Rate-Distortion Function for Finite Block Codes: Analysis of Symmetric Binary Hamming Problem

Srinivas Avasarala, Soumya Jana, Prabhakar Akella
Indian Institute of Technology
Hyderabad, India
Email: {ee13p1001,jana,akellap}@iith.ac.in

***Abstract*—Shannon's rate-distortion theory provides an asymptotic analysis, where delays are allowed to grow unbounded. In practice, real-time applications, such as video streaming and network storage, are subject to certain maximum delay. Accordingly, it is imperative to develop a finite-delay framework for analyzing the rate-distortion limit. In this backdrop, we propose an intuitive generalization of Shannon's asymptotic operational framework to finite block codes. In view of the extreme complexity of such framework, we obtain insight by specializing to the symmetric binary hamming problem. Even upon such specialization, the proposed framework is computationally so intensive that accurate evaluation of the finite-delay rate-distortion function is practical only upto a block length of three. In order to obtain further insight, we then propose a lower-complexity lower bound, based on the partition function of natural numbers, whose computation is practical upto a block length of six. Finally, using a simple combinatorial argument, we propose an upper bound to localize the desired rate-distortion function between our lower and upper bounds.**

***Keywords*—Finite block codes, Rate-Distortion function, Huffman coding, Binary source, Hamming distortion.**

I. Introduction

Since several decades, Shannon's rate-distortion (R-D) theory has been providing guidance for design of lossy source coding algorithms [1], including JPEG and MPEG [2]. However, such guidance has often been indirect, and limited to supplying the notion of R-D tradeoff, rather than concrete numerical values as reference. This happens because, while Shannon's analysis allows unbounded delay, practical applications are often sensitive to the coding delay. In fact, real-time applications come with strict delay constraints, and hence cannot directly use Shannon's R-D theorem as a theoretical basis. In fact, with the exponential growth in the use of network storage and communication, fast real-time lossless coding algorithms such as LZ4 has become indispensable [3]. However, much less effort has been directed at developing analogous real-time lossy algorithms, with vast potential for applications such as video streaming. Anticipating and with a view to facilitating the advent of such algorithms, in this paper we seek to create a benchmarking framework. In particular, we propose a plausible notion of finite-block-length (finite-delay) R-D function that is consistent with Shannon's theory in the asymptotic regime. We further analyze its complexity, which turns out to be prohibitively high, and propose lower-complexity lower and upper bounds. Finally, we study such bounds, and identify various desired improvements.

978-1-4799-6619-6/15/$31.00 © 2015 IEEE

Real-time extension of Shannon's information theory has been attempted by various authors. In the channel coding front, Sahai proposes a notion of any time capacity, indicating the ability of the encoder and the decoder to correct past errors and continuously improve upon older estimates [4]. Specifically, the decoding error probability is assumed to be an exponentially decaying function of block length. Various flavors of source and joint source-channel coding problems have also attracted substantial attention. In Table I, we provide a comparative summary of various approaches reported in the literature.

In particular, the existence and the structure of optimal real-time encoding techniques have been extensively studied. Witsenhausen took a control-theoretic viewpoint, and proved the existence of the optimal encoder for Markov sources for the finite-horizon sequential quantization problem [5]. Specifically, for any k-th order Markov source, such encoder depends on at most k previous symbols and the current state of the decoder. Later Piret studied causal sliding-block encoders, and obtained optimal schemes for both binary memoryless and binary first-order Markov sources [6]. Subsequently, Walrand and Varaiya studied a problem related to the Witsenhausen problem, where first-order Markov sources are communicated over noisy channels using encoders with finite as well as unbounded memory [7]. Neuhoff and Gilbert investigated causal sequential codes as well as causal block codes with unbounded block lengths [8]. In the process, they define optimum performance theoretically attainable function, analogous to Shannon's operational rate-distortion function. Borkar *et al.* formulated a stochastic control problem equivalent to the sequential vector quantization of a Markov source, and employs Markov decision theory for designing an encoder with finite memory [9]. Using large-deviations theory, source coding exponents for zero-delay finite-memory coding of memoryless sources were obtained by Merhav and Kontoyiannis [10]. Mahajan and Teneketzis also analyzed a problem closely related to those due to Witsenhausen, and to Walrand and Varaiya [11]. Specifically, real-time communication of a Markov source over noisy channel is performed with the help of finite encoder memory. Here they formulate a control-theoretic problem that jointly optimizes the encoder, the decoder as well as the memory update strategy. Optimal source coders with knowledge of finite number of future source symbols (lookahead), available only to the encoder, has been studied by Asnani and Weissman [12]. Lower bounds on the distortion in case of lossy source coding are derived by Leibowitz and Zamir [13]. Most of their effort is directed towards generalization of Shannon's mutual information as a rate measure so as to retain certain desirable properties.

	Encoding framework	Source model	Causality	Feedback	Encoding	Encoder memory
Witsenhausen [5]	source	Markov	causal	yes	sequential	finite
Piret [6]	source	Markov/memoryless	causal	yes	sequential	finite
Neuhoff and Gilbert [8]	source	memoryless	causal	no	sequential/block	unbounded
Walrand and Varaiya [7]	source/channel	Markov	causal	yes	sequential	finite/unbounded
Borkar *et al.* [9]	source	Markov	causal	no	sequential	unbounded
Merhav and Kontoyiannis [10]	source	memoryless	causal	no	sequential	finite
Mahajan and Teneketsis [11]	source/channel	Markov	causal	no	sequential	finite
Asnani and Weissman [12]	source/channel	memoryless	limited lookahead	yes/no	sequential	unbounded
Leibowitz and Zamir [13]	source/channel	memoryless	block causal	no	block	zero
Proposed	**source**	**memoryless**	**block causal**	**no**	**block**	**zero**

TABLE I: Summary of literature survey on real-time (finite-delay) source and source-channel coding.

However, such rate measure is not operationally motivated. Recently, Kostina and Verdu [17] derived acheivability bounds for finite block lengths; however, their distortion measure is not deterministic as they use excess distortion probability as the criterion. Pilc [18] obtained lower and upper bounds for rate-distortion function for finite alphabet source, asymptotically for large n based on the error exponent [18].

In a nutshell, three main approaches have so far been taken towards real-time source and source-channel coding problems. In one, the information-theoretic problem is mapped to a control/decision problem, and the tools and the knowledge from latter are exploited[[5], [7], [11]]. Secondly, search for new functionals that still exhibit desirable properties of known information-theoretic quantities is carried out [13]. Finally, rate-distortion function for finite block lengths is investigated in [18], [17]. In contrast, limited effort has been directed at extending Shannon's operational approach to the real-time scenario. In this paper, we attempt to fill this gap. Specifically, we propose an operational framework that seamlessly extends Shannon's to finite block lengths. We refine the usual measure of rate as log-cardinality to Hufman code rate, which is more accurate in the finite horizon scenario. In the face of extreme complexity, we explore the underlying combinatorial problem for the symmetric binary hamming specialization. Even upon such specialization, the proposed framework is computationally so intensive that accurate evaluation of the finite-delay rate-distortion function is practical only upto a block length of three. In order to obtain further insight, we then propose a lower-complexity lower bound, based on the partition function of natural numbers, whose computation is practical upto a block length of six. Finally, we propose a simple upper bound to localize the desired rate-distortion function between our lower and upper bounds. The proposed bounds are based on combinatorial properties.

This paper is organized as follows. In Section II, we revisit Shannon theory, and recast Shannon's formulation so as to allow seamless transition to a consistent real-time framework. Building on such formulation, we make certain refinements based on Huffman coding, and propose a real-time definition of achievability in Section III. In Section IV, we anticipate high complexity, and specialize to the symmetric binary hamming problem. In such special case, we provide an in-depth analysis of the proposed finite-delay R-D function, and propose lower and upper bounds. Simulation results are given in Section V. Finally, Section VI concludes the paper.

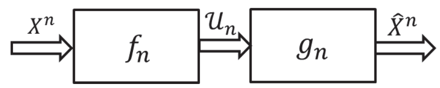


Fig. 1: Point-to-point source coding.

II. SHANNON THEORY REVISITED

Consider *i.i.d.* copies $X_1, X_2, ..., X_n, ...$ of random variable $X \sim p_X$ taking values in discrete alphabet $\mathcal{X}$. As depicted in Figure 1, encode $X^n = (X_1, X_2, ..., X_n)$ using encoder mapping $f_n : \mathcal{X}^n \to \mathcal{U}_n$, where $\mathcal{U}_n$ indicates the set of encoded indices, and decode the resulting encoded index using decoder mapping $g_n : \mathcal{U}_n \to \mathcal{X}^n$ under a bounded distortion measure $d : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ ($d_{\max} < \infty$). In this framework, Shannon's celebrated rate-distortion theorem was enunciated based on a notion of ϵ-achievability.

Definition 2.1: A rate-distortion pair (R, D) is said to be achievable if, for any $\epsilon > 0$, there exist sufficiently large n and mapping pair (f_n, g_n) such that

$$\frac{1}{n}\log|\mathcal{U}_n| \quad \leq \quad R + \epsilon \qquad (1)$$

$$\frac{1}{n}\sum_{k=1}^{n} E(d(X_k, \hat{X}_k)) \quad \leq \quad D + \epsilon, \qquad (2)$$

where $\hat{X}^n = g_n(f_n(X^n)) = (\hat{X}_1, \hat{X}_2, ..., \hat{X}_n)$. Further, the achievable set $\breve{\mathcal{A}}$ is defined by the closure of set of achievable (R, D) pairs.

Theorem 2.2: [1] The asymptotic rate-distortion (R-D) function (i.e., the lower envelop of $\breve{\mathcal{A}}$) is given by $R^*(D) = \min I(X; Z)$, where $(X, Z) \sim p_X(x)p_{Z|X}(z|x)$ such that $E(d(X, Z)) \leq D$, and the minimization is over $p_{Z|X}$'s.

In the above, $I(\cdot;\cdot)$ indicates the mutual information. Shannon's notion of ϵ-achievability provides the basis for his asymptotic Theorem 2.2, but in its current form neither (i) finitely bounds the block length n, nor (ii) provides a straightforward translation to finite block length (delay) scenarios. Accordingly, we now turn to rewriting Definition 2.1 to make such translation seamless.

Definition 2.3: $\breve{\mathcal{A}}_n^{(\epsilon)} := \text{conv}(\{(R, D) : \exists(f_n, g_n)$ so that $\frac{1}{n}\log|\mathcal{U}_n| \leq R + \epsilon, \frac{1}{n}\sum_{k=1}^{k=n} E(d(X_k, \hat{X}_k)) \leq D + \epsilon\})$, where "conv$(S)$" indicates the convex hull of set S.

Proposition 2.4: $\breve{\mathcal{A}} = \overline{\bigcup_{n\geq 1} \breve{A}_n}$, where $\breve{A}_n = \breve{A}_n^{(0)}$.

Sketch of proof: We first claim

$$\breve{\mathcal{A}} = \overline{\bigcap_{\epsilon>0} \bigcup_{n\geq 1} \breve{A}_n^{(\epsilon)}}. \tag{3}$$

Ignoring "conv$(\cdot)$" in Definition 2.3 temporarily, (3) follows from Definition 2.1 in a straightforward manner. Further, since $\breve{\mathcal{A}}$ is known to be convex (by a time sharing argument [14]), (3) continues to hold, even when we consider "conv$(\cdot)$". Next, it can be verified that $\bigcup_{n=1}^{n_0} \breve{A}_n^{(\epsilon)}$, viewed as doubly indexed by (ϵ, n_0), is uniformly convergent as $\epsilon \to 0$ and $n_0 \to \infty$ [15]. Accordingly, it is permitted to interchange the order of the intersection and the union in (3). Finally, the proof is completed via a continuity argument. □

III. Proposed Real-Time Framework

Notice that Proposition 2.4 describes $\breve{\mathcal{A}}$ in terms of increasing block-length n but does not require ϵ. Consequently, one may truncate at a finite $n = l$, and meaningfully use a quantity similar to $\breve{\mathcal{A}}_{\leq l} := \bigcup_{n=1}^{l} \breve{\mathcal{A}}_n$ as the achievable set upto delay l. However, we still require certain refinements. The first refinement involves the rate measure when the delay is finite. In (1), $\log |\mathcal{U}_n|$ measures the rate required to encode $f_n(X^n)$. This, while asymptotically tight as $n \to \infty$, can be tightened for finite n to $R_{\text{Huff}}(f_n(X^n))$, the rate when Huffman (the optimal) code is used. Specifically, we now define $\tilde{\mathcal{A}}_n$ by thus refining Definition 2.3 of $\breve{\mathcal{A}}_n$ (i.e., $\breve{\mathcal{A}}_n^{(\epsilon)}$ at $\epsilon = 0$).

Definition 3.1: $\tilde{\mathcal{A}}_n :=$ conv$(\{(R, D) : \exists (f_n, g_n)$ so that $R_{\text{Huff}}(f_n(X^n)) \leq R, \frac{1}{n}\sum_{k=1}^{k=n} E(d(X_k, \hat{X}_k)) \leq D\})$.

We call $\tilde{\mathcal{A}}_n$ the achievable set *with* delay n. Clearly, the lower envelop of the set $\tilde{\mathcal{A}}_n$ completely specifies it. Now, each extreme point of $\tilde{\mathcal{A}}_n$ corresponds to some mapping pair (f_n, g_n). Non-extreme points on the lower envelop of $\tilde{\mathcal{A}}_n$ are obtained via suitable time sharing [14]. Interestingly, in Definition 3.1, one may restrict g_n to be one-to-one without loss of generality. To see this, consider an encoder mapping f_n, which completely specifies the minimum achievable rate. So, for any many-to-one decoder mapping g_n, clearly, one can instead find some one-to-one mapping g_n' such that the average distortion does not increase.

Definition 3.2: $\tilde{\mathcal{A}}_{\leq l} :=$ conv$\left(\bigcup_{n=1}^{l} \tilde{\mathcal{A}}_n\right)$.

Here convexification is meaningful because, while each $\tilde{\mathcal{A}}_n$ is convex, their union need not be.

Property 3.3: $\overline{\lim_{l\to\infty} \tilde{\mathcal{A}}_{\leq l}} = \breve{\mathcal{A}}$.

We leave to the reader the proof, which can be completed using the property that Huffman coding achieves a rate within one bit of entropy [14]. Here we call $\tilde{\mathcal{A}}_{\leq l}$ the achievable set *upto* delay l. This nomenclature is justified by Property 3.3, i.e., as delay l increases, $\tilde{\mathcal{A}}_{\leq l}$ approaches (upto closure) Shannon's asymptotic achievable set $\breve{\mathcal{A}}$.

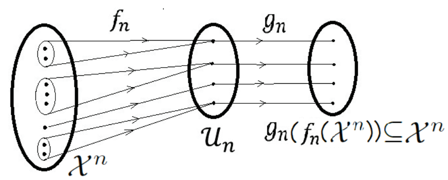


Fig. 2: Encoder and decoder mappings.

To avoid technical complications, we shall adopt compact versions of the proposed sets that preserve the lower envelop. Clearly, one incurs distortion $D = D_0 = \min_{\hat{x}\in\mathcal{X}} E(d(X, \hat{x}))$, when rate $R = 0$. Further, rate $R = \frac{1}{n}R_{\text{Huff}}(X^n)$ allows lossless reconstruction, i.e., $D = 0$. Thus, assuming monotonicity, the lower envelop of $\tilde{\mathcal{A}}_n$ always remains inside the box $\mathcal{B}_n = [0, \frac{1}{n}R_{\text{Huff}}(X^n)] \times [0, D_0]$.

Definition 3.4: $\mathcal{A}_n := \tilde{\mathcal{A}}_n \cap \mathcal{B}_n$; $\mathcal{A}_{\leq l} :=$ conv$\left(\bigcup_{n=1}^{l} \mathcal{A}_n\right)$.

Of course, the lower envelop of $\tilde{\mathcal{A}}_{\leq l}$ coincides with that of $\mathcal{A}_{\leq l}$. Henceforth, we shall use the compactified versions $\mathcal{A}_n$ and $\mathcal{A}_{\leq l}$, respectively, instead of $\tilde{\mathcal{A}}_n$ and $\mathcal{A}_{\leq l}$, which we retire at this point. Note that $\mathcal{A}_n$ is not necessarily increasing in n; however, $\mathcal{A}_{\leq l}$ is always increasing in l. Our main goal is to study the lower envelop of $\mathcal{A}_{\leq l}$, i.e., the R-D function *upto* delay l.

IV. Symmetric Binary Hamming Specialization

At this point, let us revisit Figure 2, and look closely at the source coding mechanism. Any (possibly many-to-one) encoder mapping f_n divides $\mathcal{X}^n$ into an exhaustive collection of disjoint subsets, and assigns to each such subset an index. Set of such indices is called $\mathcal{U}_n$. Finally, the decoder mapping g_n, which, as earlier argued, can be assumed one-to-one without loss of generality, assigns to each such index a sequence from $\mathcal{X}^n$. In other words, the composite mapping $g_n \circ f_n$ assigns to each aforementioned subset a representative from $\mathcal{X}^n$. Note that f_n completely determines vector of probabilities of occurrence of the associated subsets, and hence the corresponding rate of Huffman coding. Hence, given f_n, the optimal g_n would chose the aforementioned representative sequence of each subset such that the expected distortion due to such representation is minimized. In this setting, it appears unlikely that lower envelop of $\mathcal{A}_{\leq l}$ with finite l admits a general characterization analogous to that given in Shannon's Theorem 2.2. Accordingly, to obtain insight, we shall specialize to the symmetric binary hamming problem, and obtain lower and upper bounds.

A. R-D function

In particular, consider binary alphabet $\mathcal{X} = \{0, 1\}$, symmetric distribution $p_X = (\frac{1}{2}, \frac{1}{2})$ (i.e., equally likely symbols), and hamming distortion measure $d(x, \hat{x}) = \begin{cases} 0, & \forall x = \hat{x} \\ 1, & \forall x \neq \hat{x}, \end{cases}$ $(x, \hat{x}) \in \{0, 1\} \times \{0, 1\}$. Interestingly, for the symmetric binary X, we have $\frac{1}{n}R_{\text{Huff}}(X^n) = R_{\text{Huff}}(X) = 1$ and $D_0 = 0.5$, i.e., $\mathcal{B}^n = [0, 1] \times [0, 0.5]$ for any n. Now we turn to constructing $\mathcal{A}_n$ for arbitrary n, i.e., equivalently, the R-D function with delay n. Note that $\mathcal{X}^n$ consists of 2^n binary sequences of length n, each occurring with probability $\frac{1}{2^n}$. Consequently, the occurrence probability vector associated with encoder subsets depends only on the cardinalities of such subsets. Further, since Huffman coding is agnostic to permutation, the rate is completely determined by the corresponding partition of 2^n. However, while computing the expected distortion, one needs to consider specific sequences in each subset into account, i.e., take each of the $(2^n)^{(2^n)}$ encoder mappings f_n into account. Huffman coding of $f_n(\mathcal{X}^n)$ is of complexity $O(2^n \times \log(2^n))$ since the probability vector might be unsorted. For each such f_n, the optimal decoder mapping g_n is chosen based on a table of incurred distortions between pairs of sequences from $\mathcal{X}^n$. Clearly, this table has $2^n.2^n = 2^{2n}$ entries, and needs to be computed only once; however, to optimize g_n, one requires to consider each entry of the table. Thus the encoding and decoding process entails a complexity of $O((2^n)^{(2^n)} 2^{2n} 2^n \log(2^n)) = O(n 2^{n(2^n+3)})$. Hence the complexity of constructing $\mathcal{A}_{\leq l}$ is bounded by $O(l^2 2^{l(2^l+3)})$. In view of the extremely high complexity, we seek low-complexity lower and upper bounds on the corresponding finite-delay R-D functions.

B. Lower bound

First we propose a lower bound. Towards this, consider arbitrary subset $G \subseteq \mathcal{X}^n$ which is mapped by $g_n \circ f_n$ to a representative sequence $\hat{x}^n \in \mathcal{X}^n$ (see Fig. 2). The minimum expected distortion contributed by G is given by

$$d^n(G) = \frac{1}{n2^n} \sum_{x^n \in G} \sum_{k=1}^{n} d(x(k), \hat{x}(k)), \tag{4}$$

where $\hat{x}^n$ is the optimal representative that minimizes the right hand side.

Lemma 4.1: For any $G \subseteq \mathcal{X}^n$,

$$d^n(G) \geq \frac{1}{n.2^n} \sum_{m=0}^{n} m.k_m, \tag{5}$$

where $k_0 = 1$, $k_m = \max\left(0, \min\left(\binom{n}{m}, r - \sum_{i=0}^{m-1} k_i\right)\right)$, $m = 1, 2, ..., n$, and $r = |G| > 0$. Further, for any $0 < r \leq 2^n$ and any sequence $\hat{x}^n \in \mathcal{X}^n$, there exists G', with $|G'| = r$ and optimal representative $\hat{x}^n$, which satisfies (5) with equality for $G = G'$.

Proof: We shall first prove the second part of the assertion. Towards this, construct a subset $G'' \subseteq \mathcal{X}^n$ with $|G''| = r$ as follows. Arrange all sequences such that a sequence with lower hamming weight precedes one with higher weight; the order in which the sequences with the same weight is arranged is immaterial. Of course, there are $\binom{n}{m}$ sequences of weight m, $m = 0, 1, ..., n$. Now, include the first r sequences in G''. Clearly, G'' contains k_m sequences of weight m, $m = 0, 1, ..., n$. Hence, taking the all-zero sequence as the representative, one incurs the expected distortion given by the right hand side (RHS) of (5). One can verify that the all-zero sequence is a possibly nonunique, but optimal, representative. Correspondingly, we have $d(G'')$ given by the RHS of (5). Next, obtain subset G' by adding $\hat{x}^n$ modulo two to each sequence in G'' symbolwise. Clearly, such G' has $\hat{x}^n$ as the optimal representative, with $d(G')$ given by equality in (5).

To prove the first part, denote by $\hat{x}^n$ the optimal representative of G. Further, construct G' with $|G'| = |G| = r$, as above. Clearly, one can transform G' to G by a series of operations, each of which replaces a sequence $x'^n \in G' \setminus G$ by a sequence $x^n \in G \setminus G'$. However, by construction, $d(x'^n, \hat{x}^n) \leq d(x^n, \hat{x}^n)$ for any such (x'^n, x^n) pair. This proves (5). □

Denoting $\delta^n(r) = \min_{|G|=r} d^n(G)$, by Lemma 4.1, $\delta^n(r)$ is given by the RHS of (5). Hence we immediately obtain:

Theorem 4.2: Consider encoder mapping f_n partitioning $\mathcal{X}^n$ into k disjoint subsets $G_1, G_2, ...G_k$ with respective cardinalities $r_1, r_2, ...r_k$. No decoder mapping g_n achieves expected distortion lower than $D_L = \sum_{i=1}^{k} \delta^n(r_i)$.

In Theorem 4.2, apart from the said lower bound D_L on distortion D, the rate R is found by Huffman coding the probability vector $(\frac{r_1}{2^n}, \frac{r_2}{2^n}, ..., \frac{r_k}{2^n})$. We vary f_n over all possible choices, construct set $\mathcal{A}_n^L$ of resulting (R, D_L) pairs, and propose its lower convex envelop as a lower bound on the R-D function with delay n. Clearly, only the collection $\{r_1, r_2, ..., r_k\}$, which is a partition of 2^n, determines the rate. However, this bound is not necessarily tight, as simultaneous achievement of all $\delta^n(r_i)$, $i = 1, 2, ..., k$, is not guaranteed.

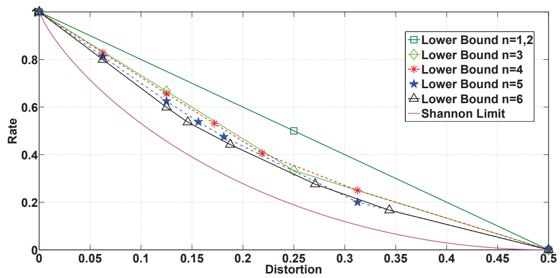
Clearly, we require $O(P(2^n))$ operations to compute all such partitions, where $P(N)$ denotes the number of partitions of integer N. One can generate such partitions such that $r_1 \geq r_2 \geq ... \geq r_k$, and divide each r_i by 2^n to obtain the probability vector for Huffman encoding, which has linear complexity for ordered probability masses. This entails a worst case complexity of $O(2^n)$, as $k \leq 2^n$. For each partition, one also computes distortion, requiring one table lookup for each subset, which is also $O(2^n)$ as above and the evaluation of distortion between two sequences of length n has a complexity of $O(n)$. Thus, using the approximation $P(N) \approx \frac{e^{\sqrt{N}}}{N}$ [16], and setting $N = 2^n$, generation of partitions is $O(2^{(2^{\frac{n}{2}} - n)})$. As a result, the complexity of lower bound computation with delay n is at most $O(2^{(2^{\frac{n}{2}} - n)} \times 2^n \times n) = O(2^{(2^{\frac{n}{2}})}.n)$. Also, the lower convex envelop of $\bigcup_{1 \leq n \leq l} \mathcal{A}_n^L$ provides a lower bound on the R-D function upto delay l, and its complexity is at most $O(l^2 2^{(2^{\frac{l}{2}})})$.

C. Upper bound

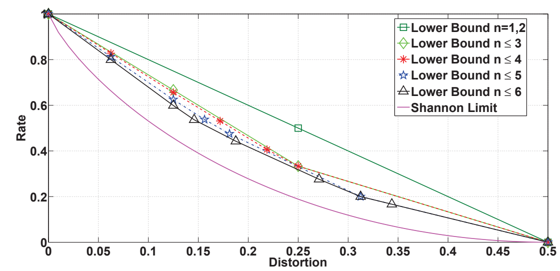
Turning to upper bound, notice that we need the lower convex envelop of a set of achievable (R, D) pairs. Of course, we have seen that irrespective of n, $(R, D) = (1, 0)$ and $(0, 0.5)$ are both trivially achievable. To obtain a meaningful upper bound, we require an achievable point that depends on n, and lies below the line joining the above two points.

Theorem 4.3: The pair $(R, D) = (\frac{1}{n}, 2\delta^n(2^{n-1}))$ is achievable with delay n.

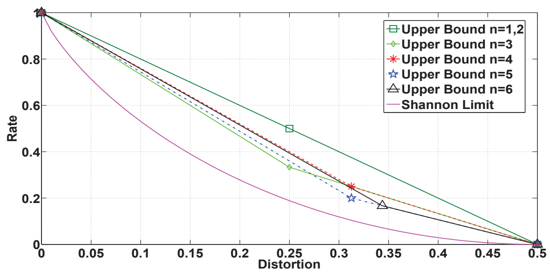
Proof: First consider odd n, and collect all sequences of hamming weight less than or equal to $\frac{n-1}{2}$ into subset G_1. Referring to the proof of Lemma 4.1, G_1 has the form of



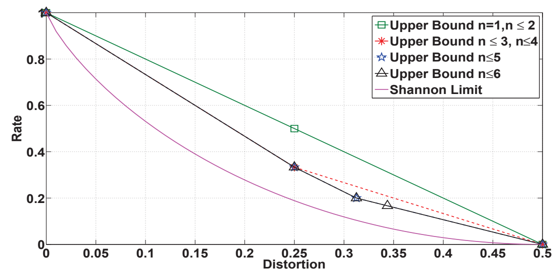
(a) Lower bound on R-D function with delay n.



(b) Lower bound on R-D function upto delay n.



(c) Upper bound on R-D function with delay n.



(d) Upper bound on R-D function upto delay n.

Fig. 3: Bounds on finite-delay R-D function.

G'', and achieves minimum distortion $\delta^n(2^{n-1})$. Similarly, the subset $G_2 = \mathcal{X}^n \setminus G_1$, consisting of all sequences with weight greater than or equal to $\frac{n+1}{2}$, also achieves distortion $\delta^n(2^{n-1})$. Further, G_1 and G_2 are equally likely, i.e., lead to probability vector $(\frac{1}{2}, \frac{1}{2})$, which needs one bit for Huffman coding. Thus one achieves a per-symbol rate of $R = \frac{1}{n}$, and distortion $D = 2\delta^n(2^{n-1})$, as stated. For even n, populate G_1 with all sequences with hamming weight upto $\frac{n}{2} - 1$ and half of those with weight equal to $\frac{n}{2}$, and construct G_2 as before. As earlier, this construction demonstrates the achievability of the said (R, D) pair. $\square$

Now define $\mathcal{A}_n^U = \{(1,0),(0,0.5),(\frac{1}{n}, 2\delta^n(2^{n-1}))\}$. Clearly, the R-D function with delay n is upper bounded by the lower convex envelop of $\mathcal{A}_n^U$. With some effort, one can obtain the closed-form expression

$$\delta^n(2^{n-1}) = \begin{cases} \frac{1}{4}(1 - \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}}), & n \text{ odd} \\ \frac{1}{4}(1 - \frac{1}{2^n}\binom{n}{\frac{n}{2}}), & n \text{ even.} \end{cases} \quad (6)$$

Since all functions involved are known to be $O(n)$, so is the complexity of the present upper bound. Further, the R-D function upto delay l is upper bounded by the lower convex envelop of $\bigcup_{1 \le n \le l} \mathcal{A}_n^U$, whose complexity is at most $O(l^2)$. To help comparative appreciation of complexities involved in computing finite-delay $R - D$ functions and the proposed lower and upper bounds, we provide the required number of operations upto a multiplicative factor in Table II. Direct computation is impractical beyond delays greater than three, while lower bounds can be computed till delays upto six. Next we turn to comparing the effectiveness of the proposed bounds.

	Exact RD function		Lower bound		Upper bound	
	delay					
m	$= m$	$\le m$	$= m$	$\le m$	$= m$	$\le m$
1	32	32	4	4	1	1
2	32768	32800	40	44	2	3
3	2.5770E+10	2.5770E+10	528	572	3	6
4	3.0223E+23	3.0223E+23	14784	15356	4	10
5	2.3945E+53	2.3945E+53	1335840	1351196	5	15
6	6.1974E+121	6.1974E+121	6.66E+8	6.66E+8	6	21

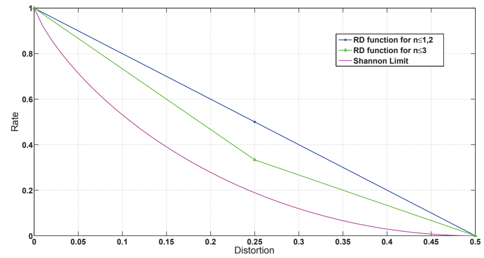
TABLE II: Comparing number (upto a factor) of operations required for computing finite-delay R-D functions and bounds.

V. Simulation Results

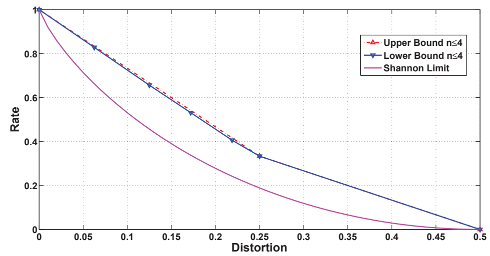
The lower and upper bounds on R-D function with delay n as well as upto delay n, $n = 1, 2, ..., 6$, are plotted in Figure 3. As a reference we also plot Shannon's (asymptotic) R-D function $R(D) = 1 + D\log D + (1 - D)\log(1 - D)$). Note that either bound for delay n is not monotone in n, as expected, whereas each such bound upto delay n is monotone.

Next we plot both the lower and upper bounds upto delay n on the same graph in Figure 4. Notice that both bounds coincide for delays upto $n \le 1, 2, 3$. While for $n \le 1, 2$, both bounds are the same trivial straight line joining $(0, 1)$ and $(0.5, 0)$, we obtain a nontrivial function for $n \le 3$. Of course, as both bounds coincide, the resulting function is the exact R-D function, which we also verify by exhaustive enumeration. For $n \le 4, 5, 6$ the bounds do not coincide, and the gap between the bounds increase with n, which appeals to intuition. However, as shown in Table II, it is impractical to evaluate the exact R-D function for these n's.

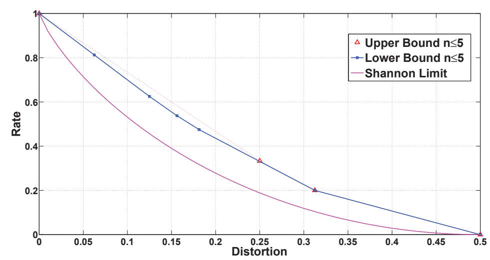
Clearly, computation of even the lower bound for larger n becomes prohibitively expensive. A lower bound, ideally with



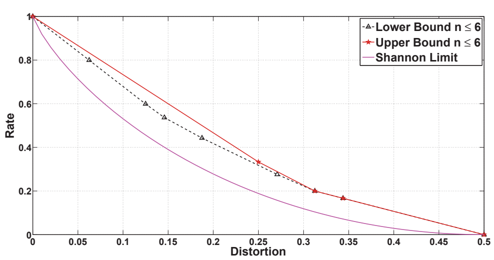
(a) $n \leq 1, 2, 3$.



(b) $n \leq 4$.



(c) $n \leq 5$.



(d) $n \leq 6$.

Fig. 4: Upper and lower bounds upto delay n.

polynomial complexity, is desirable. Further, as n increases, the cumulative upper bound only adds extreme points to the right of existing ones, while the corresponding lower bound adds extreme points more uniformly. In other words, while it remains unclear whethere the lower bound is asymptotically tight, the upper bound is clearly not asymptotically tight. Thus, while the low complexity of the upper bound is appealing, it may not be useful as n grows.

VI. Discussion

In this paper, we attempt at creating a framework for real-time R-D theory that seamlessly extends Shannon's asymptotic framework. To gain insight we specialize to the symmetric binary hamming problem, and propose lower and upper bounds based on intuitive combinatorial results. We find that the exact computation of the proposed R-D function is impractical. Further, for large n, our lower bound is too complex, and our upper bound too inaccurate to be of practical use. Accordingly, in future, we require a lower-complexity lower as well as a higher-accuracy upper bound. We believe that combinatorics and algebraic coding theory would play an important role in such analysis.

Acknowledgement

We are grateful to N. R. Aravind for useful discussion.

References

[1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion", *IRE Nat. Conv. Rec.*, pt. 4, pp. 142-163, Mar. 1959.

[2] Rafael C. Gonzalez, *Digital Image Processing*, Pearson Prentice Hall, 3rd edition, 2008.

[3] Y. Collet, "LZ4: Extremely fast compression algorithm," code.google.com, 2013.

[4] A. Sahai, *Anytime Information Theory*, Ph.D dissertation, Apr. 2001, Massachusetts Institute of Technology.

[5] H. S. Witsenhausen, "On the structure of real-time source coders", *Bell Syst. Technical Journal,* vol. 58, no. 6, Aug. 1978.

[6] P. Piret, "Causal sliding block encoders with feedback", *IEEE Trans. Inf. Theory*, vol. 25, no. 2, Mar 1979.

[7] J. C. Walrand and P. Varaiya, "Optimal causal coding-decoding problems", *IEEE Trans. Inf. Theory*, vol. 29, no. 6, Nov. 1983.

[8] D. L. Neuhoff and R. K. Gilbert, "Causal source codes", *IEEE Trans. Inf. Theory*, vol. 28, no. 5, Sept 1982.

[9] V. S. Borkar, S. K. Mitter and S. Tatikonda, "Optimal sequential vector quantization of Markov sources", *SIAM J. Contr. Optim* , vol. 40, no.1, pp. 135-148, Jan. 2001.

[10] N. Merhav and I. Kontoyiannis, "Source coding exponents for zero-delay coding with finite memory", *IEEE Trans. Inf. Theory*, vol. 49, no. 3, Mar. 2003.

[11] A. Mahajan and D. Teneketzis, "A decision theoretic framework for real-time communication", *IEEE Trans. Inf. Theory,* vol. 55, no. 11, Nov. 2009.

[12] H. Asnani and T. Weissman, "Real-time coding with limited Lookahead", *IEEE Trans. Inf. Theory*, vol. 40, no. 6, Jun. 2013.

[13] I. Leibowitz and R. Zamir, "A Ziv-Zakai-Renyi lower bound on distortion at high resolution", *Proc. Inf. Theory Workshop,* Porto, Portugal, May 2008.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Publications, 2nd edition, 1991.

[15] S. Jana and R. E. Blahut, "Canonical description for multiterminal source coding", *Intl. Symp. on Inf. Theory*, Toronto, Canada, Jul. 2008.

[16] A. Maroti, "On elementary bounds for partition function", *Integers: Elect. J. of Combinatorial Number Theory*, vol. 3, no. A10, 2003.

[17] V.Kostina and S.Verdu, "Fixed-length lossy compression in the finite blocklength regime: discrete memory less sources", *IEEE Intl. Symp. on Inf. Theory*, Feb. 2011.

[18] R.Pilc, *Coding theorems for a discrete source channel pairs*, Ph.D Thesis , M.I.T., 1967.