

## MIT Open Access Articles

*Opportunistic scheduling with limited channel state information: A rate distortion approach*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Johnston, Matthew, Eytan Modiano, and Yury Polyanskiy. "Opportunistic Scheduling with Limited Channel State Information: A Rate Distortion Approach." 2014 IEEE International Symposium on Information Theory (June 2014).

**As Published:** <http://dx.doi.org/10.1109/ISIT.2014.6875057>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <http://hdl.handle.net/1721.1/96984>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Opportunistic Scheduling with Limited Channel State Information: A Rate Distortion Approach

Matthew Johnston, Eytan Modiano, Yury Polyanskiy  
Laboratory for Information and Decision Systems  
Massachusetts Institute of Technology  
Cambridge, MA  
Email: {mrj, modiano, yp}@mit.edu

**Abstract**—We consider an opportunistic communication system in which a transmitter selects one of multiple channels over which to schedule a transmission, based on partial knowledge of the network state. We characterize a fundamental limit on the rate that channel state information must be conveyed to the transmitter in order to meet a constraint on expected throughput. This problem is modeled as a causal rate distortion optimization of a Markov source. We introduce a novel distortion metric capturing the impact of imperfect channel state information on throughput. We compute a closed-form expression for the causal information rate distortion function for the case of two channels, as well as an algorithmic upper bound on the causal rate distortion function. Finally, we characterize the gap between the causal information rate distortion and the causal entropic rate-distortion functions.

## I. INTRODUCTION

Consider a transmitter and a receiver connected by two independent channels. The state of each channel is either ON or OFF, where transmissions over an ON channel result in a unit throughput, and transmissions over an OFF channel fail. Channels evolve over time according to a Markov process. At the beginning of each time slot, the receiver measures the channel states in the current slot, and transmits (some) channel state information (CSI) to the transmitter. Based on the CSI sent by the receiver, the transmitter chooses over which of the channels to transmit.

In a system in which an ON channel and OFF channel are equally likely to occur, the transmitter can achieve an expected per-slot throughput of  $\frac{1}{2}$  without channel state information, and a per-slot throughput of  $\frac{3}{4}$  if the transmitter has full CSI before making scheduling decisions. However, the transmitter does not need to maintain complete knowledge of the channel state in order to achieve high throughput; it is sufficient to only maintain knowledge of which channel has the best state. Furthermore, the memory in the system can be used to further reduce the required CSI. We are interested in the minimum rate that CSI must be sent to the transmitter in order to guarantee a lower bound on expected throughput. This quantity represents a fundamental limit on the overhead information required in this setting.

This work was supported by NSF Grants CNS-0915988, CNS-1217048, ARO MURI Grant W911NF-08-1-0238, ONR Grant N00014-12-1-0064, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

The above minimization can be formulated as a rate distortion optimization with an appropriately designed distortion metric. The opportunistic communication framework, in contrast to traditional rate distortion, requires that the channel state information sequence be causally encoded, as the receiver causally observes the channel states. Consequently, restricting the rate distortion problem to causal encodings provides a tighter lower bound on the required CSI that must be provided to the transmitter.

Opportunistic scheduling is one of many network control schemes that requires network state information (NSI) in order to make control decisions. The performance of these schemes is directly affected by the availability and accuracy of this information. If the network state changes rapidly, there are more possibilities to take advantage of an opportunistic performance gain, albeit at the cost of additional overhead. For large networks, this overhead can become prohibitive.

This paper presents a novel rate distortion formulation to quantify the fundamental limit on the rate of overhead required for opportunistic scheduling. We design a new distortion metric for this setting that captures the impact on network performance, and incorporate a causality constraint to the rate distortion formulation to reflect practical constraints of a real-time communication system. We analytically compute a closed-form expression for the causal rate distortion lower bound for a two-channel system. Additionally, we propose a practical encoding algorithm to achieve the required throughput with limited overhead. Moreover, we show that for opportunistic scheduling, there is a fundamental gap between the mutual information and entropy-rate-based rate distortion functions, and discuss scenarios under which this gap vanishes. Proofs have been omitted for brevity.

## II. PROBLEM FORMULATION

Consider a transmitter and a receiver, connected through  $M$  independent channels. Assume a time slotted system, where at time-slot  $t$ , each channel has a time-varying channel state  $S_i(t) \in \{\text{OFF}, \text{ON}\}$ , independent from all other channels. The notation  $S_i(t) \in \{0, 1\}$  is used interchangeably.

Let  $X(t) = X_t = \{S_1(t), S_2(t), \dots, S_M(t)\}$  represent the system state at time slot  $t$ . At each time slot, the

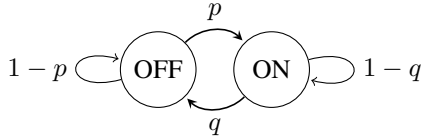


Fig. 1: Markov chain describing the channel state evolution of each independent channel.

transmitter chooses a channel over which to transmit, with the goal of opportunistically transmitting over an ON channel. Channel states evolve over time according to a Markov process described by the chain in Figure 1, with transition probabilities  $p$  and  $q$  satisfying  $p \leq \frac{1}{2}$  and  $q \leq \frac{1}{2}$ , corresponding to channels with “positive memory.”

The transmitter does not observe the state of the system. Instead, the receiver causally encodes the sequence of channel states  $X_1^n$  into the sequence  $Z_1^n$  and sends the encoded sequence to the transmitter, where  $X_1^n$  is used to denote the vector of random variables  $[X(1), \dots, X(n)]$ . The encoding  $Z(t) = Z_t \in \{1, \dots, M\}$  represents the index of the channel over which to transmit. Since the throughput-optimal transmission decision is to transmit over the channel with the best state, it is sufficient for the transmitter to restrict its knowledge to the index of the channel with the best state at each time.

The expected throughput earned in slot  $t$  is  $\mathbb{E}[\text{thpt}(t)] = S_{Z(t)}(t)$ , since the transmitter uses channel  $i = Z(t)$ , and receives a throughput of 1 if that channel is ON, and 0 otherwise. Clearly, a higher throughput is attainable with more accurate CSI, determined by the quality of the encoding  $Z_1^n$ . The average distortion between the sequences  $x_1^n$  and  $z_1^n$  is defined in terms of the per-letter distortion,

$$d(x_1^n, z_1^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, z_i), \quad (1)$$

where  $d(x_i, z_i)$  is the per-letter distortion between the  $i$ th source symbol and the  $i$ th encoded symbol at the transmitter. For the opportunistic communication framework, the per-letter distortion is defined as

$$d(x_i, z_i) \triangleq 1 - \mathbb{E}[\text{thpt}(t)] = 1 - S_{Z(t)}(t), \quad (2)$$

where  $S_{Z(t)}$  is the state of the channel indexed by  $Z(t)$ . Thus, an upper bound on expected distortion translates to a lower bound on expected throughput. Note that the traditional Hamming distortion metric is inappropriate in this setting, since the transmitter does not need to know the channel states of channels it will not transmit over.

#### A. Problem Statement

The goal in this work is to determine the minimum rate that CSI must be conveyed to the transmitter to achieve a lower bound on expected throughput. In this setting, CSI must be conveyed to the transmitter causally, in other words, the  $i$ th encoding can only depend on the channel state at time  $i$ , and previous channel states and encodings. Let  $\mathcal{Q}_c(D)$  be the family of causal encodings  $q(z_1^n|x_1^n)$

satisfying

$$\mathbb{E}[d(x_1^n, z_1^n)] = \sum_{x_1^n} \sum_{z_1^n} p(x_1^n) q(z_1^n|x_1^n) d(x_1^n, z_1^n) \leq D. \quad (3)$$

where  $p(x_1^n)$  is the PDF of the source, and the causality constraint:

$$q(z_1^i|x_1^n) = q(z_1^i|y_1^n) \quad \forall x_1^n, y_1^n \text{ s.t. } x_1^i = y_1^i, \quad (4)$$

Mathematically, the minimum rate that CSI must be transmitted is given by

$$R_c^{NG}(D) = \lim_{n \rightarrow \infty} \inf_{q \in \mathcal{Q}_c(D)} \frac{1}{n} H(Z_1^n) \quad (5)$$

where  $\frac{1}{n} H(Z_1^n)$  is the entropy rate of the encoded sequence in bits. Equation (5) is the causal rate distortion function, as defined by Neuhoff and Gilbert [1], and is denoted using the superscript *NG*. This quantity is an entropy rate distortion function, in contrast to the information rate distortion function [2], [3], [4], which will be discussed in Section III. The decision to formulate this problem as a minimization of entropy rate is based on the intuition that the entropy rate should capture the average number of bits per channel use required to convey channel state information.

#### B. Previous Work

Among the earliest theoretical works to study communication overhead in networks is Gallager’s seminal paper [5], where fundamental lower bounds on the amount of overhead needed to keep track of source and destination addresses and message starting and stopping times are derived using rate-distortion theory. A discrete-time analog of Gallager’s model is considered in [6]. A similar framework was considered in [7] and [8] for different forms of network state information.

The traditional rate-distortion problem [9] has been extended to bounds for Markov Sources in [10], [11], [12]. Additionally, researchers have considered the causal source coding problem due to its application to real-time processing. One of the first works in this field was [1], in which Neuhoff and Gilbert show that the best causal encoding of a memoryless source is a memoryless coding, or a time sharing between two memoryless codes. Neuhoff and Gilbert focus on the minimization of entropy rate, as in (5). The work in [13] studied the optimal finite-horizon sequential quantization problem, and showed that the optimal encoder for a  $k^{\text{th}}$ -order Markov source depends on the last  $k$  source symbols and the present state of the decoder’s memory (i.e. the history of decoded symbols).

A causal (sequential) rate distortion theory was introduced in [3] and [14] for stationary sources. They show that the sequential rate distortion function lower bounds the entropy rate of a causally encoded sequence, but this inequality is strict in general. Despite this, operational significance for the causal rate distortion function is developed in [3]. Lastly, [4] studies the causal rate distortion function as a minimization of directed mutual information, and computes the form of the optimal causal kernels.

### III. RATE DISTORTION LOWER BOUND

To begin, we review the traditional rate distortion problem, and define the causal *information* rate distortion function, a minimization of mutual information, which is known to lower bound  $R_c^{NG}(D)$  [14], and hence provides a lower bound on the required rate at which CSI must be conveyed to the transmitter to meet the throughput requirement.

#### A. Traditional Rate Distortion

Consider the well known rate distortion problem, in which the goal is to find the minimum number of bits per source symbol necessary to encode a source while meeting a fidelity constraint. Consider a discrete memoryless source  $\{X_i\}_{i=1}^{\infty}$ , where  $X_i$ 's are i.i.d. random variables taking values in the set  $\mathcal{X}$ , according to distribution  $p_X(x)$ . This source sequence is encoded into a sequence  $\{Z_i\}_{i=1}^{\infty}$ , with  $Z_i$  taking values in  $\mathcal{Z}$ . The distortion between a block of source symbols and encoded symbols is defined in (1) with per-letter distortion  $d(x_i, z_i)$ . Define  $\mathcal{Q}(D)$  to be the family of conditional probability distributions  $q(z|x)$  satisfying an expected distortion constraint (3).

Shannon's rate-distortion theory states that the minimum rate  $R$  at which the source can be encoded with average distortion less than  $D$  is given by the information rate distortion function  $R(D)$ , where

$$R(D) \triangleq \min_{q(z|x) \in \mathcal{Q}(D)} I(X; Z), \quad (6)$$

and  $I(\cdot; \cdot)$  represents mutual information.

#### B. Causal Rate Distortion for Opportunistic Scheduling

Consider the problem formulation in Section II. As discussed above, the *information* rate distortion is a minimization of mutual information over all stochastic kernels satisfying a distortion constraint. For opportunistic scheduling, this minimization is further constrained to include only causal kernels. Let  $\mathcal{Q}_c(D)$  be the set of all stochastic kernels  $q(z_1^n|x_1^n)$  satisfying (3) and (4). The *causal* information rate distortion function is defined as

$$R_c(D) = \lim_{n \rightarrow \infty} \inf_{q(z_1^n|x_1^n) \in \mathcal{Q}_c(D)} \frac{1}{n} I(X_1^n; Z_1^n). \quad (7)$$

The function  $R_c(D)$  is a lower bound on the Neuhoff-Gilbert rate distortion function  $R_c^{NG}(D)$  in (5), and hence a lower bound on the rate of CSI that needs to be conveyed to the transmitter to ensure expected per-slot throughput is greater than  $1 - D$ . In the traditional (non-causal) rate distortion framework, this bound is tight; however, in the causal setting this lower bound is potentially strict. Note that for memoryless sources,  $R_c(D) = R(D)$ , where  $R(D)$  is the traditional rate distortion function; however, for most memoryless sources,  $R(D) < R_c^{NG}(D)$ .

The optimization problem in (7) is solved using a geometric programming dual as in [15]. The following result gives the structure of the optimal stochastic kernel. Note that this result is also obtained in the work [4] for a similar formulation.

**Theorem 1.** *The optimal kernel  $q(z_1^n|x_1^n)$  satisfies*

$$q(z_i|z_1^{i-1}, x_1^i) = \frac{Q(z_i|z_1^{i-1}) \exp(-\lambda d(x_i, z_i))}{\sum_{z_i} Q(z_i|z_1^{i-1}) \exp(-\lambda d(x_i, z_i))} \quad (8)$$

where for all  $z_1^i$ ,  $Q(z_i|z_1^{i-1})$  and  $\lambda$  satisfy

$$1 = \sum_{x_1^n} \frac{P(x_1^n) \exp(-\sum_{i=1}^n \lambda d(x_i, z_i))}{\prod_{i=1}^n \sum_{z_i} Q(z_i|z_1^{i-1}) \exp(-\lambda d(x_i, z_i))} \quad (9)$$

Equation (9) holds for all  $z_1^n$ , and gives a system of equations from which one can solve for  $Q(z_i|z_1^{i-1})$ . Note this holds in general for any number of Markovian channels, and can be numerically solved to determine  $R_c(D)$ . Observe in (8) that  $q(z_i|z_1^{i-1}, x_1^i) = q(z_i|z_1^{i-1}, x_i)$ . In other words, the solution to the rate distortion optimization is a distribution which generates  $Z_i$  depending only on the source sequence through the current source symbol.

#### C. Analytical Solution for Two-Channel System

Consider the system in Section II with two channels ( $M = 2$ ), and a symmetric channel state Markov chain.

**Theorem 2.** *For the aforementioned system, the causal information rate distortion function is given by*

$$R_c(D) = \frac{1}{2} H_b(2p - 4pD + 2D - \frac{1}{2}) - \frac{1}{2} H_b(2D - \frac{1}{2}) \quad (10)$$

for all  $D$  satisfying  $\frac{1}{4} \leq D \leq \frac{1}{2}$ .

This result follows from evaluating (8) and (9) for a two channel system, and showing the stationarity of the optimal kernel. The information rate distortion function in (10) is a lower bound on the rate that information needs to be conveyed to the transmitter. A distortion  $D_{\min} = \frac{1}{4}$  represents a lossless encoding, since  $\frac{1}{4}$  of the time slots, both channels are OFF, and no throughput can be obtained. Additionally,  $D_{\max} = \frac{1}{2}$  corresponds to an oblivious encoder, as transmitting over an arbitrary channel requires no rate, and achieves distortion equal to  $\frac{1}{2}$ . The function  $R_c(D)$  is plotted in Figure 2 as a function of  $D$ .

### IV. HEURISTIC UPPER BOUND

In this section, we propose an algorithmic upper bound to the Neuhoff-Gilbert rate distortion function in (5). For simplicity, assume that  $p = q$ , and that  $M = 2$ , i.e. the transmitter has two symmetric channels over which to transmit. Therefore,  $X(t) \in \{00, 01, 10, 11\}$ . Observe that when  $X(t) = 11$ , no distortion is accumulated regardless of the encoding  $Z(t)$ , and a unit distortion is always accumulated when  $X(t) = 00$ . The minimum possible average distortion is  $D_{\min} = \frac{1}{4}$ , since the state of the system is 00 for a fraction  $\frac{1}{4}$  of the time.

#### A. Minimum Distortion Encoding Algorithm

Recall that a *causal* encoder  $f(\cdot)$  satisfies  $Z(t) = f(X_1^t, Z_1^{t-1})$ . Consider the following encoding policy:

$$Z(t) = \begin{cases} Z(t-1) & \text{if } X(t) = 00 \text{ or } X(t) = 11 \\ 1 & \text{if } X(t) = 10 \\ 2 & \text{if } X(t) = 01 \end{cases} \quad (11)$$

Note that  $Z(t)$  is a function of  $Z(t-1)$  and  $X(t)$ , and is therefore a causal encoding as defined in (4). The above encoding achieves expected distortion equal to  $\frac{1}{4}$ , the minimum distortion achievable. Note that the transmitter is unaware of the channel state; conveying full CSI requires additional rate at no reduction to distortion. Let  $K$  be a random variable denoting the number of time slots since the last change in the sequence  $Z(i)$ , i.e.,

$$K = \min_j \{j < i | Z(i-j) \neq Z(i-j-1)\}. \quad (12)$$

Thus, the transmitter can infer the state of the system  $K$  slots ago. Since the channel state is Markovian, the entropy rate of the sequence  $Z_1^\infty$  is expressed as

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(Z_1^n) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(Z(i) | Z_1^{i-1}) \quad (13)$$

$$= H(Z(i) | Z(i-1), K) \quad (14)$$

$$= \sum_{k=1}^{\infty} \mathbb{P}(K=k) H_b(\mathbb{P}(Z(i) \neq Z(i-1) | K=k)) \quad (15)$$

where  $H_b(\cdot)$  is the binary entropy function. Note by definition,  $Z(i-1) = Z(i-K)$  in (14). Equation (15) can be computed in terms of the transition probabilities of the Markov chain in Figure 1.

### B. Threshold-based Encoding Algorithm

In order to further reduce the rate of the encoded sequence using the encoder in (11), a higher expected distortion is required. A new algorithm is obtained by introducing a parameter  $T$ , and modifying the encoding algorithm in (11) as follows: If  $K \leq T$ , then  $Z(i) = Z(i-1)$ , and if  $K > T$ , then  $Z(i)$  is assigned according to (11). As a result, for the first  $T$  slots after the  $Z(i)$  sequence changes value, the transmitter can determine the next element of the sequence deterministically, and hence the sequence can be encoded with zero rate. After  $T$  slots, the entropy in the  $Z(i)$  process is similar to that of the original encoding algorithm. As expected, this reduction in entropy rate comes at an increase in distortion. In the first  $T$  slots after a change to  $Z(i) = 1$ , every visit to state  $X(i) = 01$  or  $X(i) = 00$  incurs a unit distortion. The accumulated distortion is equal to the number of visits to those states in an interval of  $T$  slots.

Clearly, as the parameter  $T$  increases, the entropy rate decreases, and the expected distortion increases. Consequently,  $T$  parameterizes the rate-distortion curve; however, due to the integer restriction, only a countable number of rate-distortion pairs are achievable by varying  $T$ , and time sharing is used to interpolate between these points. An example curve is shown in Figure 2. Note that as  $T$  increases, the corresponding points on the  $R(D)$  curve become more dense. Furthermore, for the region of  $R(D)$  parameterized by large  $T$ , the function  $R(D)$  is linear. The slope of this linear region is characterized by the following result.

**Proposition 1.** *Let  $R(T)$  and  $D(T)$  denote the rate and expected distortion as functions of the parameter*

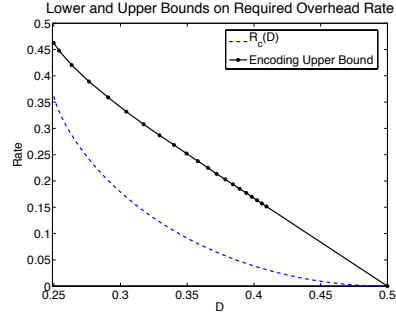


Fig. 2: Plots the causal information rate distortion function  $R_c(D)$  (Section III) and the upper bound to the rate distortion function (Section IV), computed using Monte Carlo Simulation.

$T$  respectively. For large  $T$ , the achievable  $R(D)$  curve for the above encoding algorithm, denoted by the points  $(D(T), R(T))$  has slope

$$\lim_{T \rightarrow \infty} \frac{R(T+1) - R(T)}{D(T+1) - D(T)} = \frac{-H(M)}{c + \frac{1}{4}\mathbb{E}[M]}, \quad (16)$$

where  $M$  is a random variable denoting the expected number of slots after the initial  $T$  slots until the  $Z_i$  sequence changes value, and  $c$  is a constant given by

$$c = \sum_{i=1}^T \left( \mathbb{E}[\mathbb{1}(X_i = 00 \text{ or } X_i = 01)] - \mathbb{E}[\mathbb{1}(X_i = 00 \text{ or } X_i = 01) | X_0 = 10] \right). \quad (17)$$

The constant in (17) represents the difference in expected accumulated distortion over an interval of  $T$  slots of the state processes beginning in steady state and  $X_0 = 10$ . Proposition 1 shows that the slope of  $R(D)$  is independent of  $T$  for  $T$  sufficiently large, as illustrated in Figure 2.

### V. CAUSAL RATE DISTORTION GAP

Figure 2 shows a gap between the causal information rate distortion function, and the heuristic upper bound to the Neuhoff-Gilbert rate distortion function computed in Section IV. In this section, we prove that for a class of distortion metrics including the throughput metric in (2), there exists a gap between the information and Neuhoff-Gilbert causal rate distortion functions, even at  $D = D_{\min}$ .

For example, consider a discrete memoryless source  $\{X_i\}$ , drawing i.i.d. symbols from the alphabet  $\{0, 1, 2\}$ , and an encoding sequence  $\{Z_i\}$  drawn from  $\{0, 1, 2\}$ . Consider the following distortion metrics:  $d_1(x, z) = \mathbb{1}_{z \neq x}$  and  $d_2(x, z) = \mathbb{1}_{z=x}$ , where  $\mathbb{1}$  is the indicator function. The first metric  $d_1(x, z)$  is a simple Hamming distortion measure, used to minimize probability of error, whereas the second is such that there exist two distortion-free encodings for each source symbol. The causal rate distortion functions  $R_c(D)$  for  $d_1(x, z)$  and  $d_2(x, z)$  are

$$R_1(D) = -H_b(D) - D \log \frac{2}{3} - (1-D) \log \frac{1}{3} \quad (18)$$

$$0 \leq D \leq \frac{2}{3}$$

$$R_2(D) = -H_b(D) - D \log \frac{1}{3} - (1-D) \log \frac{2}{3} \quad (19)$$

$$0 \leq D \leq \frac{1}{3}.$$

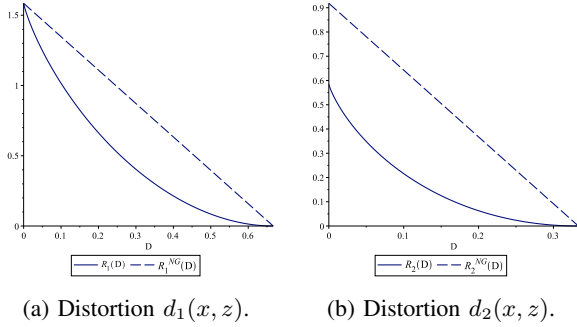


Fig. 3: Rate distortion functions for example systems.

Additionally, Neuhoff and Gilbert [1] show that for a memoryless source,  $R_c^{NG}$  equals the lower convex envelope of all memoryless encoders for this source. Thus,

$$R_{c,1}^{NG}(D) = (1 - \frac{3}{2}D) \log 3 \quad (20)$$

$$R_{c,2}^{NG}(D) = (1 - 3D)H_b(\frac{1}{3}). \quad (21)$$

The information and Neuhoff-Gilbert rate distortion functions for the two metrics are plotted in Figure 3. Note that for both distortion metrics, the causal rate distortion function is not operationally achievable. Furthermore, in the lossless encoding case ( $D = D_{\min}$ ), there is a gap in between the Neuhoff-Gilbert and information rate distortion functions when using the second distortion metric. This gap arises when for a state  $x$ , there exist multiple encodings  $z$  that can be used with no distortion penalty. This observation is formalized in the following result.

**Theorem 3.** *Let  $\{X_i\}$  represent an i.i.d. discrete memoryless source from alphabet  $\mathcal{X}$ , encoded into a sequence  $\{Z_i\}$  taken from alphabet  $\mathcal{Z}$ , subject to a per-letter distortion metric  $d(x_i, z_i)$ . Furthermore, suppose there exists  $x_1, x_2, y \in \mathcal{X}$  and  $z_1, z_2 \in \mathcal{Z}$ , such that  $z_1 \neq z_2$  and*

- $\mathbf{P}(x_1) > 0, \mathbf{P}(x_2) > 0, \mathbf{P}(y) > 0,$
- $z_1$  is the minimizer  $z_1 = \arg \min_z d(x_1, z),$
- $z_2$  is the minimizer  $z_2 = \arg \min_z d(x_2, z),$
- $d(y, z_1) = d(y, z_2) = \min_z d(y, z).$

Then  $R_c^{NG}(D_{\min}) > R_c(D_{\min}).$

*Proof:* By [1], there exists a deterministic function  $f : \mathcal{X} \rightarrow \mathcal{Z}$  such that

$$R_c^{NG}(D_{\min}) = H(f(X)) \quad (22)$$

$$\mathbb{E}[d(X, f(X))] = D_{\min} \quad (23)$$

Define a randomized encoding  $q(z|x)$ , where  $z = f(x)$  for all  $x \neq y$ , and the source symbol  $y$  is encoded randomly into  $z_1$  or  $z_2$  with equal probability. Consequently,  $H(Z|X) > 0$ , and  $H(Z) > I_q(X; Z)$  under encoding  $q(z|x)$ . Note that the new encoding also satisfies  $\mathbb{E}_q[d(X, Z)] = D_{\min}$ . To conclude,

$$\begin{aligned} R_c^{NG}(D_{\min}) &= H(f(X)) > I_q(X; Z) \\ &\geq R(D_{\min}) = R_c(D_{\min}) \end{aligned} \quad (24)$$

Theorem 3 shows that if there exists only one deter-

ministic mapping  $f : \mathcal{X} \rightarrow \mathcal{Z}$  resulting in minimum distortion, then there will be no gap between the Neuhoff-Gilbert rate distortion function and the causal information rate distortion function at  $D_{\min}$ . However, when there are multiple deterministic mappings that achieve minimum distortion, a randomized combination of them results in lower mutual information, creating a gap between the two rate distortion functions. Note that the throughput distortion metric in (2) satisfies the conditions of Theorem 3, as well as any distortion metric such that there exists a source symbol such that all per-letter encodings result in the same distortion satisfies the theorem statement.

While the above result proves that the causal information rate distortion function is not tight, it is still possible to provide an operational interpretation to  $R_c(D)$  in (10). In [3], the author proves that for a source  $\{X_{n,t}\}$ , which is Markovian across the time index  $t$ , yet i.i.d. across the spatial index  $n$ , there exist blocks of sufficiently large  $t$  and  $n$  such that the causal rate distortion function is operationally achievable, i.e. the information and Neuhoff-Gilbert rate distortion functions are equal. In the opportunistic scheduling setting, this is equivalent to a transmitter sending  $N$  messages to the receiver, where each transmission is assigned a disjoint subset of the channels over which to transmit. However, this restriction can result in a reduced throughput.

## REFERENCES

- [1] D. Neuhoff and R. Gilbert, "Causal source codes," *Information Theory, IEEE Transactions on*, vol. 28, no. 5, pp. 701–713, 1982.
- [2] A. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 12–21, 1973.
- [3] S. Tatikonda and S. Mitter, "Control under communication constraints," *Automatic Control, IEEE Transactions on*, vol. 49, no. 7, pp. 1056–1068, 2004.
- [4] P. A. Stavrou and C. D. Charalambous, "Variational equalities of directed information and applications," *arXiv preprint arXiv:1301.6520*, 2013.
- [5] R. Gallager, "Basic limits on protocol information in data communication networks," *Information Theory, IEEE Transactions on*, vol. 22, no. 4, pp. 385–398, 1976.
- [6] B. P. Dunn and J. N. Laneman, "Basic limits on protocol information in slotted communication networks," in *ISIT 2008*. IEEE, 2008, pp. 2302–2306.
- [7] A. A. Abouzeid and N. Bisnik, "Geographic protocol information and capacity deficit in mobile wireless ad hoc networks," *Information Theory, IEEE Trans. on*, vol. 57, no. 8, pp. 5133–5150, 2011.
- [8] J. Hong and V. O. Li, "Impact of information on network performance—an information-theoretic perspective," in *GLOBECOM 2009*. IEEE, 2009, pp. 1–6.
- [9] T. Berger, *Rate-Distortion Theory*. Wiley Online Library, 1971.
- [10] R. Gray, "Information rates of autoregressive processes," *Information Theory, IEEE Trans. on*, vol. 16, no. 4, pp. 412–421, 1970.
- [11] T. Berger, "Explicit bounds to  $r(d)$  for a binary symmetric markov source," *Information Theory, IEEE Transactions on*, vol. 23, no. 1, pp. 52–59, 1977.
- [12] S. Jalali and T. Weissman, "New bounds on the rate-distortion function of a binary markov source," in *ISIT 2007*. IEEE, 2007, pp. 571–575.
- [13] H. Witsenhausen, "On the structure of real-time source coders," *Bell Syst. Tech. J*, vol. 58, no. 6, pp. 1437–1451, 1979.
- [14] V. Borkar, S. Mitter, A. Sahai, and S. Tatikonda, "Sequential source coding: an optimization viewpoint," in *CDC-ECC'05*. IEEE, 2005, pp. 1035–1042.
- [15] M. Chiang and S. Boyd, "Geometric programming duals of channel capacity and rate distortion," *Information Theory, IEEE Transactions on*, vol. 50, no. 2, pp. 245–258, 2004.