Capacity of Coordinated Actions

Thomas M. Cover
Department of Electrical Engineering
Stanford University
Stanford, CA
Email: cover@stanford.edu

Haim H. Permuter
Department of Electrical Engineering
Stanford University
Stanford, CA
Email: haim1@stanford.edu

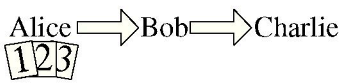
***Abstract*— We propose the problem of coordinating action over many nodes by distributed communication. The idea is to switch the emphasis from exchanging information to setting up cooperative action. Examples are given. We solve most 3-node problems but one remains open.**

I. Introduction

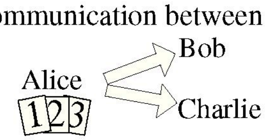
We start with two questions based on Fig. 1:

Alice, Bob and Charlie play a game with cards numbered 1,2, and 3. Alice receives a random card. The goal is for Bob and Charlie to choose cards that are different from Alice's card and different from each other. In other words, we want the cards held by Alice, Bob and Charlie to represent a fair deal, with Alice's card specified at random by nature. This is an example of distributed task assignment.

a) How much information must Alice transmit to Bob, and Bob transmit to Charlie, to achieve the goal if there is no direct communication between Alice and Charlie?



b) How much information must Alice transmit to Bob and transmit to Charlie to achieve the goal if there is no direct communication between Bob and Charlie?



Obviously, if Alice transmits her card number to Bob and Charlie, the players can achieve the goal. For instance, Bob will choose the larger number and Charlie will choose the least number from the pair of numbers that are possible. This requires $\log 3$ bits to be sent to each. However, can they do better? If we assume that the players have time synchronization, and a delay in picking the card is allowed, can we derive information theoretic bounds?

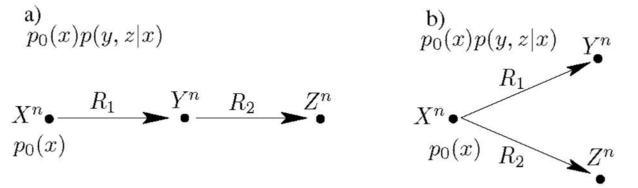


Fig. 1. Coordination capacity: What are the rates (R_1, R_2) that can achieve the joint distribution $p_0(x)p(y,z|x)$, when the actions at node X are specified by nature and distributed according to $p(x^n) = \prod_{i=1}^{n} p_0(x_i)$.

Those two questions are special cases of a more general question: How much information $\{R_{ij}\},\ i,j = 1,2,\ldots,m,$ must be conveyed between m nodes of a network to achieve a specified joint distribution $p(x_1, x_2, ..., x_m)$ at the nodes, given that the values at a certain subset of the nodes is specified? Equivalently, we ask for the set of all distributions achievable with communication rate $\{R_{ij}\}$, subject to specified random values at a subset of the nodes.

Applications might include task assignment (no two agents performing the same job), game theory (several agents taking joint action according to an optimal distribution [1]), communication (coherent relaying information), control (in a distributed environment), social planning (how do we achieve a desirable cooperation), and quantum information (quantum coding of mixed states [2], [3]).

A. Two nodes, no communication

Before developing the mathematical formulation of the problem let us discuss the case where we have two nodes and suppose that no communication between these nodes is allowed. However, the nodes can agree ahead of time how they will behave, i.e., common randomness is allowed.

We assume throughout that common randomness W is available to all nodes. Here W might be obtained

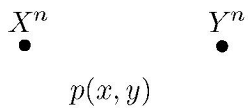


Fig. 2. No communication. Any distribution $p(x,y)$ can be achieved without communication between nodes.

from observation of some celestial event, or merely a prior agreement among the nodes to use a given randomly selected code. All that is necessary is that W be sufficiently rich ($W \sim$ uniform on the unit interval will do) and that W be independent of the nodes that are specified. A quantification of the amount of common randomness needed is given by Cuff [4]. The common randomness W plays much the same role as $\omega \in \Omega$ in the standard probability space set up $(\Omega, \mathcal{B}, P)$ with random variables $X(\omega), Y(\omega), \ldots$.

What is the set of all joint distributions $p(x,y)$ that can be achieved at these isolated nodes? The answer turns out to be any distribution whatsoever. This would seem to be the end of the problem.

But the problem changes dramatically when one of the nodes is specified to take on a certain value. Suppose that node X takes on a value x drawn according to the marginal distribution $p_0(x)$. Now what distributions $p(x,y)$ are achievable? The answer, which is given by Theorem 1 (when $R = 0$), turns out to be only those distributions of the form $p(x,y) = p_0(x)p(y)$. This makes sense. There is no communication between X and Y, so Y must be independent of X. Apparently, nature's insistence on a certain value for X restricts the set of joint distributions that can be achieved.

II. Problem Definition and Results.

A. Two nodes with communication

Here we assume that we have two nodes, where one node, say the source, produces a sequence of actions $X_1, X_2, \ldots, X_n$ i.i.d. $\sim p(x), x \in \mathcal{X}$, and the other node, say the agent, receives information from the source and can pick its own sequence of actions $Y_1, Y_2, \ldots, Y_n$ where $y_i \in \mathcal{Y}$. We assume throughout that the number of different actions $|\mathcal{X}|$ and $|\mathcal{Y}|$ are finite. The setting of distributed action with a single agent is illustrated in Fig. 3.

Definition 1: A $(2^{nR}, n)$ distributed action code consists of an encoding function,

$$f_n : \mathcal{X}^n \times \mathcal{W} \to \{1, 2, \ldots, 2^{nR}\} \tag{1}$$

and a decoding function,

$$g_n : \{1, 2, \ldots, 2^{nR}\} \times \mathcal{W} \to \mathcal{Y}^n. \tag{2}$$

The random variable $W \in \mathcal{W}$ plays the role of common randomness available at the encoder and decoder, where W is independent of X^n.

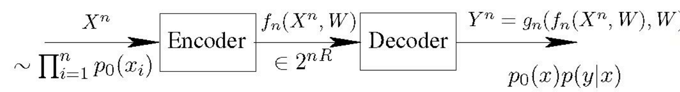


Fig. 3. Distributed action in the case of one source of actions X^n distributed according to $p(x^n) = \prod_{i=1}^n p_0(x_i)$ and one agent Y^n. The desired distribution $p_0(x)p(y|x)$ is achievable if $R > I(X;Y)$.

Definition 2: The *maximum variation* $v_n(p(x,y))$ of a $(2^{nR}, n)$ code (f_n, g_n) from the desired distribution $p(x,y) = p_0(x)p(y|x)$ is defined as

$$v_n(p(x,y)) \triangleq \max_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathbf{E}\left\{|P_{X^n,Y^n}(x,y) - p(x,y)|\right\} \tag{3}$$

where $P_{X^n,Y^n}(x,y)$ is the joint type, i.e. $P_{X^n,Y^n}(x,y) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}((X_i,Y_i) = (x,y))$ where $\mathbf{1}((X_i,Y_i) = (x,y))$ is the indicator function. The expectation is with respect to the probability distribution on (X^n, W). The sequence y^n is determined by the the relation $y^n = g_n(f_n(x^n, w), w)$.

Definition 3: We say R is an *achievable rate* for a distribution $p(x,y) = p_0(x)p(y|x)$, if there exists a sequence of $(2^{nR}, n)$ distributed action codes (f_n, g_n) with $\lim_{n\to\infty} v_n(p(x,y)) = 0$. Similarly, we say $p(x,y) = p_0(x)p(y|x)$ is an *achievable distribution* with rate R, if there exists a sequence of $(2^{nR}, n)$ distributed action codes (f_n, g_n) with $\lim_{n\to\infty} v_n(p(x,y)) = 0$.

Definition 4: The *distributed action rate* for a distribution $p(x,y) = p_0(x)p(y|x)$, is the infimum over all achievable rates for $p(x,y)$.

Definition 5: The cooperation region of rate R is the set of all distributions, $p(x,y) = p_0(x)p(y|x)$ achievable with this rate.

Based on a timesharing argument we show in [5] that the cooperation region is convex.

The following theorem relates the operational definitions of the cooperation region and the distributed action rate to single letter information measures.

Theorem 1: The distributed action rate R for an i.i.d source of actions X distributed according to $p_0(x)$, and a desired distribution $p_0(x)p(y|x)$, as shown in Fig 3, is given by

$$R = I(X;Y), \tag{4}$$

where the joint distribution of the random variables (X,Y) is given by $p(x,y) = p_0(x)p(y|x)$.

Proof of Theorem 1: For achieving a joint type $P_{X^n,Y^n}(x,y)$, common randomness W is not necessary and the proof of the achievablility follows immediately from the proof of achievability in rate distortion theory [6, Sec. 10.6].

(*Converse*) A $(2^{nR}, n)$ code (f_n, g_n) and an i.i.d source X^n distributed according to $p_0(x)$ induces a joint distribution $p(x^n, y^n)$ on (X^n, Y^n) where $Y^n = g_n(f_n(X^n, W), W)$. We denote marginal probability at time i as $p_i^{(n)}(x, y) = \Pr(X_i = x, Y_i = y)$.

The proof involves two steps. Using the sequence of inequalities in the converse for rate distortion [6, (10.58)-(10.66)] we obtain

$$nR \geq \sum_{i=1}^{n} I(X_i; Y_i). \tag{5}$$

In the second step we prove that the l_1-distance between the average of the marginal probabilities, $\frac{1}{n}\sum_{i=1}^{n} p_i^{(n)}(x, y)$, and $p(x, y)$ goes to zero, i.e.

$$\lim_{n\to\infty} \sum_{x,y\in\mathcal{X}\times\mathcal{Y}} \left| \frac{1}{n}\sum_{i=1}^{n} p_i^{(n)}(x, y) - p(x, y) \right| = 0. \tag{6}$$

Based on this property, the convexity of $I(X;Y)$ in $p(y|x)$, and the continuity of $I(X;Y)$ in $p(x,y)$, it follows that an achievable rate must satisfy

$$nR \geq I(X;Y). \tag{7}$$

■

B. Chain of two agents

Here we consider the case where three nodes are connected in a chain, where the first node in the chain is a source of actions and all the other nodes are agents who choose their actions according to the communication they receive from the previous node. The network for a chain of three nodes is illustrated in Fig. 4.

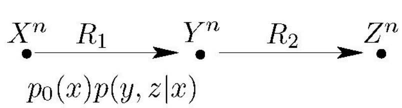


Fig. 4. Distributed action in a chain.

Definition 6: A $((2^{nR_1}, 2^{nR_2}), n)$ code for distributed action in a chain consists of two encoders,

$$f_1 : \mathcal{X}^n \times W \to 1, 2, ..., 2^{nR_1} \tag{8}$$

$$f_2 : 1, 2, ..., 2^{nR_1} \times W \to 1, 2, ..., 2^{nR_2} \tag{9}$$

and two decoders

$$g_1 : 1, 2, ..., 2^{nR_1} \times W \to \mathcal{Y}^n, \tag{10}$$

$$g_2 : 1, 2, ..., 2^{nR_2} \times W \to \mathcal{Z}^n, \tag{11}$$

where W is common randomness available at all nodes. Let

$$v_n(p(x,y,z)) \triangleq \max_{x,y,z} \mathbf{E}\left\{|P_{X^n,Y^n,Z^n}(x,y,z) - p(x,y,z)|\right\}$$

denote the maximum variation from the desired distribution $p_0(x)p(y,z|x)$ with the $((2^{nR_1}, 2^{nR_1}), n)$ code.

Theorem 2: For distributed action in a chain with an i.i.d source distributed $\sim p_0(x)$ and a desired distribution $p_0(x)p(y,z|x)$ the achievable region is given by

$$\begin{aligned} R_1 &\geq I(X;Y,Z), \\ R_2 &\geq I(X;Z). \end{aligned} \tag{12}$$

Proof: If we consider R_1 to be the rate transmitted from the source X to agents (Y, Z) then, according to Theorem 1, R_1 must be larger than $I(X;Y,Z)$ in order to achieve the desired distribution $p_0(x)p(y,z|x)$. Similarly, R_2 must be larger than $I(X;Z)$ in order to achieve the desired distribution $p_0(x)p(z|x)$.

The idea of the proof for achievability is first to cover X^n by 2^{nR_2} codewords of $Z^n(X^n)$ such that $(X^n, Z^n(X^n))$ is jointly typical with high probability and then, given that node Y knows the codeword $Z^n(X^n)$ we need $2^{n(R_1-R_2)}$ additional codewords where $R_1 - R_2 > I(X;Y|Z)$ in order to insure that (X^n, Y^n, Z^n) is jointly typical with high probability. ■

Example 1: Consider question (a) from the introduction. The joint distribution of the actions of Alice, Bob and Charlie (X, Y, Z) is the uniform distribution over all the six permutations of $\{1, 2, 3\}$ where X is specified by nature. Hence by Theorem 2 the minimum communication rate for achieving the goal is for Alice to transmit to Bob $I(X;Y,Z) = H(X) - H(X|Y,Z) = \log 3 - 0 = \log 3$ bits, and for Bob to transmit to Charlie $I(X;Z) = H(X) - H(X|Z) = \log 3 - \log 2 = \log\frac{3}{2}$ bits. So $(R_1, R_2) = (\log(3), \log(\frac{3}{2}))$ is achievable.

C. Three nodes, one rate

Consider the case where X communicates with Y at rate R as shown in Fig. 5 and there is no communication to Z. It makes sense that the set of achievable joint distributions would be $p_0(x)p(y|x)p(z)$ over all $p(y|x)$ such that $I(X;Y) \leq R$. It is true that this distribution is achievable. However a larger region is achievable.

Corollary 3: The achievable region for the distributed action problem in Fig. 5 is the set of all $p(x,y,z) = p_0(x)p(z)p(y|z,x)$ such that

$$I(X;Y|Z) \leq R. \tag{13}$$

This result can be derived directly from Theorem 2 where $R_2 = 0$, i.e. $I(X;Z) = 0$ and $R_1 = I(X;Y,Z) = I(X;Y|Z)$.

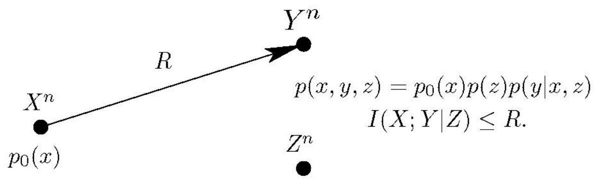


Fig. 5. Three nodes, one link: Here $p_0(x)$ is given and $X^n \sim \prod_{i=1}^{n} p_0(x_i)$ is specified at node X. Node X communicates at rate R to node Y. Node Z receives no information. A distribution $p(x,y,z) = p_0(x)p(z)p(y|x,z)$ is achievable if and only if $I(X;Y|Z) < R$.

D. Multi-agent chain

Theorem 2 can be extended to a chain with $N+1$ nodes (see Fig. 6) where only the first node X is a source of action and the rest of the nodes $(Y_1,...,Y_N)$ choose actions according to the communication between the nodes.

Corollary 4: The achievable region for the distributed action problem given in Fig. 6 with a desired distribution $p(y_1,y_2,...,y_N|x)p_0(x)$ and an i.i.d source $\sim p_0(x)$ is given by

$$R_i \geq I(X;Y_i,...,Y_N), \; i \in 1,...,N, \qquad (14)$$

where the distribution of the random variable $(X,Y_1,...,Y_N)$ is given by $p_0(x)p(y_1,y_2,...,y_N|x)$.

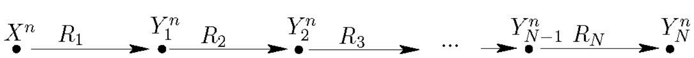


Fig. 6. Distributed action in a chain with $N+1$ nodes

E. Broadcast distributed action

This next setup generalizes problem b) from the introduction. Here we consider a distributed action problem where the source X communicates separately with agents Y, Z, and the agents do not communicate with each other. The setting is illustrated in Fig. 7 and it is related to the multiple description problem [7] [8].

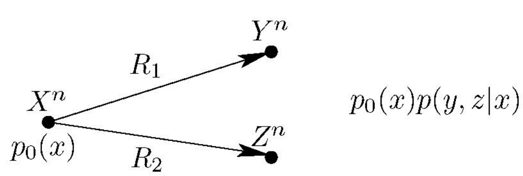


Fig. 7. Broadcast distributed action

Definition 7: A $((2^{nR_1}, 2^{nR_2}), n)$ code for broadcast distributed action consists of two encoders,

$$f_1 : \mathcal{X}^n \times W \rightarrow 1,2,...,2^{nR_1} \qquad (15)$$

$$f_2 : \mathcal{X}^n \times W \rightarrow 1,2,...,2^{nR_2} \qquad (16)$$

and two decoders

$$g_1 : 1,2,...,2^{nR_1} \times W \rightarrow \mathcal{Y}^n, \qquad (17)$$

$$g_2 : 1,2,...,2^{nR_2} \times W \rightarrow \mathcal{Z}^n. \qquad (18)$$

Theorem 5: A distribution $p_0(x)p(y,z|x)$ is achievable in the broadcast setting (Fig. 7) with rate (R_1,R_2) if the inequalities

$$\begin{aligned} R_1 &> I(X;U,Y), \\ R_2 &> I(X;U,Z), \end{aligned} \qquad (19)$$

are satisfied for some auxiliary random variable U that forms the Markov chain $Y - (X,U) - Z$, i.e. $p(u,x,y,z) = p_0(x)p(u|x)p(y|u,x)p(z|u,x)$, and satisfies $\sum_u p(u,x,y,z) = p_0(x)p(y,z|x)$.

The idea of the proof is the following. First cover X^n by $2^{n(I(X;U)+\epsilon)}$ codewords from U^n and send the appropriate codeword to both nodes, Y and Z. Given that node Y knows the codeword $U^n(X^n)$ one needs $2^{n(I(X;Y|U)+\epsilon)}$ codewords for any codeword of $U^n(X^n)$ to ensure that with high probability $(X^n, U^n(X^n), Y^n(U^n,X^n))$ is jointly typical. Finally, in order to cover both X^n, Y^n by Z^n given that all nodes know $U^n(X^n)$ one needs $2^{n(I(X,Y;Z|U)+\epsilon)}$ codewords of Z^n for any $U^n(X^n)$, and because of the Markov relation $Y-(X,U)-Z$ we have $I(X,Y;Z|U) = I(X;Z|U)$.

For some specific cases (Examples 2 and 3) we can prove optimality by using a simple lower bound that follows from the converse for two nodes in Section II-A.

Example 2: Markov chain $X-Y-Z$. The random variables X,Y,Z form the Markov chain $X-Y-Z$. We choose the auxiliary random variable $U=Z$, and therefore we have the Markov chain $Z-(Z,X)-Y$. According to Theorem 5 if R_1, R_2 satisfy

$$\begin{aligned} R_1 &\geq I(X;U,Y) = I(X;Z,Y) = I(X;Y), \\ R_2 &\geq I(X;U,Z) = I(X;Z) \end{aligned} \qquad (20)$$

then the distribution $p(x,y,z) = p_0(x)p(y|x)p(z|x)$ is achievable . According to the converse for two nodes, Theorem 1, if there is a code at rate (R_1,R_2) that achieves a distributed action distribution $p(x,y,z) = p_0(x)p(y|x)p(z|x)$ then (20) must be satisfied.

Example 3: Markov chain $Y-X-Z$: The random variables X,Y,Z form a Markov chain $Y-X-Z$. For this case U can be chosen to be null and the region given in (19) becomes

$$\begin{aligned} R_1 &> I(X;Y), \\ R_2 &> I(X;Z). \end{aligned}$$

Similar to the previous case this region is optimal.

Example 4: Question (b) from the introduction: Alice Bob and Charlie's actions X, Y, Z are distributed uniformly over the six possible permutations of $\{1, 2, 3\}$. Consider the deterministic scheme in Fig. 8. Node Y takes only the actions $1, 2$ and node Z takes only the action $2, 3$. For example, in Fig. 8 if $X = 1$ then $Y = 2$ and $Z = 3$. This scheme is asymmetric, but by using six different schemes of this form and using the random variable U for timesharing between them we can achieve a uniform distribution over all six permutations. It is sufficient that

$$\begin{aligned} R_1 &> I(X;U,Y) = I(X;Y|U) \\ &= H(Y|U) - H(Y|U,X) = H(\frac{1}{3}) - 0 = 0.918, \end{aligned}$$

and similarly, $R_2 > H(\frac{1}{3})$.

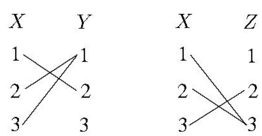


Fig. 8. The dependency of random variables X, Y, Z that achieves $R_1 = R_2 = H(\frac{1}{3})$ in Example 4.

It is possible to extend the cooperation region presented in Theorem 5 by introducing an additional auxiliary random variable V such that (U, V, X, Y, Z) forms a Markov chain $Y - (V, U, X) - Z$. For this case an achievable region for the distribution $p_0(x)p(y, z|x)$ is

$$\begin{aligned} R_1 &> I(X;V) + \min\{I(X;U,Y|V), I(X,U;Y|V)\}, \\ R_2 &> I(X;V) + \min\{I(X;U,Z|V), I(X,U;Z|V)\}. \end{aligned}$$

The proof is omitted; see [5].

III. Distributed Action and Rate Distortion

There are similarities between rate distortion problems and distributed action problems. Distributed action considers the coordination of all the nodes, while in rate distortion the pairwise distortions between source and reconstructions are considered.

In the rate distortion problem, the goal is to achieve $\lim_{n\to\infty} Ed(X^n, Y^n) \leq D$ where the distortion between sequence x^n and the reconstruction y^n is defined by $d(x^n, y^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, y_i)$. Then distribution D is achievable at rate R if there exists a distribution $p_0(x)p(y|x)$ such that

$$\sum_{x,y} p_0(x)p(y|x)d(x,y) \leq D \quad (21)$$

and $I(X;Y) \leq R$. Hence, the distortion is a linear function of the distribution. Eq. (21) defines a half space where the distributions satisfy the distortion criteria. Fig. 9 shows a typical cooperation region when $R = 0.1$ for a binary source and two nodes as defined in Section II-A. The figure also shows the corresponding half-space for Hamming distortion less then D.

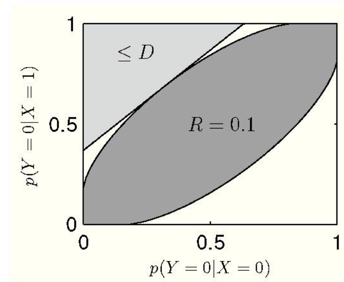


Fig. 9. Typical cooperation region for the case of two nodes. The dark region is the cooperation region for rate $R = 0.1$ and $X \sim Bernoulli(0.5)$. The shaded region defines the set of all joint distributions $p_0(x)p(y|x)$ with Hamming distortion less then D, i.e. $E(d(X^n, Y^n)) \leq D$ where D is chosen to satisfy $R(D) = 0.1$.

We know that the set $\mathcal{P}_R$ of achievable distributions $p_0(x)p(y_1, y_2, ..., |x)$ at rate R is convex. Thus these distributions can be characterized by the envelope of tangent hyperplanes

$$D(d(x, y_1, y_2, ...)) = \max_{p\in\mathcal{P}_R} \sum_{x,y_1,y_2...} p(x, y_1, y_2, ...)d(x, y_1, y_2, ...)$$

defined for every "distortion" $d(x, y_1, y_2, ...)$.

Acknowledgment

This work was supported by National Science Foundation (NSF) through the grants CCF-0635318 and CCF-0515303. The authors would like to thank Paul Cuff, Young-Han Kim and Tsachy Weissman for helpful discussions.

References

[1] V. Anantharam and V. Borkar, "Common randomness and distributed control; a counterexample," August 2005, submitted to Systems and Control Letters.

[2] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, "On quantum coding for ensembles of mixed states ," *J. of Phys. A Math. Gen.*, vol. 34, pp. 6767–6785, 2001.

[3] G. Kramer and S. A. Savari, "Communicating probability distributions," *IEEE Trans. Inform. Theory*, vol. 53, pp. 518–525, 2007.

[4] P. Cuff, "Communication requirements for generating joint distributions," submitted to ISIT 2007, Nice, France.

[5] T. M. Cover and H. H. Permuter, "Capacity of coordinated actions," Jan. 2006, in preparation.

[6] T. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New-York: New York: Wiley, 2006.

[7] A. E. Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 28, p. 851, 1982.

[8] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 33, pp. 502–521, 1987.