

# The source coding game with a cheating switcher

Hari Palaiyanur  
hpalaiya@eecs.berkeley.edu

Cheng Chang  
cchang@eecs.berkeley.edu

Anant Sahai  
sahai@eecs.berkeley.edu

Department of Electrical Engineering and Computer Sciences  
University of California at Berkeley  
Berkeley, CA 94720

**Abstract**—Berger’s paper ‘The Source Coding Game’, *IEEE Trans. Inform. Theory*, 1971, considers the problem of finding the rate-distortion function for an adversarial source comprised of multiple known IID sources. The adversary, called the ‘switcher’, was allowed only causal access to the source realizations and the rate-distortion function was obtained through the use of a type covering lemma. In this paper, the rate-distortion function of the adversarial source is described, under the assumption that the switcher has non-causal access to all source realizations. The proof utilizes the type covering lemma and simple conditional, random ‘switching’ rules. The rate-distortion function is once again the maximization of the  $R(D)$  function for a region of attainable IID distributions.

## I. INTRODUCTION

The rate distortion function,  $R(D)$ , specifies the number of codewords, on an exponential scale, needed to represent a source to within a distortion  $D$ . Shannon [1] showed that for an additive distortion function  $d$  and a known discrete source that produces independent and identically distributed (IID) letters according to a distribution  $p$ ,

$$R(D) = R_p(D) \triangleq \min_{W: \sum_{x,y} p(x)W(y|x)d(x,y) \leq D} I(p, W) \quad (1)$$

where  $I(p, W)$  is the mutual information for an input distribution  $p$  and probability transition matrix  $W$ .

Sakrison [2] studied the rate distortion function for the class of *compound* sources. That is, the source is assumed to come from a known set of distributions and is fixed for all time. If  $G$  is the set of possible sources, Sakrison showed that planning for the worst case source is both necessary and sufficient in the discrete memoryless source case. Hence, for compound sources,

$$R(D) = \max_{p \in G} R_p(D) \quad (2)$$

In Berger’s ‘source coding game’ [3], the source is assumed to be an adversarial player called the ‘switcher’ in a statistical game. In this setup, the switcher is allowed to choose any source from  $G$  at any time, but must do so in a causal manner without access to the current step’s source realizations. The conclusion of [3] is that under this scenario,

$$R(D) = \max_{p \in \bar{G}} R_p(D) \quad (3)$$

where  $\bar{G}$  is the convex hull of  $G$ . In his conclusion, Berger poses the question of what happens to the rate-distortion

function when the rules of the game are tilted in favor of the switcher. Suppose that the switcher were given access to the source realizations before having to choose the switch positions. The main result of this paper is that under these rules,

$$R(D) = \max_{p \in \mathcal{C}} R_p(D) \quad (4)$$

where

$$\mathcal{C} = \left\{ p \in \mathcal{P} : \begin{array}{l} \sum_{i \in \mathcal{V}} p(i) \geq \prod_{l=1}^m \sum_{i \in \mathcal{V}} p_l(i) \\ \forall \mathcal{V} \text{ such that} \\ \mathcal{V} \subseteq \mathcal{X} \end{array} \right\} \quad (5)$$

Here, the  $p_l$  are the distributions of the  $m$  sources and  $\mathcal{P}$  is the set of all probability distributions on  $\mathcal{X}$ .

Section II sets up the notation for the paper, and is followed by a description of the source coding game in Section III. The main result is stated in Section IV, and an example illustrating the main ideas is given in Section V. The proofs are located in Section VI and some concluding remarks are made in Section VII.

## II. DEFINITIONS

We work in essentially the same setup as Berger’s source coding game [3], and with most of the same notation. There are two finite alphabets  $\mathcal{X}$  and  $\mathcal{Y}$ . Without loss of generality,  $\mathcal{X} = \{1, 2, \dots, |\mathcal{X}|\}$  is the source alphabet and  $\mathcal{Y} = \{1, 2, \dots, |\mathcal{Y}|\}$  is the reproduction alphabet. Let  $\vec{x} = (x_1, \dots, x_n)$  denote an arbitrary vector from  $\mathcal{X}^n$  and  $\vec{y} = (y_1, \dots, y_n)$  an arbitrary vector from  $\mathcal{Y}^n$ . When needed,  $\vec{x}^k = (x_1, \dots, x_k)$  will be used to denote the first  $k$  symbols in the vector  $\vec{x}$ .

Let  $d : \mathcal{X} \times \mathcal{Y} \rightarrow [0, \infty)$  be a distortion measure<sup>1</sup> (any nonnegative function) on the product set  $\mathcal{X} \times \mathcal{Y}$ . Then define  $d_n : \mathcal{X}^n \times \mathcal{Y}^n \rightarrow [0, \infty)$  for  $n \geq 1$  to be

$$d_n(\vec{x}, \vec{y}) = \frac{1}{n} \sum_{k=1}^n d(x_k, y_k) \quad (6)$$

Let  $\mathcal{P}$  be the set of probability distributions on  $\mathcal{X}$ ,  $\mathcal{P}_n$  the set of types of length  $n$  strings from  $\mathcal{X}$ , and let  $\mathcal{W}$  be the set of probability transition matrices from  $\mathcal{X}$  to  $\mathcal{Y}$ . The rate

<sup>1</sup>We could allow for infinite distortions and require that the probability that the distortion exceed  $D + \epsilon$  go to zero for all  $\epsilon > 0$ . The main result would hold in this setup as well.

distortion function of  $p \in \mathcal{P}$  with respect to distortion measure  $d$  is defined to be

$$R_p(D) = \min_{w \in \mathcal{W}(p,D)} I(p, w) \quad (7)$$

where

$$\mathcal{W}(p, D) = \left\{ w \in \mathcal{W} : \sum_{i=1}^{|\mathcal{X}|} \sum_{j=1}^{|\mathcal{Y}|} p(i)w(j|i)d(i, j) \leq D \right\} \quad (8)$$

and  $I(p, w)$  is the mutual information<sup>2</sup>

$$I(p, w) = \sum_{i=1}^{|\mathcal{X}|} \sum_{j=1}^{|\mathcal{Y}|} p(i)w(j|i) \log_2 \left[ \frac{w(j|i)}{\sum_{i'=1}^{|\mathcal{X}|} p(i')w(j|i')} \right] \quad (9)$$

The only interesting domain of values for  $R_p(D)$  is  $D \in (D_{\min}(p), D_{\max}(p))$  where

$$D_{\min}(p) = \sum_{i=1}^{|\mathcal{X}|} p(i) \min_j d(i, j) \quad (10)$$

$$D_{\max}(p) = \min_j \sum_{i=1}^{|\mathcal{X}|} p(i)d(i, j) \quad (11)$$

Let  $B = \{\vec{y}_1, \dots, \vec{y}_K\}$  be a codebook of length  $n$  vectors in  $\mathcal{Y}^n$ . Define

$$d_n(\vec{x}; B) = \min_{\vec{y} \in B} d_n(\vec{x}, \vec{y}) \quad (12)$$

If  $B$  is used to represent an IID source with distribution  $p$ , then the average distortion of  $B$  is defined to be

$$d(B) = \sum_{\vec{x} \in \mathcal{X}^n} P(\vec{x})d_n(\vec{x}; B) = E[d_n(\vec{x}; B)] \quad (13)$$

where

$$P(\vec{x}) = \prod_{k=1}^n p(x_k) \quad (14)$$

Let  $K(n, D)$  be the minimum number of codewords needed in a codebook  $B \subset \mathcal{Y}^n$  so that  $d(B) \leq D$ . Then, Shannon's Rate-Distortion Theorem ([1], [4]) says that if the source is IID with distribution  $p$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 K(n, D) = R_p(D) \quad (15)$$

### III. THE SOURCE CODING GAME

We suppose as in Berger's paper that a 'switcher' is a player in a two person game with access to the position of a switch which can be in one of  $m$  positions. The switch position  $l$ ,  $1 \leq l \leq m$  corresponds to a memoryless source with distribution  $p_l(\cdot)$  that is independent of all the other sources<sup>3</sup>. Let  $\mathbf{s} = (s_1, s_2, \dots, s_n)$  be the vector of switch positions chosen by the switcher. Let  $x_k$  be the switcher's output at time  $k$  and let

<sup>2</sup>We use  $\log_2$  in the report, but any base can be used.

<sup>3</sup>There can be multiple copies of the same source. For example, there can be any number of copies of a Bernoulli (1/10) source, so long as they are all independent. In that sense, the switcher has access to a list of  $m$  sources, rather than a set of  $m$  different distributions.

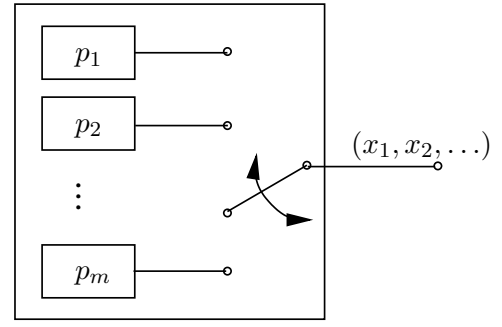


Fig. 1. The source coding game.

$x_{l,k}$  be the output of the  $l^{\text{th}}$  source at time  $k$ . When needed,  $\vec{x}_l$  will denote the block of  $n$  symbols for the  $l^{\text{th}}$  source.

The other person in the game is called the 'coder'. The coder's goal is to construct a codebook of minimal size to ensure the average distortion between the switcher's output and reconstruction in the codebook is at most  $D$ . Fix  $n$  and  $D \geq 0$ . Let  $B$  denote the codebook chosen by the coder, and  $d_n(\vec{x}; B)$  be the distortion between a vector  $\vec{x}$  and the best reproduction of  $\vec{x}$  in  $B$ ; in the sense of least distortion. The payoff of the game is the average distortion, which for a particular switching strategy is

$$E[d(\vec{x}; B)] = \sum_{\vec{x} \in \mathcal{X}^n} P_S(\vec{x})d_n(\vec{x}; B) \quad (16)$$

Here  $P_S(\vec{x})$  is the probability of the switcher outputting the sequence  $\vec{x}$  averaged over any randomness the switcher chooses to use, as well as the randomness in the sources. Let  $P(\mathbf{s}, \vec{x})$  be the probability of the switcher using a switching vector  $\mathbf{s}$  and outputting a string  $\vec{x}$ . Then,

$$P_S(\vec{x}) = \sum_{\mathbf{s} \in \{1, \dots, m\}^n} P(\mathbf{s}, \vec{x}) \quad (17)$$

In Berger's original game, the coder chooses a codebook that is revealed to the switcher. The switcher must then choose the switch position at every integer time  $k$  without access to the actual letters that the sources produce at that time. The switcher, however, has access to the previous outputs of the switch. So in [3], an admissible joint probability rule for  $P(\mathbf{s}, \vec{x})$  is of the form

$$P(\mathbf{s}, \vec{x}) = \prod_{k=1}^n P(s_k | \mathbf{s}^{k-1}, \vec{x}^{k-1}) P_{s_k}(x_k) \quad (18)$$

In this discussion, we consider the case when the switcher gets to see the outputs of the  $m$  sources and then has to output a letter from one of the letters that the sources produced. The switcher outputs a letter,  $x_k$ , which must come from the (possibly proper) subset of  $\mathcal{X}$ ,  $\{x_{1,k}, \dots, x_{m,k}\}$ . Hence, for this 'cheating' switcher, allowable strategies are of the form

$$P(\mathbf{s}, \vec{x} | \vec{x}_1, \dots, \vec{x}_m) = P(\mathbf{s} | \vec{x}_1, \dots, \vec{x}_m) \mathbf{1}(x_k = x_{s_k, k}, \mathbf{1} \leq k \leq n) \quad (19)$$

Since the sources are still IID,

$$P(\vec{x}_1, \dots, \vec{x}_m) = \prod_{l=1}^m \prod_{k=1}^n p_l(x_{l,k}) \quad (20)$$

Define the minimum number of codewords needed by the coder to guarantee average distortion  $D$  as  $M(n, D)$ .

$$M(n, D) = \min \left\{ |B| : \begin{array}{l} B \subset \mathcal{Y}^n, E[d(\vec{x}; B)] \leq D \\ \text{for all allowable} \\ \text{switcher strategies} \end{array} \right\} \quad (21)$$

We are interested in the exponential rate of growth of  $M(n, D)$  with  $n$ . Define

$$R(D) = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 M(n, D) \quad (22)$$

Let  $G = \{p_1(\cdot), \dots, p_m(\cdot)\}$  be the set of  $m$  distributions on  $\mathcal{X}$  the switcher has access to. Let  $\bar{G}$  be the convex hull of  $G$ . Then let

$$R^*(D) = \max_{p \in \bar{G}} R_p(D)$$

The conclusion of [3] is that  $R(D) = R^*(D)$  when the switcher is not allowed to witness the source realizations until committing to a switch position.

#### IV. MAIN RESULT

The main result is the determination of  $R(D)$  in the case when the switcher gets to see the entire block of  $mn$  source outputs ahead of choosing the switching sequence.

*Theorem 1:* Let the switcher ‘cheat’ and have access to the  $n$  outputs of all  $m$  sources before choosing a symbol for each time  $k$ . Then,

$$R(D) = \tilde{R}(D) \triangleq \max_{p \in \mathcal{C}} R_p(D) \quad (23)$$

where  $\mathcal{C}$  is defined in (5).

Here, we have defined  $\tilde{R}(D) = \max_{p \in \mathcal{C}} R_p(D)$ . The theorem’s conclusion is that when the switcher is allowed to ‘cheat’,  $R(D) = \tilde{R}(D)$ . The number of constraints in the set  $\mathcal{C}$  is exponential in the size of  $\mathcal{X}$ . Depending on the source distributions, a large number of these constraints could be inactive. Unfortunately,  $R_p(D)$  is generally not concave in  $p$  for a fixed  $D$ , so computation of  $\tilde{R}(D)$  may not be easy.

Qualitatively, allowing the switcher to ‘cheat’ gives access to distributions  $p \in \mathcal{C}$  which may not be  $\bar{G}$ . Quantitatively, the conditions placed on the distributions in  $\mathcal{C}$  are precisely those that restrict the switcher from producing symbols that do not occur often enough on average. For example, let  $\mathcal{V} = \{1\}$ . Then for every  $p \in \mathcal{C}$ ,

$$p(1) \geq \prod_{l=1}^m p_l(1)$$

Since the sources are independent,  $\prod_{l=1}^m p_l(1)$  is the probability that all  $m$  sources produce the letter 1 at a given time. In this case, the switcher has no option but to output the letter 1, hence any distribution the switcher mimics must

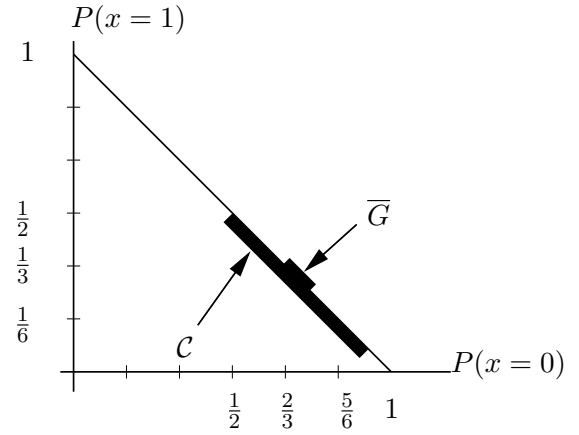


Fig. 2. The binary distributions the switcher can mimic.  $\bar{G}$  is the set of distributions the switcher can mimic without cheating, and  $\mathcal{C}$  is the set attainable with cheating.

have  $p(1) \geq \prod_{l=1}^m p_l(1)$ . The same logic can be applied to all subsets  $\mathcal{V}$  of  $\mathcal{X}$ .

As commented in Section V of [3],  $\tilde{R}(D) = R^*(D)$  if  $R^*(D) = \max_{p \in \mathcal{P}} R_p(D)$ . Before giving the proof of the result, an example is presented.

#### V. AN EXAMPLE

Suppose the switcher has access to two IID binary sources. Source 1 outputs 1 with probability  $1/3$  and source 2 outputs 1 with probability  $1/4$ . Then, since the sources are IID across time and independent of each other, for any time  $k$ ,

$$P(x_{1,k} = x_{2,k} = 0) = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2} \quad (24)$$

Similarly,

$$P(x_{1,k} = x_{2,k} = 1) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12} \quad (25)$$

Hence,

$$P(\{x_{1,k}, x_{2,k}\} = \{0, 1\}) = 1 - \frac{1}{2} - \frac{1}{12} = \frac{5}{12} \quad (26)$$

If at time  $k$ , the switcher has the option of choosing either 0 or 1, suppose the switcher chooses 1 with probability  $f_1$ . This strategy is memoryless, but it is an allowable strategy for the ‘cheating’ switcher. The coder then sees an IID binary source with a probability of a 1 occurring being equal to:

$$p(1) = \frac{1}{12} + \frac{5}{12} f_1 \quad (27)$$

By using  $f_1$  as a parameter, the switcher can produce 1’s with a probability between  $1/12$  and  $1/2$ . The attainable distributions are shown in Figure 2. This kind of memoryless, ‘conditional’ switching strategy will be used for half of the proof of the main result. If the distortion measure is Hamming distortion, clearly the switcher will choose  $f_1 = 1$  and produce a Bernoulli  $1/2$  process. Regardless of the distortion measure,  $\mathcal{C}$  contains all the distributions on  $\mathcal{X}$  that the switcher can mimic.

## VI. PROOFS

## A. Achievability for the coder

First, the main tool of this section is stated.

*Lemma 1 (Type Covering [3]):* Let  $\mathcal{P}_n$  denote the set of types for length  $n$  sequences from  $\mathcal{X}$ . Let  $S_D(\vec{y}) \triangleq \{\vec{x} \in \mathcal{X}^n : d_n(\vec{x}, \vec{y}) \leq D\}$  be the set of  $\mathcal{X}^n$  strings that are within distortion  $D$  of a  $\mathcal{Y}^n$  string  $\vec{y}$ . Fix a  $p \in \mathcal{P}_n$  and an  $\epsilon > 0$ . Then there exists a codebook  $B = \{\vec{y}_1, \vec{y}_2, \dots, \vec{y}_M\}$  where  $M < \exp_2(n(R_p(D) + \epsilon))$  and

$$T_p^n \subseteq \bigcup_{k=1}^M S_D(\vec{y}_k)$$

where  $T_p^n$  is the set of  $\mathcal{X}^n$  strings with type  $p$  for  $n$  large enough.

We now show how the coder can get arbitrarily close to  $\tilde{R}(D)$  for large enough  $n$ . For  $\delta > 0$ , define  $\mathcal{C}_\delta$  as

$$\mathcal{C}_\delta \triangleq \left\{ p \in \mathcal{P} : \begin{array}{l} \sum_{i \in \mathcal{V}} p(i) \geq \prod_{l=1}^m \sum_{i \in \mathcal{V}} p_l(i) - \delta \\ \forall \mathcal{V} \text{ such that} \\ \mathcal{V} \subseteq \mathcal{X} \end{array} \right\}$$

*Lemma 2 (Converse for switcher):* Let  $\epsilon > 0$ . For all  $n$  sufficiently large

$$\frac{1}{n} \log_2 M(n, D) \leq \tilde{R}(D) + \epsilon$$

*Proof:* We know  $R_p(D)$  is a continuous function of  $p$  ([5]). It follows then that because  $\mathcal{C}_\delta$  is monotonically decreasing (as a set) with  $\delta$  that for all  $\epsilon > 0$ , there is a  $\delta > 0$  so that

$$\max_{p \in \mathcal{C}_\delta} R_p(D) \leq \max_{p \in \mathcal{C}} R_p(D) + \epsilon/2$$

We will have the coder use a codebook such that all  $\mathcal{X}^n$  strings with types in  $\mathcal{C}_\delta$  are covered within distortion  $D$ . The coder can do this for large  $n$  with at most  $M$  codewords where

$$M < (n+1)^{|\mathcal{X}|} \exp_2(n \max_{p \in \mathcal{C}_\delta} R_p(D)) \quad (28)$$

$$\leq (n+1)^{|\mathcal{X}|} \exp_2(n(\max_{p \in \mathcal{C}} R_p(D) + \epsilon)) \quad (29)$$

Explicitly, this is done by taking a union of the codebooks provided by the type covering lemma and noting that the number of types is less than  $(n+1)^{|\mathcal{X}|}$ . Next, we will show that the probability of the switcher being able to produce a string with a type not in  $\mathcal{C}_\delta$  goes to 0 exponentially with  $n$ .

Consider a type  $p \in \mathcal{P}_n \cap (\mathcal{P} - \mathcal{C}_\delta)$ . By definition, there is some  $\mathcal{V} \subseteq \mathcal{X}$  such that  $\sum_{i \in \mathcal{V}} p(i) < \prod_{l=1}^m \sum_{i \in \mathcal{V}} p_l(i) - \delta$ . Let  $\alpha_k(\mathcal{V})$  be the indicator function

$$\alpha_k(\mathcal{V}) = \prod_{l=1}^m \mathbf{1}(x_{l,k} \in \mathcal{V})$$

$\alpha_k$  indicates the event that the switcher cannot output a symbol outside of  $\mathcal{V}$  at time  $k$ . Then  $\alpha_k(\mathcal{V})$  is a Bernoulli random variable with a probability of being 1 equal to  $Q(\mathcal{V}) \triangleq \prod_{l=1}^m \sum_{i \in \mathcal{V}} p_l(i)$ . That is, we can envision  $\alpha_k(\mathcal{V})$  as being

a sequence of IID binary random variables with distribution  $q' \triangleq (1 - Q(\mathcal{V}), Q(\mathcal{V}))$ .

Now for our type  $p \in \mathcal{P}_n \cap (\mathcal{P} - \mathcal{C}_\delta)$ , we have that for all strings  $\vec{x}$  in the type class  $T_p$ ,  $\frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i \in \mathcal{V}) < Q(\mathcal{V}) - \delta$ . Let  $p'$  be the binary distribution  $(1 - Q(\mathcal{V}) + \delta, Q(\mathcal{V}) - \delta)$ , assuming  $\delta$  is small enough to make this a distribution (if not, make delta small enough). Therefore  $\|p' - q'\|_1 = 2\delta$ , and hence  $D(p' || q') \geq \delta / \ln 2$  by Pinsker's inequality. Using standard types properties [6] gives

$$\begin{aligned} P\left(\frac{1}{n} \sum_{k=1}^n \alpha_k(\mathcal{V}) < Q(\mathcal{V}) - \delta\right) &\leq \exp_2(-nD(p' || q')) \\ &\leq \exp_2(-n\delta / \ln 2) \end{aligned}$$

If we let  $E$  be the event that  $\vec{x}$  has a type which is not in  $\mathcal{C}_\delta$ , we just sum over types not in  $\mathcal{C}_\delta$  to get

$$\begin{aligned} P(E) &\leq \sum_{p \in \mathcal{P}_n \cap (\mathcal{P} - \mathcal{C}_\delta)} \exp_2(-n\delta / \ln 2) \\ &\leq (n+1)^{|\mathcal{X}|} \exp_2(-n\delta / \ln 2) \\ &= \exp_2\left(-n\left(\frac{\delta}{\ln 2} - |\mathcal{X}| \frac{\ln(n+1)}{n}\right)\right) \end{aligned}$$

Now let  $d^* = \max_{x,y} d(x,y) < \infty$ . Then, regardless of the switcher strategy,

$$E[d(\vec{x}; B)] \leq D + d^* \cdot \exp_2\left(-n\left(\frac{\delta}{\ln 2} - |\mathcal{X}| \frac{\ln(n+1)}{n}\right)\right)$$

So for large  $n$  we can get arbitrarily close to distortion  $D$  while the rate is at most  $\tilde{R}(D) + \epsilon$ . Using the fact that the rate-distortion function is continuous in  $D$  gives us that the coder can achieve at most distortion  $D$  on average while the rate is at most  $\tilde{R}(D) + \epsilon$ . Since  $\epsilon$  is arbitrary,  $R(D) \leq \tilde{R}(D)$ . ■

## B. Achievability for the switcher

This section considers why  $R(D) \geq \tilde{R}(D)$ . We will show that the switcher can target any distribution  $p \in \mathcal{C}$  and produce a sequence of IID symbols with distribution  $p$ . In particular, the switcher can target the distribution that yields  $\max_{p \in \mathcal{C}} R_p(D)$  and Shannon's rate distortion theorem gives  $R(D) \geq \tilde{R}(D)$ .

The switcher will use a memoryless randomized strategy. Let  $\mathcal{V} \subseteq \mathcal{X}$  and suppose that at some time  $k$  the set of symbols available to choose from for the switcher is exactly  $\mathcal{V}$ . That is  $\{x_{1,k}, \dots, x_{m,k}\} = \mathcal{V}$ . Define  $\beta(\mathcal{V}) \triangleq P(\{x_{1,1}, \dots, x_{m,1}\} = \mathcal{V})$  to be the probability that at any time the switcher can choose any element of  $\mathcal{V}$  and no other symbols. Then let  $f(i|\mathcal{V})$  be a probability distribution on  $\mathcal{X}$  with support  $\mathcal{V}$ , i.e.  $f(i|\mathcal{V}) \geq 0$ ,  $\forall i \in \mathcal{X}$ ,  $f(i|\mathcal{V}) = 0$  if  $i \notin \mathcal{V}$ , and  $\sum_{i \in \mathcal{V}} f(i|\mathcal{V}) = 1$ . The switcher will have such a randomized rule for every nonempty subset  $\mathcal{V}$  of  $\mathcal{X}$  such that  $|\mathcal{V}| \leq m$ . Let  $\mathcal{D}$  be the set of distributions on  $\mathcal{X}$  that can be achieved with these kinds of rules, so

$$\mathcal{D} \triangleq \left\{ p \in \mathcal{P} : \begin{array}{l} p(\cdot) = \sum_{\mathcal{V} \subseteq \mathcal{X}, |\mathcal{V}| \leq m} \beta(\mathcal{V}) f(\cdot|\mathcal{V}), \\ \forall \mathcal{V} \text{ s.t. } \mathcal{V} \subseteq \mathcal{X}, |\mathcal{V}| \leq m, \\ f(\cdot|\mathcal{V}) \text{ is a PMF on } \mathcal{V} \end{array} \right\}$$

It is clear from the construction of  $\mathcal{D}$  that  $\mathcal{D} \subseteq \mathcal{C}$  because the conditions in  $\mathcal{C}$  are those that prevent the switcher only from producing symbols that do not occur enough, but put no further restrictions on the switcher. So we need only show that  $\mathcal{C} \subseteq \mathcal{D}$ . The following gives such a proof by contradiction.

*Lemma 3 (Achievability for switcher):* The set relation  $\mathcal{C} \subseteq \mathcal{D}$  is true.

*Proof:* Suppose  $p \in \mathcal{C}$  but  $p \notin \mathcal{D}$ . It is clear that  $\mathcal{D}$  is a convex set. Let us view the probability simplex in  $\mathbb{R}^{|\mathcal{X}|}$ . Since  $\mathcal{D}$  is a convex set, there is a hyperplane through  $p$  that does not intersect  $\mathcal{D}$ . Hence, there is a vector  $(a_1, \dots, a_{|\mathcal{X}|})$  such that  $\sum_{i=1}^{|\mathcal{X}|} a_i p(i) = t$  for some real  $t$  but  $t < \min_{q \in \mathcal{C}} \sum_{i=1}^{|\mathcal{X}|} a_i q(i)$ . Without loss of generality, assume  $a_1 \geq a_2 \geq \dots \geq a_{|\mathcal{X}|}$  (otherwise permute symbols). Now, we will construct  $f(\cdot|\mathcal{V})$  so that the resulting  $q$  has  $\sum_{i=1}^{|\mathcal{X}|} a_i p(i) \geq \sum_{i=1}^{|\mathcal{X}|} a_i q(i)$ , which contradicts the initial assumption. Let

$$f(i|\mathcal{V}) \triangleq \begin{cases} 1 & \text{if } i = \max(\mathcal{V}) \\ 0 & \text{else} \end{cases}$$

For example, if  $\mathcal{V} = \{1, 5, 6, 9\}$ , then  $f(9|\mathcal{V}) = 1$  and  $f(i|\mathcal{V}) = 0$  if  $i \neq 9$ . Call  $q$  the distribution on  $\mathcal{X}$  induced by this choice of  $f(\cdot|\mathcal{V})$ . Recall that  $Q(\mathcal{V}) = \prod_{l=1}^m \sum_{i \in \mathcal{V}} p_l(i)$ . Then, we have

$$\begin{aligned} \sum_{i=1}^{|\mathcal{X}|} a_i q(i) &= a_1 Q(\{1\}) + a_2 [Q(\{1, 2\}) - Q(\{1\})] + \\ &\dots + a_{|\mathcal{X}|} [Q(\{1, \dots, |\mathcal{X}|\}) - Q(\{1, \dots, |\mathcal{X}| - 1\})] \end{aligned}$$

By the constraints in the definition of  $\mathcal{C}$ , we have the following inequalities for  $p$ :

$$\begin{aligned} p(1) &\geq Q(\{1\}) = q(1) \\ p(1) + p(2) &\geq Q(\{1, 2\}) = q(1) + q(2) \\ &\vdots \\ \sum_{i=1}^{|\mathcal{X}|-1} p(i) &\geq Q(\{1, \dots, |\mathcal{X}| - 1\}) = \sum_{i=1}^{|\mathcal{X}|-1} q(i) \end{aligned}$$

Therefore, the difference of the objective is

$$\begin{aligned} \sum_{i=1}^{|\mathcal{X}|} a_i (p(i) - q(i)) &= \\ &a_{|\mathcal{X}|} \left[ \sum_{i=1}^{|\mathcal{X}|} p(i) - q(i) \right] + \\ &(a_{|\mathcal{X}|-1} - a_{|\mathcal{X}|}) \left[ \sum_{i=1}^{|\mathcal{X}|-1} p(i) - q(i) \right] + \\ &\dots + (a_1 - a_2) \left[ p(1) - q(1) \right] \\ &= \sum_{i=1}^{|\mathcal{X}|-1} (a_i - a_{i+1}) \left[ \sum_{j=1}^i p(j) - \sum_{j=1}^i q(j) \right] \\ &\geq 0 \end{aligned}$$

The last step is true because of the monotonicity in the  $a_i$  and the inequalities we derived earlier. Therefore, we see that  $\sum_{i=1}^{|\mathcal{X}|} a_i p(i) \geq \sum_{i=1}^{|\mathcal{X}|} a_i q(i)$  for the  $p$  we had chosen at the beginning of the proof. This contradicts the assumption that  $\sum_{i=1}^{|\mathcal{X}|} a_i p(i) < \min_{q \in \mathcal{D}} \sum_{i=1}^{|\mathcal{X}|} a_i q(i)$ , therefore it must be that  $\mathcal{C} \subseteq \mathcal{D}$ . ■

## VII. CONCLUSION

The rate-distortion function for the ‘cheating’ switcher has been described. It is the maximization of the IID rate-distortion function over the distributions the switcher can simulate. It was assumed the switcher had access to all source outputs ahead of time, but the proof required only that the switcher had access to the source realizations for one step ahead at each time.

In this paper, the sources were independent and memoryless. A minor tweak to the argument also gets the rate-distortion function if the sources are dependent but still memoryless. The region  $\mathcal{C}$  would just be modified to become:

$$\mathcal{C} = \left\{ p \in \mathcal{P} : \begin{array}{l} \sum_{i \in \mathcal{V}} p(i) \geq P(\cup_{l=1}^m x_{l,1} \subset \mathcal{V}) \\ \forall \mathcal{V} \text{ such that} \\ \mathcal{V} \subseteq \mathcal{X} \end{array} \right\}$$

A more interesting problem is to consider what happens when the sources are independent but have memory. Apparently, Dobrushin [7] has analyzed the case of the non-cheating switcher with independent sources with memory. One could imagine that, perhaps, giving the switcher access to all source realizations could result in the ability to simulate memoryless sources from a collection of sources with memory.

Similar techniques might also prove useful in considering a cheating ‘jammer’ for an arbitrarily varying channel. While the problem is mathematically well defined, it seems unphysical in the usual context of jamming or channel noise. The idea may make more sense in the context of watermarking, where the adversary can try many different attacks on different letters of the input before deciding to choose one for each.

## ACKNOWLEDGMENT

The authors would like to thank the NSFGRFP for partial support of this research. Also, we thank Prof. Michael Gastpar, the students of the Fall 2006 EE290S course at UC Berkeley, and the reviewers for helping to refine the presentation of this work.

## REFERENCES

- [1] C. Shannon, “Coding theorems for a discrete source with a fidelity criterion,” in *IRE Natl. Conv. Rec.*, 1959, pp. 142–163.
- [2] D. Sakrison, “The rate distortion function for a class of sources,” *Information and Control*, vol. 15, pp. 165–195, Mar. 1969.
- [3] T. Berger, “The source coding game,” *IEEE Trans. Inform. Theory*, vol. 17, pp. 71–76, Jan. 1971.
- [4] J. Wolfowitz, “Approximation with a fidelity criterion,” in *5th Berkeley Symp. on Math. Stat. and Prob.*, vol. 1. Berkeley, California: University of California, Press, 1967, pp. 565–573.
- [5] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York, NY: Academic Press, 1997.
- [6] T. Cover and J. Thomas, *Elements of Information Theory*. New York, NY: John Wiley and Sons, 1991.
- [7] R. Dobrushin, “Unified methods for the transmission of information: The general case,” *Sov. Math.*, vol. 4, pp. 284–292, 1963.