On Locally Decodable Source Coding

Ali Makhdoumi, Shao-Lun Huang, Muriel Médard, Yury Polyanskiy
Massachusetts Institute of Technology,
Email: {makhdoum, shaolun, medard, yp}@mit.edu

Abstract—With the boom of big data, traditional source coding techniques face the common obstacle to decode only a small portion of information efficiently. In this paper, we aim to resolve this difficulty by introducing a specific type of source coding scheme called locally decodable source coding (LDSC). Rigorously, LDSC is capable of recovering an arbitrary bit of the unencoded message from its encoded version, by only feeding a small number of the encoded message to the decoder, and we call the decoder \( t \)-local if only \( t \) encoded symbols are required. We consider both almost lossless (block error) and lossy (bit error) cases for LDSC. First, we show that using linear encoder and a decoder with bounded locality, the reliable compress rate can not be less than one. More importantly, we show that even with a general encoder and 2-local decoders \((t = 2)\), the rate of LDSC is still one. On the contrary, the achievability bounds for almost lossless and lossy compressions with excess distortion suggest that optimal compression rate is achievable when \( O(\log n) \) encoded symbols is queried by the decoder with block-length \( n \). We also show that, rate distortion is achievable when the number of query is scaled over \( n \) with a bound on the rate in finite-length regime. Although the achievability bounds are simply based on the concatenation of code blocks, they outperform the existing bounds in succinct data structures literature.

I. INTRODUCTION

A. Motivation

The basic communication problem may be expressed as transmitting source symbols with the highest fidelity without exceeding a given bit rate, or expressed as transmitting the source symbols using the lowest bit rate possible while maintaining a given reproduction fidelity \([1]\). In either case, a fundamental trade-off is made between bit rate and distortion/error level. Therefore, source coding is primarily characterized by rate and distortion/error of the code. However, in practical communication systems, many issues such as memory access requirements (both updating memory and querying from memory) must be considered. In traditional compression algorithms, in order to retrieve one symbol of the source, accessing all the encoded symbols are often required. This issue, referred to as the memory access issue, is the main topic of this paper.

The memory access issue appears in many applications in distributed data management ranges from costly storage to data privacy. In costly storage, assume that we wish to compress a given source and store it on some storage cells. If we use the traditional source coding, then in order to recover only one bit of the original source, we would need to read the entire encoded data on all data storage cells. Since reading from the storage cells is generally costly, we may want to design a compression scheme, where one only needs to read part of the storage cells to recover one bit of the source. In data privacy, assume that we encode a source and then store it on some data storage cells. We are asked to reveal information about one symbol of the source to some party, but, we do not want to reveal the information about the entire source symbols. If we use a conventional source coding, we may have to reveal all the encoded data. Thus, a honest but curious party may have access to the entire original source sequence. In this paper, we introduce a class of source codes with a type of decoder called local decoder, in which the local decoder is restricted to reproduce any source symbol by only querying a limited number of encoded symbols. With the local decoder, we provide only a small part of the encoded data to the party, so that the party can recover its desired information, while keeping the rest information secretly to that party.

B. Mathematical Formulation and Contributions

In this paper, we aim to investigate a source coding scheme called locally decodable source coding (LDSC): A source sequence \( x_n \) (this denotes the vector \((x_1, \ldots, x_n)\)) takes values from the source alphabet \( \mathcal{X} \) and is mapped into a sequence \( y_n \) of encoded symbols taking values in the alphabet \( \mathcal{Y} \). These symbols are then used to generate the reproduction sequence \( \hat{x}_n \). Note that we assume \( x_1, \ldots, x_n \) are drawn i.i.d. according to a probability distribution denoted by \( P_X \). A scheme is called \( t \)-locally decodable, if for any \( i \) \( (1 \leq i \leq n) \), the reproduced symbol \( \hat{x}_i \) is a function of at most \( t \) of the symbols \( y_1, \ldots, y_t \). In other words, the decoder asks \( t \) queries from \( y_i \) to decode any symbol of \( x_i \). We shall define this notion formally in Section II. The number of queries to decode any source symbol, is called locality and is shown by \( t \). This is different from traditional source coding, as we are restricting the way that \( y_i \) can be mapped to the \( \hat{x}_n \) sequence. Similarly, locally decodable lossy source coding (LDLSC) is defined as a lossy source coding with local decoder. Throughout this paper, \( X \) denotes a random variable taking values in \( \mathcal{X} \), where \( x \) denotes an outcome of \( X \). The same notation holds for other letters such as \( Y \) and \( \hat{X} \). Also, for any subset \( S \subset \{1, \ldots, n\} \), \( X^S \) is defined as the vector \((X_i : i \in S)\).

Our contributions are both converse bounds and achievability bounds for the rate of LDSC. We provide a converse bound on the rate of LDSC with linear encoder and show that for binary sources, the rate of any linear LDSC is one, rather than the entropy rate. In other words, we show that using linear LDLSC, compression is not possible. More importantly, we consider binary sources with a general encoder (not necessarily linear) and a 2-local decoder \((t = 2)\), and we show a key result that the rate of compression is one, which implies that source compression is not possible with the 2-local decoder. This result potentially suggests an intuition of the unachievability of source compression for local decoders when the number of query-able symbol is finite w.r.t. the growing block-length.

Then, we study the rate of LDSC in the regime where the number of queries, \( t \), scales with the code block-length, \( n \),
and provide achievability bounds on the rate. In particular, the main achievability result is that with \( O(\log n) \) queries, any rate above entropy rate is achievable. Furthermore, we consider LDLSC for binary sources and provide achievability bound on the rate of which for both scaling number of queries and bounded number of queries. In the case where \( t \) scales with \( n \), we show that the rate distortion is achievable for any number of queries. We provide an upper bound on the rate in the finite block-length regime (finite \( n \)). The achievability results are obtained by dividing a sequence of source symbols into small blocks, applying an encoder-decoder on each block and then concatenate them. Although the achievability bounds are simply based on the concatenation of code blocks, the comparison with the existing results in the data structure literature shows that our achievable bound is tighter than the existing bounds. In the case where \( t \) is bounded, we show that any given rate above the rate distortion can be achieved with large enough locality, \( t \). We also show that for lossy source coding with excess distortion, any rate above the rate distortion is achievable with locality \( O(\log n) \).

C. Related Works and Organization of the Paper

In particular, there exists many researches addressing similar problems from a data structure perspective. For example, Bloom filters ([2]) are data structures for storing a set in a compressed form while allowing membership queries to be answered in bounded time. The rank/select problem and dictionary problem ([3], [4]) in the field of succinct data structures are also examples of problems involving both compression and the ability to recover efficiently a single symbol of the input message. In particular, Patrascu ([5]) provides a succinct data structure, supporting efficient recovery of source symbols. Moreover, Chandar et al. ([6]) introduce a data structure that is efficient in both updating and querying. In most of these works, the efficiency is interpreted in terms of the decoding time. However, in this work the efficiency is interpreted in terms of the number of memory access requirement.

Causal Source Coding is a related topic where the constraint on the decoder is not locality, but, causality (see [7], [8]). Locally decodable codes (LDC) ([9]) is a counter part of LDSC in the error-correction world. Also, Mazumdar et al. [10] study the issue of memory access requirement in channel coding and introduce update efficient codes. Another recent variation is locally repairable codes ([11]).

The problem of source coding with local encoder is studied in several works in both data structure and information theory literatures. This line of research addresses the update efficiency issue. Varshney et al. [12] analyze continuous source codes from an information theoretic point of view. Also, Mossel and Montanari [13] construct source codes with local encoder based on nonlinear graph codes. Sparse linear codes is studied by Mackay [14], who introduces a class of local linear encoders.

The organization of the paper is as follows. In Section II, we give the problem formulation and the converse bounds on the rate of LDSC. We also provide an achievability bound in case of scaling number of queries with block-length. LDLSC is defined in Section III, where we provide achievability bounds on the rate of LDLSC with both bounded and scaling number of queries. We conclude the paper in Section IV.

II. Locally Decodable Source Coding (LDSC)

In this section, we define LDSC and its fundamental limits. We then show converse bounds on the rate of LDSC with bounded number of queries and conclude the section by providing an achievability bound on the rate of LDSC with scaling number of queries.

An almost lossless LDSC is defined as a pair, consisting of an encoder, \( f \), and a decoder, \( g \), such that \( f: X^n \rightarrow \{0,1\}^k \) and \( g: \{0,1\}^k \rightarrow X^n \). The decoder is called local if each coordinate of the output is affected by a bounded number of input coordinates. Formally, Let \( g_a \), for \( a \in \{1,\ldots,n\} \), be the \( a \)-th component of the decoding function. Assume \( g_a \) depends on \( y^k = \{0,1\}^k \) only through the vector \( y^N_a \) for some \( N_a \subset \{1,\ldots,k\} \). In other words, we have:

For any \( y^k \) and \( y^k \), \( g_a(y^k) = g_a(y^k) \) if \( y^N_a = y^N_a \).

For any given \( t \), a decoder is called \( t \)-local if \( \lvert N_a \rvert \leq t \) for any \( a \in \{1,\ldots,n\} \).

Definition 1. A \((n,k,t,\epsilon)\)-LDSC is a pair, consisting of an encoder, \( f: X^n \rightarrow \{0,1\}^k \), and a \( t \)-local decoder, \( g: \{0,1\}^k \rightarrow X^n \), such that

\[
\mathbb{P}[g(f(X^n)) \neq X^n] \leq \epsilon.
\]

Let

\[
k^*(n,\epsilon, t) = \min\{k: \exists (n,k,\epsilon, t)\text{-LDSC}\}.
\]

We define the best rate of a LDSC with locality \( t \) as

\[
R^*(t) = \lim_{\epsilon \to 0} R^*(\epsilon, t),
\]

where

\[
R^*(\epsilon, t) = \limsup_{n \to \infty} \frac{k^*(n,\epsilon, t)}{n}.
\]

Note 1. We assume both encoder and decoder are deterministic, because, for a given \((n,k,\epsilon, t)\)-LDSC with randomized encoder and decoder, there exists an \((n,k,\epsilon, t)\)-LDSC code with deterministic encoder and decoder.

Next, we show a converse bound for the rate of LDSC with linear encoder.

A. Linear Encoder

We focus on binary sources, where \( X = \{0,1\} \). First, we define LDSC with linear encoder and then we show that \( R^*(t) = 1 \), for any locality, \( t \). However, using linear encoder one can achieve any rate above the entropy rate as shown in Theorem 2 of [14] and [13].

Definition 2. An encoder, \( f: X^n \rightarrow \gamma^k (X = \gamma = \{0,1\}) \), is linear if it can be represented by a generating matrix, \( G \in \mathbb{F}_2^{n \times k} \), such that \( G \) is a mapping from \( \{0,1\}^n \) to \( \{0,1\}^k \). The encoding is as \( x \mapsto xG \), where all the operations are over \( \mathbb{F}_2 \).

We use the following lemma to establish a converse bound.

Lemma 1. Let \( \mathbb{F}_2^n \) be a vector space over \( \mathbb{F}_2 \) and \( P_X \) be a Bernoulli(\( p \)) distribution. We define a probability distribution over \( \mathbb{F}_2^n \) according to \( n \)-fold product of \( P_X \). If \( U \) is a \( k \)-dimensional sub-space of \( \mathbb{F}_2^n \), we have

\[
\max\{p, 1-p\}^{n-k} \geq \mathbb{P}[U] \geq \min\{p, 1-p\}^{n-k}.
\]
have mapping \( \pi \) is 0. \( \pi \) → \( Y \)

Consider an arbitrary \((n, \epsilon, t) = (n, \epsilon, t) = (n, \epsilon, t) = (n, \epsilon, t) \) of dimension \( \epsilon < (\min\{p, 1 - p\})^2 \), then we have \( k \geq n \). In particular, we have \( R^*(t) = 1 \).

Proof: We prove this by contradiction. Without loss of generality assume \( p \leq \frac{1}{2} \). For the sake of contradiction, assume \( n > k \). We shall show \( \epsilon \geq p^2 \).

The claim is that if the decoder can recover \( X^{k+1} \) i.i.d. Bern(p) with a 2-local decoder on a set with probability \( p(k) \), then \( p(k) \leq 1 - p^2 \). This implies \( \epsilon \geq p^2 \).

By induction on \( k \) we show \( p(k) \leq 1 - p^2 \). For \( k = 1 \), by considering all 16 possible encoder functions \( (X^2 \rightarrow Y_1) \), it can be seen that \( p(1) \leq 1 - p^2 \). Assume \( p(k - 1) \leq 1 - p^2 \) for any 2-local decoder, we will show \( p(k) \leq 1 - p^2 \), for any 2-local decoder. Consider a 2-local decoder which recovers \( X_{k+1}^k \) from \( Y_{k}^k \). Let \( X_1 \) be recovered by \( Y_1 \) and \( Y_2 \).

Without loss of generality assume \( g_1(0, 0) = 0 \), where for any \( 1 \leq i \leq n \), \( g_i \) is a mapping with two inputs from \( Y_{k}^k \), and output \( X_i \) (the reproduction of \( X_i \)). We consider all cases:

1) \( g_1(0, 1) = 0 \). If we consider the induced mapping from \( Y_2^k \) to \( X_{k+1}^k \), by replacing 0 with \( Y_1 \) in all the mappings that use \( Y_1 \) as one of their inputs, we obtain a local mapping on a set with maximum probability of \( p(k-1) \). Similarly, since \( g_1(1, 1) = g_1(1, 0) = 1 \), if we replace 1 with \( Y_1 \), we obtain another local mapping on a set with maximum probability \( p(k-1) \). Therefore, we have \( p(k) \leq pp(k-1)+(1-p)p(k-1) = p(k-1) \leq 1 - p^2 \).

2) \( g_1(1, 0) = 0 \). In this case, we replace 0 with \( Y_2 \) and construct a mapping from \( (Y_1^k, Y_{k}^k) \) to \( X_{k+1}^k \). Similarly, it can be shown that \( p(k) \leq 1 - p^2 \).

3) \( g_1(1, 1) = 0 \). In this case, we replace \( Y_1 \) by \( Y_2 \) in all the mappings that use \( Y_1 \) as one of their inputs. Similarly, we obtain \( p(k) \leq 1 - p^2 \).
4) \(g_1(1,0) = g_1(0,1) = g_1(1,1) = 0\). In this case, \(X_1 = 1\) cannot be decoded correctly. Thus, \(p(k) \leq 1 - p \leq 1 - p^2\).
5) \(g_1(1,0) = g_1(0,1) = g_1(1,1) = 1\). For a binary variable, \(Y\), let \(\bar{Y}\) denote its complement (\(\bar{Y} = Y + 1\), mode 2). In this case, \(g_1(Y_1, Y_2) = \bar{Y}_1, \bar{Y}_2\). In general, we refer to \(Y_1, Y_2, \bar{Y}_1, Y_2, \bar{Y}_1, \bar{Y}_2\), and their complements as product forms. Next, we consider this case.

Note that if only one of the \(k + 1\) decoding functions, \(g_i\) (\(1 \leq i \leq k + 1\)), is not of the product form, then considering the corresponding mapping and the above argument, by induction we obtain \(p(k) \leq 1 - p^2\). Now, assume all the mappings are in the product form. If \(Y_i\) is appeared in one of the decoding functions as \(X_{i_1} = Y_i Y_j\) and in another one as its complement, i.e., \(X_{i_2} = \bar{Y}_i Y_k\), then \(X_1 = X_2 = 1\) cannot be recovered and we have \(p(k) \leq 1 - p^2\). Therefore, without loss of generality we assume that all the mappings are of the form \(Y_i Y_j\) and no complement is used.

Consider a bipartite graph demonstrating the relation between the variables \(X_1, \ldots, X_{n+1}\) and \(Y_1, \ldots, Y_k\). On the Y-side of it we have \(k\) nodes corresponding to \(Y_1\)'s and on X-side of it we have \(k + 1\) nodes corresponding to \(X_1\)'s. The degree of each node on X-side is 2 indicating the variables on the Y-side that are involved in decoding of that node. If two nodes on the X-side have the same neighbors on the Y-side, then we have \(X_{i_1} = Y_i Y_j\) and \(X_{i_2} = \bar{Y}_i Y_k\) for some \(i, j, i_1\), and \(i_2\). Thus, only \(X_{i_1} = X_{i_2}\) is recoverable and \(p(k) \leq 1 - p^2\). Therefore, there exist nodes such that \(X_{i_1} = Y_i Y_j\), \(X_{i_2} = \bar{Y}_i Y_k\), and \(X_{i_3} = Y_i Y_j\) (note that \(l\) might be equal to \(k\)). If \(X_{i_1} = 1\), then we can find a mapping form \((Y_1^{i_1}, \ldots, Y_k^{i_1})\) to \((X_1^{i_1}, \ldots, X_k^{i_1}, 1)\) on a set with probability \(p(k - 1) \leq 1 - p^2\). Also, note that \(X_{i_1} = 0, X_{i_2} = X_{i_3} = 1\) is not possible to recover. Therefore,

\[
p(k) \leq pp(k - 1) + (1 - p)(1 - p^2) \leq 1 - p^2.
\]

As a result, we have \(R^*(t = 2) = 1\).

We established converse bounds on the rate of LDSC when the number queries is bounded. We concluded that, with bounded number of queries, no compression is possible with linear encoder and general decoder with \(t = 2\). Next, we will focus on the case where the number of queries scales with the block-length. In this case, we will provide achievability bounds on the rate of LDSC.

**C. Scaling Number of Queries**

The number of queries, \(t\), can be a growing function of \(n\). We show that, with \(t = O(\log(n))\), any rate above the entropy is achievable. Note that, in the conventional source coding (not necessarily local), \(t(n)\) scales linearly with \(n\). The following proposition is motivated by the approach given in [10], to establish achievability bounds for update efficient codes.

**Proposition 1.** Let \(X\) be an i.i.d. source with distribution \(P_X\). For any \(R > H(X)\) and \(\epsilon > 0\), there exist \(n_0\) such that for any \(n > n_0\), there exist a \((n, nR, R_{\text{opt}}(\log n, \epsilon))\)-LDSC, where \(E(R) = \min_{Q:H(Q) \geq R} D(Q||P)\).

**Proof:** In order to establish the achievable bound on LDSC with scaling number of queries, we use the following result on the error exponent of source coding. For an i.i.d. source with probability distribution \(P_X\) and a given rate \(R > H(X)\), there exist an \(n_1\) such that for any \(n \geq n_1\) there exists an encoder-decoder pair \(f_n\) and \(g_n\), such that

\[
\mathbb{P}[g_n(f_n(X^n)) \neq X^n] \leq 2^{-nE(R)},
\]

for some constant \(\ell\) (Theorem 2.15 of [15]). Consider the following construction of an encoder-decoder pair for a source sequence of length \(n\), where the distribution of the source is \(P_X\). Let \(X^n\) be a sequence of source symbols. Divide this sequence into blocks of length \(t(n)\) and apply the encoder-decoder pair, found by Theorem 2.15 of [15] to each block separately. Form an encoder-decoder for \(X^n\) by concatenating these \(n^2\) pairs (for the sake of presentation, we drop ceiling and floor in this discussion) pairs of encoder-decoder. The constructed decoder has locality no greater than \(t(n)\). We now remove the error of the concatenated source codes. Using a union bound, we obtain

\[
\mathbb{P}[^\hat{X}^n \neq X^n] = \mathbb{P}\left[\bigcup_{i=1}^{n/t(n)} \{[^\hat{X}^n_{(i-1)t(n)+1}] \neq X_{(i-1)t(n)+1}\}\right] \leq \frac{n}{t(n)} \mathbb{P}[^\hat{X}^n \neq X^n].
\]

Using (4), we have

\[
\mathbb{P}[^\hat{X}^n \neq X^n] \leq \frac{n}{t(n)} 2^{-nE(R)}.
\]

Since \(E(R) > 0\), this bound goes to zero if \(t(n) > \frac{1}{E(R)} \log n\). Therefore, we choose \(n_0\) greater than \(n_1\) and also large enough to make \(\frac{n}{t(n)} 2^{-nE(R)}\) smaller than \(\epsilon\).

Next, we will turn our attention to LDLSC and establish achievability results for the rate of LDLSC.

**III. Locally Decodable Lossy Source Coding**

(LDSC)

We first define LDLSC and the fundamental limits of it. Then we provide achievability bounds on the rate of LDLSC for both scaling number of queries with \(n\), and bounded number of queries. Consider a separable distortion metric defined as \(d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)\), where \(d: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+\) is a distortion measure.

**Definition 3.** A \((n, k, d, t)\)-LDLSC is a pair containing an encoder, \(f: \mathcal{X}^n \to \mathcal{Y}^k\), and a decoder, \(g: \mathcal{Y}^k \to \mathcal{X}^n\), where the decoder is \(t\)-local and the distortion is bounded,

\[
\mathbb{E}[d(X^n, g(f(X^n)))] \leq d.
\]

Let \(k^*(n, d, t) = \min(k, \text{such that } \exists (n, k, d, t) - \text{LDLSC})\), and \(R^*(d, t) = \lim_{n \to \infty} \frac{k^*(n, d, t)}{n}\).

**Note 2.** For a binary source, if we let \(d(x, \hat{x}) = 1\{x \neq \hat{x}\}\), then we have \(\mathbb{E}[d(X^n, g(f(X^n)))] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}[X_i \neq \hat{X}_i] \leq d\), which is the same as assuming the bit error rate is bounded (comparing to block error rate in the definition of LDLSC).

**A. Scaling Number of Queries**

In this section, we consider a binary source with Bernoulli distribution and assume the distortion level, \(d\), is below \(p\). Let \(t(n)\) be a growing function of \(n\). The following is an achievable bound on the rate of LDLSC in the finite block-length regime.
Proposition 2. For a Bern(p) source, a distortion level d, and any growing number of queries t(n), we have
\[
R^*(n, d, t(n)) \leq h(p) - h(d) + \frac{\log t(n)}{t(n)} + o\left(\frac{\log t(n)}{t(n)}\right) \tag{5}
\]

Proof: Since d \leq p, we have R(d) = h(p) - h(d).
Using Theorem 5 of [16], we have: for a Bern(p) source, and distortion level d, there exists a source coding such that
\[
R(n, d) \leq h(p) - h(d) + \frac{\log n}{n} + o\left(\frac{\log n}{n}\right). \tag{6}
\]
Now, divide the sequence X^n into \( \frac{n}{t(n)} \) blocks of length t(n) (for the sake of presentation, we drop ceiling and floor in this argument). Apply the encoder-decoder obtained from (6) to each block. Concatenate these \( \frac{n}{t(n)} \) pairs to obtain an encoder-decoder for X^n. The locality of the constructed decoder is no greater than t(n). The average distortion of the overall code is also bounded by d, and its rate is bounded by h(p) - h(d) + \( \frac{\log t(n)}{t(n)} + o\left(\frac{\log t(n)}{t(n)}\right) \), which concludes the proof. \qed

Proposition 2 shows that, for any number of queries, t(n), if \( \lim_{n \to \infty} \frac{t(n)}{n} = \infty \), then \( R^*(t(n)) = h(p) - h(d) \), which is the rate distortion.

Corollary 1. For the special case of t(n) = t \log n, we have the following upper-bound on \( R^*(n, d, t \log n) \).
\[
h(p) - h(d) + \frac{\log(t \log n)}{t \log n} + o\left(\frac{\log(t \log n)}{t \log n}\right) \tag{7}
\]

Proof: We apply Proposition 2 for t(n) = t \log n. \qed

Reference [5] studies the problem of storage of bits with local recovery (with the same definition of locality we use here). Their results are based on a generic transformation of augmented B-trees to succinct data structures.

Theorem 3 (Theorem 2, [5]). We can represent a binary sequence of length n, with pn ones, using \( \log \binom{n}{pn} + \frac{n}{4} + O(n^{3/4} \log n) \) bits, where a decoder exists querying only \( t \log n \) bits to decode any bit of the sequence.

We now compare the bound given in Corollary 1 with the bound suggested by Theorem 3. Using Theorem 3 and the identity \( \log \binom{n}{pn} = nh(p) + O(\log n) \), for any d, we obtain
\[
R^*(n, d, t \log n) \leq h(p) + O\left(\frac{\log n}{n}\right) + \frac{1}{(t/n)^t} + \frac{1}{n} O(n^{3/4} \log n). \tag{8}
\]
For any fixed d, the bound given by (7) is asymptotically (in n) better than (8). Note that the bound given in (8) does not gain from the fact that an encoding-decoding scheme can tolerate a distortion of d.

Remark 1. One may consider the case where d goes to zero as n goes to infinity. In order to compare the bounds given in (8) and (7), without proof, assume both bounds hold for this case as well. If \( d(n) = O\left(\frac{1}{\log n}\right) \), then (8) is tighter than (7).

In order to show (8) is tighter than (7), we need to have
\[
h(p) + O\left(\frac{\log n}{n}\right) + \frac{1}{(t/n)^t} + \frac{1}{n} O(n^{3/4} \log n) \leq h(p) - h(d(n)) + \frac{\log\log n}{\log n}.
\]
This inequality holds if \( h(d(n)) \leq \frac{\log\log n}{\log n} - O\left(\frac{\log n}{n}\right) - \frac{1}{\log n} \). It can be seen that, for d(n) = O(\frac{1}{\log n}), this inequality holds.

Next, we provide an achievability for LDLSC with bounded number of queries.

B. Bounded Number of Queries
For a given number of queries, we show that one can achieve any rate above the rate distortion function with a properly large locality. Consider a binary source with Bern(p) distribution and rate distortion R(d), for any distortion level, d. We show the following result.

Proposition 3. Consider a binary source, a distortion level, d, and rate distortion, R(d). For a given rate R > R(d), there exist t_0 such that for any t \geq t_0 there exist a LDLSC of rate R with t-local decoder.

Proof: Consider the following construction. Using (6), we obtain the bound \( R(t(n), d) \leq R(d) + 2\log t(n) \), for large enough t_0. Also, let t_0 be large enough such that
\[
\frac{2\log t_0}{t_0} \leq R - R(d).
\]
Therefore, for any t \geq t_0, we have \( R(t, d) \leq R \). Thus, there exists an encoder-decoder pair for X^t, such that the rate of the code is less than R and the distortion is bounded by d. Now, consider \( \frac{t}{2} \) pairs of the same encoder-decoder (we drop ceiling and floor in this discussion). Concatenate these encoder-decoder pairs to form an encoder-decoder for X^n. The locality of the constructed decoder is no greater than t. The distortion of the constructed encoder-decoder pair for X^n is bounded by d, because we have \( \sum_{i=1}^{n} d(x_i, \hat{x}_i) \leq \frac{2}{t} \sum_{j=1}^{\frac{t}{2}} \mathbb{E} d(X^t_{ij}, X^t_{ij+1}) \leq d \), and each term of the summation is bounded by d. The rate of the constructed encoder-decoder pair is \( R^*(nt, d, t) \leq R \). Therefore, for any block length, there exists a t-local LDLSC with rate R and average distortion bounded by d as long as t \geq t_0. This completes the proof. \qed

C. LDLSC for Excess Distortion
In this section, we will consider locally decodable lossy source coding for excess distortion, where instead of bounding the average distortion, we bound the probability of the event that the distortion exceeds a level d \geq 0. We will define LDLSC for excess distortion and provide an achievable bound on its rate.

Definition 4. A (n, k, d, t, \epsilon)-LDLSC for excess distortion is a pair consisting of an encoder \( f : X^n \to \{0,1\}^k \) and a decoder \( g : \{0,1\}^k \to X^n \), where the decoder is t-local and the excess distortion is bounded,
\[
\mathbb{P}(d(X^n, g(f(X^n))) > d) \leq \epsilon.
\]
For a given $\epsilon$, $d$ and block-length $n$, let $k^*_e(n, d, t, \epsilon)$ be
\[
\min \{ k : \exists (n, k, d, r, \epsilon) - \text{LDLSC for excess distortion} \},
\]
where $e$ stands for excess distortion. The rate of LDLSC for excess distortion is defined as $R^*_e(d, t) = \lim_{n \to \infty} R^*_e(n, d, t, \epsilon)$, where $R^*_e(d, t, \epsilon) = \limsup_{n \to \infty} k^*_e(n, d, t, \epsilon)$.

We use the following results on the error exponent of excess distortion lossy source coding.

**Theorem 4** (Theorem 1, [17]). For source with distribution $P_X$ and a distortion level $d$, we have:

for any $\epsilon > 0$, $\exists K_\epsilon$ such that for any $n \geq 0$ there exists encoding-decoding pair $f_n$ and $g_n$, such that
\[
P[\|g_n(f_n(X^n)) - X^n\| > d] \leq K_\epsilon 2^{-n(F_d(R) - \epsilon)},
\]
where, $F_d(R) = \min \{ D(Q||P) : R(Q, d) \geq R \}$, and $R(Q, d)$ is defined as the following where the probability distribution over $X$ is $Q$.
\[
R(Q, d) = \min_{P_X : \mathbb{E}[d(X, \hat{X})] \leq d} I(X; \hat{X}).
\]

Next, we show that with $O(\log n)$ number of queries, LDLSC with excess distortion can achieve any rate above rate distortion.

**Proposition 4.** Let $X$ be a Bernoulli source and $f : X^n \to \{0, 1\}^k$ and $g : \{0, 1\}^k \to X^n$ be encoder and decoder respectively. For any $\epsilon > 0$ and $R > R_d(d)$, there exists a constant $C$ and $n_0$ such that for $n > n_0$ there exists a $(n, nR, C \log n, d, \epsilon) - \text{LDLSC for excess distortion}$.

**Proof:** We use Theorem 4 to design locally decodable codes for excess distortion. Let $\epsilon < F_d(R)$. Consider a source sequence of length $n$. Divide the sequence into sequences of length $t(n)$. Therefore, we have $\frac{n}{t(n)}$ blocks of length $t(n)$ (for the sake of presentation, we omit ceiling and floor in this discussion). Consider the corresponding encoder-decoder pair to each block of length $t(n)$, obtained from Theorem 4. Form an encoder-decoder pair for the whole sequence by concatenating these encoder-decoder pairs. The resulting decoder is $t(n)$-local. Using the union bound, we obtain
\[
P[\|d(g(f(X^n)) - X^n\| > d] \leq \sum_{j=1}^{n/t(n)} P[d(g(f(X^{jt(j-1)+1}) - X^{jt(j-1)+1}) > d] \leq \frac{n}{t(n)} K_\epsilon 2^{-t(n)(F_d(R) - \epsilon)}.
\]

We need $t(n)$ to be such that
\[
\lim_{n \to \infty} K_\epsilon \frac{n}{t(n)} 2^{-t(n)(F_d(R) - \epsilon)} = 0.
\]
Therefore, if $t(n) > C \log n$, for some $C > \frac{1}{F_d(R) - \epsilon}$, then the error goes to zero. This completes the proof. \hfill $\Box$

**IV. CONCLUSIONS AND FUTURE WORK**

We introduced locally decodable source coding in both almost lossless and lossy cases. The following summarizes the main results.

- Almost lossless source coding:
  - bounded locality: We conclude that, with bounded number of queries, no compression is possible with linear encoder, linear decoder, and general encoder-decoder with $t = 2$. To establish a converse bound for a general encoder-decoder with $t \geq 3$, is postponed to future works.
  - scaling locality with block-length: any rate above the entropy is achievable with locality $O(\log n)$, where $n$ is block-length.

- Lossy source coding:
  - scaling locality with block-length: for a binary source, the rate distortion is achievable with any scaling locality ($\lim_{n \to \infty} t(n) = \infty$) and the rate of convergence is upper bounded as given in Proposition 2. The bound given in Proposition 2 outperforms the existing bound in reference [5].
  - bounded locality: for a binary source, any given rate above the rate distortion is achievable with a bounded locality as given in Proposition 3.
  - excess distortion: any rate above the rate distortion is achievable with locality $O(\log n)$, where $n$ is the block-length.

**REFERENCES**