is defined by a cost function Λ: $\mathcal{A} \to [0, \Lambda_{\max}]$ with $0 \leq \Lambda_{\max} < \infty$, as $\Lambda(a^n) = \sum_{i=1}^n \Lambda(a_i)$. The estimated sequence $\hat{X}_1^n \in \hat{\mathcal{X}}_1^n$ is then obtained as a function of M and Y^n.

Decoder 2 observes a function of the action sequence A^n, thus obtaining $\mathrm{f}(A^n) = (\mathrm{f}(A_1), ..., \mathrm{f}(A_n)) \in \mathcal{B}^n$. Based on $\mathrm{f}(A^n)$, Decoder 2 obtains an estimate $\hat{X}_2^n \in \hat{\mathcal{X}}_2^n$ within given distortion requirements. The estimated sequences $\hat{X}_j^n$ for $j = 1,2$ must satisfy distortion constraints defined by functions $d_j(x,\hat{x}_j)$: $\mathcal{X} \times \hat{\mathcal{X}}_j \to [0, D_{j,\max}]$ with $0 \leq D_{j,\max} < \infty$ for $j = 1,2$, respectively. A formal description of the operations at encoder and decoder follows.

Definition 1. An (n,R,D_1,D_2,Γ) code for the set-up of Fig. 1 consists of a source encoder

$$\mathrm{h}^{(e)}\colon \mathcal{X}^n \to [1, 2^{nR}], \tag{1}$$

which maps the sequence X^n into a message M at the Encoder; an "action" function

$$\mathrm{h}^{(a)}\colon \mathcal{X}^n \to \mathcal{A}^n, \tag{2}$$

which maps the sequence X^n into an action sequence A^n at the Encoder; two decoders, namely

$$\mathrm{h}_1^{(d)}\colon [1, 2^{nR}] \times \mathcal{Y}^n \to \hat{\mathcal{X}}_1^n, \tag{3}$$

which maps the message M and the measured sequence Y^n into the estimated sequence $\hat{X}_1^n$ at Decoder 1; and

$$\mathrm{h}_2^{(d)}\colon \mathcal{B}^n \to \hat{\mathcal{X}}_2^n, \tag{4}$$

which maps the observed sequence $\mathrm{f}(A^n)$ into the the estimated sequence $\hat{X}_2^n$ at Decoder 2; such that the action cost constraint Γ and distortion constraints D_j for $j = 1,2$ are satisfied, i.e.,

$$\frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[\Lambda(A_i)\right] \leq \Gamma \tag{5}$$

$$\text{and } \frac{1}{n}\sum_{i=1}^n \mathrm{E}\left[d_j(X_i, \hat{X}_{ji})\right] \leq D_j \text{ for } j = 1,2. \tag{6}$$

Given a distortion-cost tuple (D_1, D_2, Γ), a rate R is said to be achievable if, for any $\epsilon > 0$, and sufficiently large n, there exists a $(n, R, D_1+\epsilon, D_2+\epsilon, \Gamma+\epsilon)$ code. The *rate-distortion-cost function* $R(D_1,D_2,\Gamma)$ is defined as $R(D_1,D_2,\Gamma) = \inf\{R : \text{the tuple } (R,D_1,D_2,\Gamma) \text{ is achievable}\}$. In the rest of this section, for simplicity of notation, we drop the subscripts from the definition of the pmfs, thus identifying a pmf by its argument.

B. Rate-Distortion-Cost Function

As discussed in [1], even in the absence of Decoder 2, the problem at hand, in which the action sequence is taken by the encoder, is challenging. Here we observe that the problem involves source-channel coding since the action channel $p(y|x,a)$ provides a communication link to Decoder 1. To facilitate the analysis, we thus focus first on the special case in which the side information channel $p(y|x,a)$ is such that Y is a deterministic function of A, i.e., $Y = \mathrm{f}_Y(A)$, and $\mathrm{f}(A) = A$ in Proposition 1. The result is then generalized to all deterministic functions f for which $H(\mathrm{f}_Y(A)|\mathrm{f}(A)) = 0$. Proposition 2 finally considers the case $H(\mathrm{f}(Y)|\mathrm{f}_Y(A)) = 0$.

Proposition 1. *The rate-distortion-cost function $R(D_1,D_2,\Gamma)$ with $\mathrm{f}(A) = A$ and $Y = \mathrm{f}_Y(A)$ is given by*

$$R(D_1,D_2,\Gamma) = \min_{p(u|x),\, p(\hat{x}_1|u,x),\, p(\hat{x}_2|u,x),\, p(a)} \{I(X;\hat{X}_1,U) - H(\mathrm{f}_Y(A))\}^+, \tag{7}$$

where the information measures are evaluated with respect to the joint pmf

$$p(x,u,\hat{x}_1,\hat{x}_2,a) = p(x)p(u|x)p(\hat{x}_1|u,x)p(\hat{x}_2|u,x)p(a), \tag{8}$$

for some pmfs $p(u|x)$, $p(\hat{x}_1|u,x)$, $p(\hat{x}_2|u,x)$ and $p(a)$ such that the inequalities

$$\mathrm{E}[d_j(X,\hat{X}_j)] \leq D_j, \text{ for } j = 1,2, \tag{9a}$$
$$\mathrm{E}[\Lambda(A)] \leq \Gamma, \tag{9b}$$
$$I(X;U) \leq H(\mathrm{f}_Y(A)) \tag{9c}$$
$$\text{and } I(X;\hat{X}_2|U) \leq H(A|\mathrm{f}_Y(A)) \tag{9d}$$

are satisfied. Finally, U is an auxiliary random variable whose alphabet cardinality can be constrained as $|\mathcal{U}| \leq |\mathcal{X}||\hat{\mathcal{X}}_1||\hat{\mathcal{X}}_2| + 3$ without loss of optimality.

Remark 1. The results above generalize a number of known single-letter characterizations. Notably, if $D_2 = D_{2,\max}$, so that the distortion requirements of Decoder 2 are immaterial to the system performance, the result reduces to [1, Theorem 7]. Moreover, in the special case in which $A = (A_0, A_2)$, $Y = A_0$, $R = R_1$, $|\mathcal{A}_0| = 2^{R_0}$, $|\mathcal{A}_2| = 2^{R_2}$, the model coincides with the lossy Gray-Wyner problem [12][1].

Proposition 1 establishes the optimality of separate source-channel coding for the set-up in Fig. 1 under the stated conditions. In particular, in the optimal strategy, the encoder compresses using a standard successive refinement source code in which U represents the coarse description and $\hat{X}_1, \hat{X}_2$ two independent refinements. The indices of the coarse description U and of the refined description $\hat{X}_2$ are sent on the degraded deterministic broadcast channel with input A and outputs $(A, \mathrm{t}(A))$ using superposition coding. Reliable compression and communication is guaranteed by the two bounds (9c)-(9d). A further refined description $\hat{X}_1$ is produced for Decoder 1, and the corresponding index is sent partly over the mentioned broadcast channel and partly over the link of rate R, leading to the rate (7). Details of the achievability proof and the proof of the converse are given in [13, Appendix C].

Remark 2. Following the discussion above, specializing Proposition 1 to the case $R = 0$ shows the optimality of source-channel coding separation for the lossy transmission of a source over a deterministic degraded broadcast channel (see [14, Chapter 14] for a review of scenarios in which the optimality of separation holds for lossless transmission over a broadcast channel).

The scenario solved above is when the action observation is perfect, i.e., $\mathrm{f}(A) = A$. The result also carries verbatim for the more general case where $\mathrm{f}(A)$ is a generic function as long as $H(\mathrm{f}_Y(A)|\mathrm{f}(A)) = 0$. The expressions of the rate region remain the same as in the proposition above except that A is replaced by $\mathrm{f}(A)$.

As mentioned, Proposition 1 characterizes the optimal performance for the case when Decoder 2 has a better information about the actions taken by the encoder than Decoder 1 in the sense that $H(\mathrm{f}_Y(A)|\mathrm{f}(A)) = 0$. We note here that a similar characterization can be given also for the dual setting in which $H(\mathrm{f}(A)|\mathrm{f}_Y(A)) = 0$ so that Decoder 1 has the better observation about the actions.

Proposition 2. *The rate-distortion-cost function $R(D_1, D_2, \Gamma)$ with $H(\mathrm{f}(A)|\mathrm{f}_Y(A)) = 0$ is given by*

$$R(D_1, D_2, \Gamma) = \min_{p(a),\, p(\hat{x}_1, \hat{x}_2|x)} \{I(X; \hat{X}_1, \hat{X}_2) - H(\mathrm{f}_Y(A))\}^+, \tag{10}$$

where the information measures are evaluated with respect to the joint pmf

$$p(x, \hat{x}_1, \hat{x}_2, a) = p(x)p(\hat{x}_1, \hat{x}_2|x)p(a), \tag{11}$$

such that the following inequalities are satisfied,

$$\mathrm{E}[d_j(X, \hat{X}_j)] \leq D_j, \text{ for } j = 1, 2, \tag{12a}$$

$$\mathrm{E}[\Lambda(A)] \leq \Gamma, \tag{12b}$$

$$I(X; \hat{X}_2) \leq H(\mathrm{f}(A)). \tag{12c}$$

[1]Note that here 2^{R_0} and 2^{R_2} are constrained to be integers.

Referring to [13] for details, here we just outline the achievability. A successive refinement codebook is generated with coarse source description $\hat{X}_2$ and refinement $\hat{X}_1$. The index of the coarse description and part of the index of the refined description are sent through the degraded broadcast channel with input A and outputs $(\mathrm{f}_Y(A), \mathrm{f}(A))$. The remaining part of the rate of the index for $\hat{X}_1$ is conveyed to Decoder 1 on the link of rate R. As a result, reliability of compression and communication over the "action" broadcast channel is guaranteed if the inequalities

$$I(X; \hat{X}_2) \leq H(\mathrm{f}(A))$$
$$I(X;\hat{X}_2) + I(X; \hat{X}_1|\hat{X}_2) - R \leq H(\mathrm{f}(A), \mathrm{f}_Y(A)) = H(\mathrm{f}_Y(A))$$

are satisfied, where the latter inequality implies $R \geq I(X; \hat{X}_1, \hat{X}_2) - H(\mathrm{f}_Y(A))$.

III. Actions for Channel State Control and Probing

In this section, we consider the impact of information embedding on actions for the set-up of channel coding with actions of [2]. To this end, we consider the model in Fig. 2, in which Decoder 1, based on the observation of a deterministic function of the actions, wishes to retrieve part of the information destined to Decoder 2. Note that for simplicity of notation here the additional encoder that observes the actions is denoted as Decoder 1, rather than Decoder 2 as done above.

A. System Model

The system is defined by the pmfs $p(x)$, $p(y|x, s, a)$, $p(s|a)$, function f: $\mathcal{A} \to \mathcal{B}$ and by discrete alphabets $\mathcal{X}, \mathcal{A}, \mathcal{B}, \mathcal{S}$, and $\mathcal{Y}$. Given the messages (M_1, M_2), selected uniformly from the set $\mathcal{M}_1 \times \mathcal{M}_2 = [1, 2^{nR_1}] \times [1, 2^{nR_2}]$, an action sequence $A^n \in \mathcal{A}^n$ is selected by the Encoder. Decoder 1 observes the signal $B^n = \mathrm{f}(A^n)$ as a deterministic function of the actions, and estimates message M_1. Moreover, the state sequence $S^n \in \mathcal{S}^n$ is generated as the output of a memoryless channel $p(s|a)$ and we have $p(b^n, s^n|a^n) = \prod_{i=1}^n p(s_i|a_i)\mathbf{1}_{\{b_i = \mathrm{f}(a_i)\}}$ for an action sequence $A^n = a^n$. The input sequence $X^n \in \mathcal{X}^n$ is selected on the basis of both messages (M_1, M_2) and of the state sequence S^n by the Encoder. The action sequence A^n and the input X^n have to satisfy an average cost constraint defined by a function $\gamma : \mathcal{A} \times \mathcal{X} \to [0, \infty)$, so that the cost for the input sequences a^n and x^n is given by $\gamma(a^n, x^n) = \frac{1}{n}\sum_{i=1}^n \gamma(a_i, x_i)$. Given $X^n = x^n$, $S^n = s^n$ and $A^n = a^n$, the received signal is distributed as $p(y^n|x^n, s^n, a^n) = \prod_{i=1}^n p(y_i|x_i, s_i, a_i)$. Decoder 2, having received the signal Y^n, estimates both messages (M_1, M_2).

The setting includes the semi-deterministic broadcast channel with degraded message sets [15] (see also [14, Ch. 8]) as a special case by setting X to be constant and $Y = S$, and the channel with action-dependent states studied in [2] for $R_1 = 0$.

Definition 2. An $(n, R_1, R_2, \Gamma, \epsilon)$ code for the model in Fig. 2 consists of an action encoder

$$\mathrm{h}^{(a)}\colon \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{A}^n, \tag{13}$$

which maps message (M_1, M_2) into an action sequence A^n; a channel encoder

$$\mathrm{h}^{(e)}\colon \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{S}^n \to \mathcal{X}^n, \tag{14}$$

which maps message (M_1, M_2) and the state sequence S^n into the sequence X^n; two decoding functions

$$\mathrm{h}_1^{(d)}\colon \mathcal{B}^n \to \mathcal{M}_1, \tag{15}$$

$$\text{and } \mathrm{h}_2^{(d)}\colon \mathcal{Y}^n \to \mathcal{M}_1 \times \mathcal{M}_2, \tag{16}$$

which map the sequences B^n and Y^n into the estimated messages $\hat{M}_1$ and $(\hat{M}_1, \hat{M}_2)$, respectively; such that the probability of error in decoding the messages (M_1, M_2) is small,

$$\Pr[\mathrm{h}_1^{(d)}(B^n) \neq M_1] \leq \epsilon, \tag{17}$$

$$\text{and } \Pr[\mathrm{h}_2^{(d)}(Y^n) \neq (M_1, M_2)] \leq \epsilon, \tag{18}$$

and the cost constraint is satisfied, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\gamma(A_i, X_i)\right] \quad \leq \quad \Gamma + \epsilon. \tag{19}$$

Given a cost Γ, a rate pair (R_1, R_2) is said to be achievable for a cost-constraint Γ if, for any $\epsilon > 0$ and sufficiently large n, there a exists a $(n, R_1, R_2, \Gamma, \epsilon)$ code. We are interested in characterizing the capacity-cost region $\mathcal{C}(\Gamma)$, which is the closure of all achievable rate pairs (R_1, R_2) for the given cost Γ.

B. Capacity-Cost Region

In this section, a single-letter characterization of the capacity-cost region is derived.

Proposition 3. *The capacity-cost region $\mathcal{C}(\Gamma)$ is given by the union of all rate pairs (R_1, R_2) such that the inequalities*

$$R_1 \quad \leq \quad H(\mathrm{f}(A)) \tag{20a}$$

$$\text{and } R_1 + R_2 \quad \leq \quad I(A, U; Y) - I(U; S|A), \tag{20b}$$

are satisfied, where the mutual informations are evaluated with respect to the joint pmf

$$p(a, s, u, x, y) = p(a)p(s|a)p(u|s, a)\mathbf{1}_{\{x=\mathrm{g}(u,s)\}}p(y|x, s, a), \tag{21}$$

for some pmfs $p(a)$, $p(u|s, a)$ and function $\mathrm{g}\colon \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that

$$\mathrm{E}[\gamma(A, X)] \quad \leq \quad \Gamma. \tag{22}$$

Finally, we can set $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{S}||\mathcal{A}| + 1$ without loss of optimality.

The proof of converse is an immediate consequence of cut-set arguments and of the proof of the upper bound obtained in [2, Theorem 1]. Specifically, inequality (20a) follows by considering the cut around Decoder 1, while the inequality (20b) coincides with the bound derived in [2, Theorem 1] on the rate that can be communicated between the Encoder and Decoder 2 with no regards for Decoder 1[2]. For achievability, the rate region

$$R_1 \quad \leq \quad H(\mathrm{f}(A)) \tag{23a}$$

$$R_1 + R_2 \quad \leq \quad I(A; Y) + I(U; Y|A) - I(U; S|A), \tag{23b}$$

$$\begin{aligned} R_2 \quad \leq \quad & I(A; Y|\mathrm{f}(A)) + I(U; Y|A) \\ & -I(U; S|A) \end{aligned} \tag{23c}$$

can be seen to be equivalent to (20) since, by the problem definition, the rate R_1 can be transferred to R_2. Region (23) can be achieved by using a superposition codebook defined using the pmf of the variables $(\mathrm{f}(A), A)$, which encode messages M_1 and M_2 and further superimposing a codebook U, which is used in the same manner as in [2, Theorem 1]. Further details can be found in [13, Appendix D].

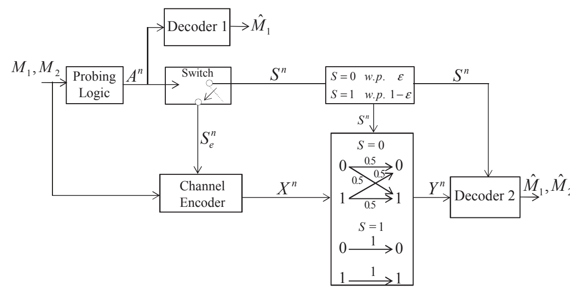

Figure 3. Channel coding with actions for channel state probing and with information embedding on actions.

C. Probing Capacity

Here we provide an example that illustrates the effect of the communication requirements of the action-cribbing decoder on the system performance. Consider the communication system shown in Fig. 3, where the state sequence S^n, i.i.d. and distributed according to pmf $p(s)$, is known to Decoder 2. We further assume binary actions, such that, if $A = 1$, the channel encoder observes the state S, and, if $A = 0$, it does not obtain any information about S. We model this aspect by defining the state information available at the encoder as $S_e = \mathrm{u}(S, A)$, where $\mathrm{u}(S, 1) = S$ and $\mathrm{u}(S, 0) = \mathrm{e}$, where e represents as "erasure" symbol. Following [3], we refer to this problem as having a "probing" encoder.

The channel encoder maps the state information S_e^n and messages M_1, M_2 into a codeword X^n (see Fig. 3). Moreover, two cost constraints, namely $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\gamma_a(A_i)\right] \leq \Gamma_A$ and $\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\gamma_x(X_i)\right] \leq \Gamma_X$ are imposed for given action input cost functions $\gamma_a : \mathcal{A} \to [0, \Lambda_{a,max}]$ and $\gamma_x : \mathcal{X} \to [0, \Lambda_{x,max}]$ with $0 \leq \Lambda_{a,max} < \infty$ and $0 \leq \Lambda_{x,max} < \infty$, respectively. In [3, Theorem 1], a correspondence was proved between the set-up of a probing encoder and that of action dependent states. Using [3, Theorem 1] and Proposition 3, we can easily obtain that the capacity-cost region $\mathcal{C}(\Gamma_A, \Gamma_X)$ for

[2]The cardinality constraints follow from [2, Theorem 1]

the system in Fig. 3 is given by the union of all rate pairs (R_1, R_2) such that the inequalities

$$R_1 \leq H(A|Q) \tag{24a}$$

$$\text{and } R_1 + R_2 \leq I(X;Y|S,Q), \tag{24b}$$

are satisfied, where the mutual informations are evaluated with respect to the joint pmf

$$p(q,a,s,s_e,x,y) = p(q)p(a|q)p(s)\mathbf{1}_{\{s_e=\mathrm{u}(s,a)\}} \\ p(x|s_e,a,q)p(y|x,s), \tag{25}$$

for some pmfs $p(q)$, $p(a|q)$, $p(x|s_e,a,q)$ such that $\mathrm{E}[\gamma_a(A)] \leq \Gamma_A$ and $\mathrm{E}[\gamma_x(X)] \leq \Gamma_X$.

We now apply (24a)-(24b) to the channel shown in Fig. 3 in which alphabets are binary $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0,1\}$, S is a Bern$(1-\epsilon)$ variable for $0 \leq \epsilon \leq 1$ and the channel is a binary symmetric with flipping probability 0.5 if $S = 0$ ("bad" channel state) and 0 if $S = 1$ ("good" channel state).

To evaluate the maximum achievable sum-rate $R_1 + R_2$ for a given rate R_1, we define $\Pr[A=1] = \gamma$, $\Pr[X=1|S_e = 1, A=1] = p_1$ and $\Pr[X=1|S_e = \mathrm{e}, A=0] = p_2$, and we set $\Pr[X=1|S_e=0, A=1] = 0$ without loss of optimality. The maximum sum-rate $R_1 + R_2$ for a given rate R_1 is then obtained from (24b) by solving the problem

$$R_1 + R_2 = \\ \max_{0 \leq p_1, p_2, \gamma \leq 1} \gamma(1-\epsilon)H(p_1) + (1-\gamma)(1-\epsilon)H(p_2), \tag{26}$$

under the constraint $\mathrm{E}[X] = p_1\gamma(1-\epsilon) + p_2(1-\gamma) \leq \Gamma_X$, $\mathrm{E}[A] = \gamma \leq \Gamma_A$ and $H(A) = H(\gamma) \geq R_1$. Note that the last constraint imposes that the rate achievable by the Decoder 1 is larger than R_1 as per (24a).

The sum-rate in (26) is shown in Fig. 4 for $\epsilon = 0.5$, $\Gamma_A = 1$ and different values of R_1. It can be seen that, for sufficiently small values of the cost constraint Γ_X, increasing the communication requirements, i.e., R_1, of the Decoder 1, reduces the achievable sum-rate $R_1 + R_2$. This is due to the fact that increasing R_1 requires to encode more information in the action sequence, which in turn reduces the portion of the actions that can be set to $A = 1$, i.e., $\Pr[A=1]$. As a result, the encoder is less informed about the state sequence and thus bound to waste some power on bad channel states.

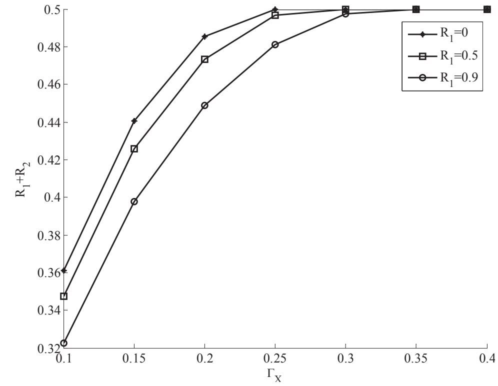

Figure 4. Sum-rate $R_1 + R_2$ versus the input cost constraint Γ_X for values of $R_1 = 0$, $R_1 = 0.5$ and $R_1 = 0.9$.

IV. Concluding Remarks

There is a profound interplay between actuation and communication in that both actuation can be instrumental to improve the efficiency of communication, and, vice versa, communication, implicit or explicit, can provide an essential tool to improve control tasks. This work has focused on the first type of interplay, and has investigated the implications of embedding information directly in the actions for the aim of communicating with a separate decoder. The communication requirements of this decoder are generally in conflict with the goal of improving the efficiency of the given communication link. This performance trade-off has been studied here for both source and channel coding. The results provided in the paper allow to give a quantitative answer to the questions posed in Sec. I regarding the impact of the requirements of action information embedding on the system performance.

V. Acknowledgement

The work of O. Simeone was supported in part by the U.S. National Science Foundation under Grant No. 0914899. H. Asnani was supported by Scott A. and Geraldine D. Macomber Stanford Graduate Fellowship. H. H. Permuter was supported in part by the Marie Curie Reintegration fellowship.

References

[1] H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 57, pp. 4530–4544, Jul 2011.

[2] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inform. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.

[3] H. Asnani, H. Permuter, and T. Weissman, "Probing capacity," *IEEE Trans. Inform. Theory*, vol. 57, no. 11, pp. 7317–7332, Nov. 2011.

[4] K. Kittichokechai, T. J. Oechtering, and M. Skoglund, "Source coding with common reconstruction and action-dependent side information," in *Proc. IEEE Information Theory Workshop* (ITW 2010), pp.1-5, Aug. 30 2010-Sept. 3 2010.

[5] H. Asnani and T. Weissman, "On real time coding with limited lookahead," submitted to *IEEE Trans. Inform. Theory,* arXiv:1105.5755.

[6] C. Choudhuri and U. Mitra, "How useful is adaptive action?," in *Proc. IEEE Globecom 2012*.

[7] H. Asnani, H. Permuter, and T. Weissman, "To feed or not to feed back," submitted to *IEEE Trans. Inform. Theory,* arXiv:1011.1607.

[8] Y. Chia, H. Asnani, and T. Weissman, "Multi-terminal source coding with action dependent side information," submitted (arXiv:1110.6916).

[9] Y. Steinberg and T. Weissman, "The degraded broadcast channel with action-dependent states," in *Proc. IEEE International Symposium on Information Theory* (ISIT 2012), Cambridge, MA, USA, July 1-6, 2012.

[10] B. Ahmadi, O. Simeone, C. Choudhuri, and U. Mitra, "On cascade source coding with a side information "Vending Machine"," submitted to *IEEE Trans. Inform. Theory,* arXiv:1204.1548.

[11] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563- 593, Mar. 2003.

[12] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell System Tech. J.*, vol. 53, pp. 1681–1721, 1974.

[13] B. Ahmadi, H. Asnani, O. Simeone and H. H. Permuter, "Information Embedding on Actions," submitted to *IEEE Trans. Inform. Theory,* arXiv:1207.6084.

[14] A. El Gamal and Y. Kim, *Network Information Theory*, Cambridge University Press, Dec 2011.

[15] J. Korner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. on Information Theory*, vol.23, no.1, pp. 60-64, Jan 1977.