Local Decoding in Distributed Compression

Shashank Vatedka®, Member, IEEE, Venkat Chandar, and Aslan Tchamkerten®

Abstract—A recent result says that the lossless compression of a single source \( X^n \) is achievable with a strong locality property; any \( X_i \) can be decoded from a constant number of compressed bits, with a vanishing in \( n \) probability of error. By contrast, we show that for two separately encoded sources \( (X^n, Y^n) \), lossless compression and strong locality is generally not possible. Specifically, we show that for the class of “confusable” sources, strong locality cannot be achieved whenever one of the sources is compressed below its entropy. Irrespective of \( n \), there always exists at least one index \( i \) for which the probability of incorrectly decoding \( (X_i, Y_i) \) is lower bounded by \( 2^{-\Omega(d)} \), where \( d \) denotes the worst-case (over indices) number of compressed bits accessed by the local decoder. Conversely, if the source is not confusable, strong locality is possible even if one of the sources is compressed below its entropy. Results extend to an arbitrary number of sources.

Index Terms—Distributed data compression, source coding, local decoding.

I. INTRODUCTION

The storage of massive datasets has highlighted the growing need for compression schemes with locality properties [2], [3], [4], [5], [6], [7], [8]. One such basic property, referred to as “local decoding,” is the ability to efficiently decode part of the dataset by accessing only a small portion of its compressed version. Classic compression schemes such as Lempel-Ziv [9], [10] are not local since the decoding of the message symbols, and in particular, hence, only “weak” locality holds in the sense that for the local decoder error probability to vanish, the number of probed bits—here equal to the sub-block codeword length \( b \cdot R \)—must grow with \( n \).

In this paper, we address the question whether strong locality extends to the Slepian-Wolf distributed compression of two sources: if \( X^n \) and \( Y^n \) are compressed at rates \( R_1 \) and \( R_2 \), respectively, is it possible to design a fix-length compressor and a local decompressor which can decode any message symbol from a constant number of compressed bits, with an error probability that decays as \( n \) grows?

If each source is compressed above its entropy, then strong locality obviously holds simply by duplicating the results of [11], [12] separately for each of the sources. We also observe that the concatenation scheme—wherein \( (X^n, Y^n) \) is decomposed into consecutive sub-blocks of size \( b \) that are encoded via Slepian-Wolf coding—achieves weak locality at any \((R_1, R_2)\) within the Slepian-Wolf rate region. So the interesting question is: does strong locality hold when at least one of the sources is compressed below its entropy?

The answer depends on the source distribution. Suppose \( p_{XY} \) is “confusable” in the sense that, for every \( x_1 \) and \( x_2 \) in \( \mathcal{X} \) there exists \( y \in \mathcal{Y} \) such that \( p_{XY}(x_1, y) > 0 \) and \( p_{XY}(x_2, y) > 0 \). In this case, we show that if \( R_1 < H(X) \), then for some index \( 1 \leq i \leq n \) the probability of wrongly decoding \( (X_i, Y_i) \) is lower bounded by \( 2^{-\Omega(d)} \), where \( d \) denotes the (worst-case) number of bits probed by the local decoder. This conclusion holds even if the decoder tries to only decode \( X_i \), with the full cooperation of the \( Y \)-transmitter that provides \( Y^n \) uncompressed. Conversely, if \( p_{XY} \) is not confusable, then strong locality is possible for some \( R_1 < H(X) \) and \( R_2 = H(Y) \).

Hence, when the source is confusable, only the weak locality as achieved by the concatenation scheme is possible. However, a drawback of the concatenation scheme is that its encoding procedure is tied to the sub-block length \( b \) which governs the error probability of the local decoder. Even if the local decoder probes all \( n(R_1 + R_2) \) bits of the \( X \)- and \( Y \)-codewords, the local error probability remains the same as if only \( b(R_1 + R_2) \) were probed.

Thus, for the local decoder to achieve a lower error probability, the encoding procedure must be changed. We address this limitation through a hierarchical (weakly local) compression scheme whose local decoding error probability decreases as more symbols get probed—without modifying the encoding. Specifically, for any \((R_1, R_2)\) within the Slepian-Wolf rate region, and for any constant \( \eta \) such that \( 2^{-2^{O(1/\eta)}} < \eta < 1 \), the local decoder achieves \( p_{loc} \leq \eta \) with \( d = \text{poly}(\log(1/\eta)) \).

Manuscript received 15 April 2022; revised 23 September 2022; accepted 16 December 2022. Date of publication 2 February 2023; date of current version 13 June 2023. The work of Shashank Vatedka was supported in part by the Seed Grant from IIT Hyderabad and in part by the Start-Up Research Grant from the Science and Engineering Research Board, India, under Grant SRG/2020/0000910. This work was presented in part at the 2022 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT50566.2022.9834371]. (Corresponding author: Shashank Vatedka.)

Shashank Vatedka is with the Department of Electrical Engineering, Indian Institute of Technology Hyderabad, Hyderabad 502285, India (e-mail: shashankvatedka@ee.iith.ac.in).

Venkat Chandar is with DE Shaw, New York, NY 10036 USA.

Aslan Tchamkerten is with the Department of Communications and Electronics, Telecom Paris, Institute Polytechnique de Paris, 91120 Palaiseau, France.

Digital Object Identifier 10.1109/JSAIT.2023.3240187
A. Literature on Locally Decodable Compression

Local decoding has been studied extensively in the context of compressed data structures by the computer science community; see, e.g., [13], [14], [15], [16], [17], [18] and the references therein. Most of these results hold under the word-RAM model which assumes that operations (memory access, arithmetic operations) on $w$-bit words take constant time. The word size $w$ is typically chosen to be $\Theta(\log n)$ bits, motivated in part by on-chip type of applications where data transfer happens through a common memory bus for both data and addressing (hence $w = \Theta(\log n)$ bits), and partly by the fact that certain proof techniques work only when $w = \Omega(\log n)$.

In the word-RAM model, it is possible to compress any sequence to its empirical entropy and still be able to locally decode any message symbol in constant time [13], [14]. Most approaches modify the Lempel-Ziv class of algorithms to provide efficient local decodability [19], [20], [21]. Similar results also hold for compression of correlated data [22], and efficient recovery of short substrings of the message [19], [23], [24], [25]. However, all of these schemes require the local decoder to probe at least $O(\log n)$ compressed bits to recover any source symbol.

In this work, the decoding cost is measured by the number of compressed bits that need to be accessed in order to recover a single source symbol. This is sometimes referred to as the local decodability [26], or the bit-probe complexity in the literature [27].

The problem of locally decodable source coding of random sequences with a vanishing (in the literature [27]) probability of error was first studied in [26], [28]. These works showed that any compressor with $\varepsilon = 2$ cannot achieve a rate below the trivial rate $\log |\mathcal{X}|$, and similarly for linear source code that achieves $\varepsilon = \Theta(1)$. Later, Mazumdar et al. [11] showed that for any $\varepsilon > 0$, rate $H(X) + \varepsilon$ is achievable with local decodability $\varepsilon = \Theta(\frac{1}{\log \frac{1}{\varepsilon}})$. Moreover, for non-dyadic sources, $\varepsilon = \Omega(\log(1/\varepsilon))$ for any compression scheme that achieves rate $H(X) + \varepsilon$. Inspired by [11], a compressor of Markov sources was given in [12] which achieves a rate-locality trade-off ($R = H(X) + \varepsilon$, $\varepsilon = \Theta(\frac{1}{\log \frac{1}{\varepsilon}})$). A common feature of the code construction in [11], [12] is the use of the bitvector compressor of Buhrman et al. [29] which is based on a nonexpicit construction of expander graphs.

The above references on the bit-probe model consider fixed-length block coding, Variable-length coding was investigated by Pananjady and Courtade [30] who gave compression rate upper and lower bounds for the compression of sparse sequences under local decodability constraints. The works [31], [32] considered simultaneous local decodability and update efficiency. In particular, [31] exhibited a compressor whose average-case local decodability (defined as the expected number of bits that need to be probed to recover any $X_i$) and the average-case update efficiency (the expected number of bits that need to be read and written in order to update a single $X_i$) both scale as $O(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon})$. In fact, our scheme for distributed compression with locality is inspired by the multilevel compression scheme in [31]. The paper [32] designed a compression scheme whose worst-case local decodability and update efficiency scales as $O(\log \log n)$.

More recently, [33], [34] implemented different versions of the concatenation scheme and evaluated its performance on practical datasets.

B. Paper Organization

In Section II, we introduce notions of localities and formally define the problem. In Section III, we present our results, and prove them in Sections IV, V, and VI. In Section VII, we discuss extensions to more than two sources and, in Section VIII, we draw concluding remarks.

II. PRELIMINARIES AND PROBLEM STATEMENT

A. Distributed Compression Without Locality

Let $(X^n, Y^n)$ be $n$ independent copies of a pair of random variables $(X, Y) \sim p_{XY}$ defined over some finite alphabet $\mathcal{X} \times \mathcal{Y}$, with $|\mathcal{X}| \geq 2$, $|\mathcal{Y}| \geq 2$. Without loss of generality, we assume that $\mathcal{X} = \{x : p_X(x) > 0\}$ and $\mathcal{Y} = \{y : p_Y(y) > 0\}$.

Sequences $X^n$ and $Y^n$ represent two sources of information separately encoded into binary codewords $C^{nR_1}$ and $C^{nR_2}$ at rates $R_1$ and $R_2$, respectively. Upon receiving these codewords, a receiver outputs an estimate $(\hat{X}^n, \hat{Y}^n)$ of the sources and makes an error with probability

$$P_e \overset{\text{def}}{=} \Pr[(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)].$$

The optimal rate region is the closure of the set of rate pairs $(R_1, R_2)$ for which $P_e \to 0$ as $n \to \infty$, and it is given by: Theorem 1 (Slepian-Wolf, [35], [36]): The optimal rate region of a source specified by $p_{XY}$ is the set of pairs $(R_1, R_2)$ that satisfy

$$R_1 \geq H(X|Y) \quad R_2 \geq H(Y|X) \quad R_1 + R_2 \geq H(X, Y).$$

Moreover, for any $(R_1, R_2)$ in the interior of the optimal rate region, and any $\varepsilon > 0$, there exist a sequence of coding schemes operating at rates at most $R_1 + \varepsilon$ and $R_2 + \varepsilon$ such that

$$P_e \leq 2^{-n(E-\varepsilon)},$$

where $E$ is a constant that depends on $R_1, R_2$ and $p_{XY}$.

B. Distributed Compression With Locality

1) Local Decoder: Given encodings $C^{nR_1}$ and $C^{nR_2}$, a local decoder takes as input $i \in [n]^1$ probes/reads a fix set $I_i$ of components from $C^{nR_1}$ and $C^{nR_2}$, which we denote as $C_{I_i}$, and outputs an estimate

$$(\hat{X}_i, \hat{Y}_i) = (\hat{X}_i(C_{I_i}), \hat{Y}_i(C_{I_i}))$$

of $(X_i, Y_i)$.

The worst-case local decodability and error probability are defined as

$$\alpha \overset{\text{def}}{=} \max_{1 \leq i \leq n} d(i),$$

$$P_e \overset{\text{def}}{=} \Pr[(\hat{X}^n, \hat{Y}^n) \neq (X^n, Y^n)].$$
where \(d(i) \overset{\text{def}}{=} |I_i|\), and
\[
P_e^{\text{loc}} \overset{\text{def}}{=} \max_{1 \leq i \leq n} \Pr \left[ (\hat{X}_i, \hat{Y}_i) \neq (X_i, Y_i) \right].
\]
A few words concerning the properties of the set \(I_i\). First, it may contain different components from \(CnR_1\) and \(CnR_2\). Second, it should be non-randomized, and this is without loss of generality. Finally, the components of \(I_i\) should be chosen non-adaptively; conditioned on the index \(i\), set \(I_i\) should be independent of \((X^n, Y^n)\). The reason for this restriction is that our results relate to error probability bounds for local decoders and that they immediately translate to bounds for local decoders with adaptive queries. To see this note that a sequence of \(d \geq 1\) adaptive queries generates \((\text{at most})\) \(2d-1\) different sequences of probed bits.\(^2\) Because of this correspondence, bounds on the probability of error for locality-\(d\) decoders with adaptive probes, translate into lower bounds for locality-\(2d-1\) non-adaptive decoders.

**Remark 1:** The notations \(I_i\) and \(C_{I_i}\) leave out any explicit reference to the underlying code. In the sequel, it is understood that if both sources \(X^n\) and \(Y^n\) are compressed, then \(C_{I_i}\) may contain components from both \(CnR_1\) and \(CnR_2\). Instead, if only source \(X^n\) is compressed (say, with \(Y^n\) revealed to the local decoder uncompressed), then \(C_{I_i}\) contains components from \(CnR_1\) only. From the context there should be no confusion.

2) **Strong vs. Weak Locality:** A rate pair \((R_1, R_2)\) is said to be achievable with **strong locality** if
\[
P_e^{\text{loc}} = o(1) \quad \text{with} \quad d = \Theta(1) \quad \text{as} \quad n \to \infty.
\]
That is, by probing only a constant number (independent of \(n\)) of symbols, the error probability of the local decoder goes to zero as the blocklength increases. By contrast, \((R_1, R_2)\) is said to be achievable with **weak locality** if
\[
P_e^{\text{loc}} = o(1) \quad \text{with} \quad d = o(1) \quad \text{as} \quad n \to \infty.
\]
Weak locality is always achievable through the concatenation scheme where source sequences \(X^n\) and \(Y^n\) are decomposed into length \(b\) sequences
\[
X^b(j) \overset{\text{def}}{=} X_{(j-1)b+1}^{jb} \overset{\text{def}}{=} (X_{(j-1)b+1}, X_{(j-1)b+1}, \ldots, X_{jb})
\]
\[
y^b(j) \overset{\text{def}}{=} Y_{(j-1)b+1}^{jb} \overset{\text{def}}{=} (Y_{(j-1)b+1}, Y_{(j-1)b+1}, \ldots, Y_{jb})
\]
for \(j = 1, 2, \ldots\) and where each block \((X^b(j), y^b(j))\) is independently compressed using a Slepian-Wolf code operating at the desired \((R_1, R_2)\). Given \(i \in [n]\), the local decoder decodes block \(j = [i/b]\), and outputs the estimates of the \(i\)-th block of \(X^n\) and \(Y^n\). For this scheme we have \(d = b(R_1 + R_2)\) and from Theorem 1 we get:

**Corollary 1 (Concatenation):** For any source \(p_{XY}\) and any \((R_1, R_2)\) in the interior of the optimal rate region (1), the concatenation scheme achieves weak locality:
\[
P_e^{\text{loc}} \leq 2^{-\Theta(d)}.
\]

\(^2\)A probing scheme for \(d\) adaptive non-randomized queries can be represented as a complete binary decision tree of depth \(d - 1\). Each node in this tree (including the root and the leaves) is labeled with a codeword component (among the \(nR_1 + R_2\) possible), and each edge is labeled 0 or 1. Any instance of \(d\) adaptive queries describes one of the \(2^{d-1}\) paths from the root to the leaves, branching at each node towards the “0” or “1” edge depending on the node value.

C. **Statement of the Problem**

By contrast with weak locality, whether strong locality is generally achievable is much less clear. In fact, it is only recently that strong locality was shown to be achievable for the single source setup at any lossless compression rate \(R > H(X)\) (see [11], [12]). For the Slepian-Wolf setup at hand, this result implies that strong locality holds for any \((R_1, R_2)\) such that \(R_1 > H(X)\) and \(R_2 > H(Y)\). In this regime, sources can be encoded using the single source strongly local codes of [11], [12], separately for source \(X^n\) and source \(Y^n\)—and ignore dependency between \(X^n\) and \(Y^n\).

The central question addressed in this paper is: does strong locality extend to the regime where at least one of the sources is encoded at a rate below its entropy?

III. **Main Results**

Our main result answers the above question in the negative: if the source is “confusable,” strong locality is impossible whenever one of the sources is compressed below its entropy.

**Definition 1 (Source Confusability):** Source \(p_{XY}\) is said to be \(X\)-confusable if, for every \(x_1, x_2 \in X\), there exists \(y \in Y\) such that \(p_{XY}(x_1|y) > 0\) and \(p_{XY}(x_2|y) > 0\)—recall that \(p_Y(y) > 0\) for any \(y \in Y\), see Section II-A. \(Y\)-confusability is defined similarly.

Any source with full support, i.e., such that \(p_{XY}(x, y) > 0\) for all \((x, y)\) in \(X \times Y\), is both \(X\)- and \(Y\)-confusible. An example of an \(X\)-confusible source which does not have full support is \(p_{XY}\) where \(p_Y(y) = \text{Bernoulli}(p), 0 < p < 1, \) and where \(p_{XY}\) is a Z-channel with crossover parameter \(0 < \epsilon < 1\). Instead, if \(p_{XY}\) is the erasure channel, source \(p_{XY}\) is not \(X\)-confusible. If \(X = Y = \{0, 1\}\), source \(p_{XY}\) is always \(X\)-confusible unless \(p_{XY}\) is the noiseless channel.

**Theorem 2:** Suppose source \(p_{XY}\) is \(X\)-confusible and that \(X^n\) is encoded into codeword \(CnR_1\) with \(R_1 < H(X)\). Suppose that to decode \(X_i, i \in [n]\), the local decoder \(\hat{X}_i(C_{I_i}, Y^n)\) has access to both \(C_{I_i}\) and \(Y^n\). Then,
\[
\max_{1 \leq i \leq n} \Pr \left[ \hat{X}_i(C_{I_i}, Y^n) \neq X_i \right] \geq 2^{-\Theta(d)}
\]
where \(d\) denotes the worst-case local decodability.\(^3\)

This result says that if the source is \(X\)-confusible, then strong locality is impossible whenever \(R_1 < H(X)\); not even the full cooperation of the \(Y\)-transmitter through the uncompressed source \(Y^n\) allows to achieve strong locality. Under adaptive probing, the lower bound given in Theorem 2 becomes (see Section II-B)
\[
\max_{1 \leq i \leq n} \Pr \left[ \hat{X}_i(C_{I_i}, Y^n) \neq X_i \right] \geq 2^{-\Theta(2d)}
\]
Hence, if the source is \(X\)-confusible and if \(R_1 < H(X)\), then strong locality cannot be achieved even under adaptive probing.

The \(X\)-confusability property turns out to be necessary for Theorem 2 to hold:

**Theorem 3:** Suppose source \(p_{XY}\) is not \(X\)-confusible. Then, strong locality is achievable at some \(R_1 < H(X)\) and \(R_2 = H(Y)\).

\(^3\)See Remark 1.
From Corollary 1, for any \((R_1, R_2)\) in the interior of the optimal rate region, the concatenation scheme achieves a local error probability that decays as \(2^{-\Theta(n)}\), and this is order optimal by Theorem 2 for confusable sources. One drawback of this scheme, though, is that its local decoding error-probability is tied to its encoding procedure through the sub-block length \(b\). In particular, if \(b = \Theta(1)\), it is impossible to recover \((X^n, Y^n)\) with vanishing probability of error as \(n\) grows, even after probing the entire compressed sequences—because each sub-block is encoded independently. As a consequence, to reduce the local decoding error-probability, the parameter \(b\), hence the encoding procedure, should be modified accordingly.

Our second contribution is a compression scheme whose local decoder has an error probability that decreases as the number of probed symbols increases, without changing the encoding. The performance of this scheme is given in the following theorem:

**Theorem 4:** For any \((R_1, R_2)\) in the interior of the optimal rate region, there exists a rate \((R_1, R_2)\) encoder and a local decoder such that, for every constant \(\eta\) such that

\[
2^{-2\Theta(\log(n))} < \eta < 1,
\]

the local decoder achieves

\[
P_e^\text{loc} \leq \eta
\]

while probing \(d = \text{poly}(\log(1/\eta))\) bits.

Theorems 2, 3, and 4 easily generalize to more than two sources, see Section VII.

**Note:** The present paper differs from the ISIT paper [1] mainly in that it establishes the impossibility of strong locality (Theorem 2) for the most general class of sources (confusable sources), and not only for the specific class of doubly symmetric binary sources. In fact, the arguments used in [1] do not extend beyond sources with full-support. Theorem 3 is new and [1] contains mostly a sketch of the proof of Theorem 4. Theorems 5, 6, and 7 that extend the above results to more than two sources (see Section VII) did not formally appear in [1].

**IV. PROOF OF THEOREM 2**

In Section IV-A we provide the three main ingredients, Lemmas 1, 2, and 3, used to prove Theorem 2 in Section IV-B. These lemmas are proved in Section IV-C.

**A. Preliminaries**

The first ingredient relies on the following coupling. Given \(p_{XY}\), define random variable \(\tilde{X}\) so that

\[
X - Y \sim \tilde{X}
\]

forms a Markov chain, and so that

\[
p_{X|Y} = p_{\tilde{X}|Y}.
\]

Observe that if \(p_{XY}\) is \(\mathcal{X}\)-confusable then, for any \((x, \tilde{x}) \in \mathcal{X} \times \mathcal{X}\), there exists \(y\) with \(p_Y(y) > 0\) (recall that without loss of generality \(p_Y(y) > 0\) for any \(y \in \mathcal{Y}\)) such that

\[
p_{X|Y}(x, \tilde{x}|y) = p_{X|Y}(x|y)p_{\tilde{X}|Y}(\tilde{x}|y) > 0.
\]

Hence, if \(p_{XY}\) is \(\mathcal{X}\)-confusable, then \(p_{X\tilde{X}}\) has full support. In turn, since distributions with full support (and finite alphabet) are reverse hypercontractive [37, Th. 1], we have:

**Lemma 1:** If \(p_{XY}\) is \(\mathcal{X}\)-confusable then, for every \(\mathcal{A}, \mathcal{B} \subset \mathcal{X}^n\), we have

\[
\Pr[X^n \in \mathcal{A}, \tilde{X}^n \in \mathcal{B}] \geq (\Pr[X^n \in \mathcal{A}])^\alpha (\Pr[\tilde{X}^n \in \mathcal{B}])^\beta
\]

for some finite constants \(\alpha, \beta\)—here \((X^n, \tilde{X}^n) = (X_1, \tilde{X}_1), \ldots, (X_n, \tilde{X}_n)\) is i.i.d. according to the distribution \(p_{X\tilde{X}}\) induced by \(p_{XY}\).

The second ingredient used to prove Theorem 2 is the following lemma which also exploits the above coupling:

**Lemma 2:** Fix source \(p_{XY}\). Suppose \(X^n\) is encoded into codeword \(C^n_{\tilde{X}}\) at some rate \(R_1 \geq 0\). Let \(\tilde{X}(C^n_{\tilde{X}}, Y^n)\) be a local decoder for \(X_i\) given \(C_{\tilde{X}}\) and \(Y^n\). Then, for any realization \(c\) of \(C_{\tilde{X}}\) and any \(x \in \mathcal{X}\), we have

\[
\Pr[\tilde{X}(C_{\tilde{X}}, Y^n) \neq X_i, C_{\tilde{X}} = c] \
\geq \Pr[X_i = x, C_{\tilde{X}} = c, \tilde{X} \neq x, \tilde{C}_{\tilde{X}} = c],
\]

where \(\tilde{C}\) is obtained by encoding \(\tilde{X}^n\) using the same encoding/local decoding procedure as for \(X^n\).

The last ingredient used to prove Theorem 2 follows from a basic rate-distortion argument:

**Lemma 3:** Suppose \(X^n\) is encoded into codeword \(C^n_{\tilde{X}}\) at some rate \(R_1 < H(X)\). For any local decoder \(\tilde{X}(C_{\tilde{X}}, Y^n)\) of worst-case local decodability \(d\) with \(\tilde{d} > 0\) that depends only on \(R_1\) and \(p_X\), an index \(i \in [n]\), and a realization \(c\) of \(C_{\tilde{X}}\), such that

\[
\Pr[X_i \neq \tilde{X}(C_{\tilde{X}}), C_{\tilde{X}} = c] \geq \delta 2^{-d}.
\]

Note that \(\tilde{X}\), in Lemma 3 has only \(C_{\tilde{X}}\) as argument while the “wide hat” \(\tilde{X}\) in Lemma 2 has two arguments, \(C_{\tilde{X}}\) and \(Y^n\). Lemmas 2 and 3 are proved in Section IV-C.

**B. Proof of Theorem 2**

Suppose the source \(p_{XY}\) is \(\mathcal{X}\)-confusable and suppose \(X^n\) is compressed at rate \(R_1 < H(X)\). Let \(\tilde{X}(C_{\tilde{X}}, Y^n)\) be a local decoder for \(X_i\) given \(C_{\tilde{X}}\) and \(Y^n\). Let \(c\) be a realization of \(C_{\tilde{X}}\) and let \(x \in \mathcal{X}\). Using Lemma 2 then Lemma 1, we have

\[
\Pr(\tilde{X}(C_{\tilde{X}}, Y^n) \neq X_i, C_{\tilde{X}} = c) \geq (\Pr[X_i = x, C_{\tilde{X}} = c])^\alpha \times (\Pr[\tilde{X} \neq x, \tilde{C}_{\tilde{X}} = c])^\beta
\]

(3)

for some finite constants \(\alpha\) and \(\beta\). Now, using Lemma 3, we show that for some index \(i \in [n]\), realization \(c\) of \(C_{\tilde{X}}\), and \(x \in \mathcal{X}\), each of the two probability terms on the right-hand side of the inequality (3) is lower bounded by \(2^{-\Theta(d)}\), where \(d = \max_{1 \leq i \leq n} |I_{\tilde{X}}|\). This then implies that

\[
\Pr(\tilde{X}(C_{\tilde{X}}, Y^n) \neq X_i, C_{\tilde{X}} = c) \geq 2^{-\Theta(d)},
\]

and therefore

\[
\Pr(\tilde{X}(C_{\tilde{X}}, Y^n) \neq X_i) \geq 2^{-\Theta(d)},
\]

which yields the desired result.

4More precisely, \(\alpha\) and \(\beta\) are in \((1, \infty)\) (see [37]), but for our purpose the values of \(\alpha\) and \(\beta\) (as functions of \(p_{XY}\)) are irrelevant.
Let us first consider the second term on the right-hand side of the inequality (3). Since \( \hat{X}^n \) has the same distribution as \( X^n \), from Lemma 3 with
\[
\hat{X}_i(C_{I_i} = c) = \arg \max_x \Pr(X_i = x|C_{I_i} = c) \defdot \hat{x}_c,
\]
there exist a constant \( \delta > 0 \), an index \( i \in [n] \), and a realization \( c \in \{0, 1\}^{d(i)} \) of \( C_{I_i} \) with \( d(i) \leq \delta \), such that
\[
\Pr[\hat{X}_i \neq \hat{x}_c, \hat{C}_{I_i} = c] \geq \delta 2^{-d} \geq \delta |\mathcal{X}|^{-d},
\]
(4)
where the second inequality holds since \( |\mathcal{X}| \geq 2 \) (see Section II-A).

For the first probability term on the right-hand side of inequality (3), since \( \hat{X}^n \) and \( X^n \) have the same distribution, from (4) we get
\[
\Pr[C_{I_i} = c] \geq \Pr[X_i \neq \hat{x}_c, C_{I_i} = c] \geq \delta |\mathcal{X}|^{-d}.
\]
(5)
Note also that
\[
\Pr[X_i = \hat{x}_c, C_{I_i} = c] \geq \frac{1}{|\mathcal{X}|},
\]
(6)
for otherwise the probabilities would not sum to one. From (5) and (6), it then follows that
\[
\Pr[X_i = \hat{x}_c, C_{I_i} = c] = \Pr[X_i = \hat{x}_c|C_{I_i} = c] \Pr[C_{I_i} = c] \geq \delta |\mathcal{X}|^{-d-1},
\]
(7)
which establishes the desired claim.

C. Proofs of Lemmas 2 and 3

Proof of Lemma 2: Let \( c \) be a non-zero probability instance of \( C_{I_i} \)—for otherwise there is nothing to prove—and let \( x \in \mathcal{X} \). Let \( \hat{X}_i(C_{I_i}, Y^n) \) be an estimator of \( X_i \) given \( C_{I_i} \) and \( Y^n \). Define
\[
E(X_i) \defdot \begin{cases} 0 & \text{if } X_i = x \\ 1 & \text{if } X_i \neq x \end{cases}
\]
and \( \hat{E}(C_{I_i}, Y^n) \defdot \begin{cases} 0 & \text{if } \hat{X}_i(C_{I_i}, Y^n) = x \\ 1 & \text{if } \hat{X}_i(C_{I_i}, Y^n) \neq x \end{cases} \)
and let \( E^*(C_{I_i}, Y^n) \) be the MAP estimator of \( E(X_i) \) given \( C_{I_i} \) and \( Y^n \). We have for any \( y^n \)
\[
\Pr[\hat{X}_i(C_{I_i}, Y^n) \neq X_i|C_{I_i} = c, Y^n = y^n] \geq \Pr[\hat{E}(C_{I_i}, Y^n) \neq E(X_i)|C_{I_i} = c, Y^n = y^n] \geq \Pr[E^*(C_{I_i}, Y^n) \neq E(X_i)|C_{I_i} = c, Y^n = y^n] = \min \left\{ \Pr[X_i = x|Y^n = y^n, C_{I_i} = c], \Pr[X_i = \tilde{x}|Y^n = y^n, C_{I_i} = c] \right\} \geq \Pr[X_i = x|Y^n = y^n, C_{I_i} = c] \times \Pr[X_i \neq \tilde{x}|Y^n = y^n, C_{I_i} = c] = \Pr[X_i = \tilde{x}|Y^n = y^n, C_{I_i} = c] \times \Pr[\tilde{X}_i \neq \tilde{x}|Y^n = y^n, \tilde{C}_{I_i} = c],
\]
where the last equality holds since \( \tilde{X}^n \) and \( \tilde{X}^n \) have the same distribution. Therefore,
\[
\Pr[\hat{X}_i(C_{I_i}, Y^n) \neq X_i|C_{I_i} = c, Y^n = y^n] \geq \Pr[X_i = x|Y^n = y^n, C_{I_i} = c] \times \Pr[X_i \neq \tilde{x}|Y^n = y^n, \tilde{C}_{I_i} = c] \times \Pr[\tilde{C}_{I_i} = c|Y^n = y^n],
\]
equivalently,
\[
\Pr[\hat{X}_i(C_{I_i}, Y^n) \neq X_i, C_{I_i} = c|Y^n = y^n] \geq \Pr[X_i = x, C_{I_i} = c|Y^n = y^n] \times \Pr[\hat{X}_i \neq \tilde{x}, \hat{C}_{I_i} = c] \times \Pr[\tilde{C}_{I_i} = c|Y^n = y^n] = \Pr[X_i = x, C_{I_i} = c, \tilde{X}_i \neq \tilde{x}, \tilde{C}_{I_i} = c|Y^n = y^n].
\]
(8)
where the last equality holds since \( X - Y - \tilde{X} \) forms a Markov chain. Multiplying both sides of (8) by \( \Pr[Y^n = y^n] \) and summing over \( y^n \) gives
\[
\Pr[\hat{X}_i(C_{I_i}, Y^n) \neq X_i, C_{I_i} = c] \geq \Pr[X_i = x, C_{I_i} = c, \tilde{X}_i \neq \tilde{x}, \tilde{C}_{I_i} = c],
\]
which concludes the proof. ■

Proof of Lemma 3: The converse to Shannon’s lossy source coding theorem implies that if \( R < H(X) \), then there exists a \( \delta = \delta(R) > 0 \) such that
\[
\mathbb{E}d_H(X^n, \tilde{X}^n) = \sum_{i=1}^n \Pr[\hat{X}_i(C_{I_i}) \neq X_i] \geq n\delta,
\]
where \( d_H(X^n, \tilde{X}^n) \) denotes the Hamming distance between \( X^n \) and \( \tilde{X}^n \). Hence,
\[
\delta \leq \Pr[\hat{X}_i(C_{I_i}) \neq X_i]
\]
for at least one index \( i \in [n] \).

Expanding the right-hand side and assuming a worst-case local decodability of \( \tilde{c} \in [n] \), we have
\[
\delta \leq \Pr[\hat{X}_i(C_{I_i}) \neq X_i] = \sum_{c \in \{0, 1\}^{d(i)}} \Pr[\hat{X}_i(C_{I_i}) \neq X_i, C_{I_i} = c] 
\leq 2^d \max_{c \in \{0, 1\}^{d(i)}} \Pr[\hat{X}_i(C_{I_i}) \neq X_i, C_{I_i} = c]
\]
which concludes the proof. ■

V. PROOF OF THEOREM 3

If \( p_{X|Y} \) is not \( \mathcal{X} \)-confusable, then there exists \( x_1, x_2 \in \mathcal{X} \) such that, for any \( y \in \mathcal{Y} \), either \( p_{X|Y}(x_1, y) > 0 \) or \( p_{X|Y}(x_2, y) > 0 \) (recall that without loss of generality \( p_Y(y) > 0 \) for any \( y \in \mathcal{Y} \)). Therefore, conditioned on \( X \in \{x_1, x_2\} \), the knowledge of \( Y \) reveals \( X \).

Let \( \mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\} \), and suppose without loss of generality that \( (x_1, x_2) = (1, 2) \). Define the new source \( U^n \) over the reduced alphabet \( \{2, 3, \ldots, |\mathcal{Y}|\} \) as
\[
U_i = \begin{cases} 2 & \text{if } X_i \in \{1, 2\} \\ X_i & \text{if } X_i \notin \{1, 2\}. \end{cases}
\]
Clearly, \( H(U) < H(X) \) and \( (U, Y) \) determines \( (X, Y) \). We can therefore compress \( U^n \) and \( Y^n \) independently at rates \( R_1 = H(U) < H(X) \) and \( R_2 = H(Y) \) using the compressors of [11], [12] to achieve strong locality.
VI. PROOF OF THEOREM 4

We now design a scheme that achieves the following: For any fixed \( \delta > 0 \) and \((R_1, R_2)\) within the optimal rate region, 

- The sequences \( (X^n, Y^n) \) are independently compressed to rates \((R_1 + \delta, R_2 + \delta)\) respectively. 
- For any \( i \in [n] \) and \( 1 > \eta > 2^{-20|\log n|} \), specified at the receiver, the local decoder probes \( \text{poly}(\log(1/\eta)) \) compressed bits, and outputs \((\hat{X}_i, \hat{Y}_i)\) which satisfies 
  \[
  p_{\text{loc}}^i = \Pr[(\hat{X}_i, \hat{Y}_i) \neq (X_i, Y_i)] \leq \eta. 
  \]

Our coding scheme is inspired by that in [31], and is a hierarchical compression scheme. The compressed bits consist of various blocks that are spread across multiple “levels” \( 1 \leq \ell \leq \ell_{\text{max}} \). The compressed bits at level \( \ell = 0 \) is obtained by applying the concatenation scheme defined in Section II-B2 with \( b = O(1) \). This guarantees that any pair of source symbols can be recovered with \( 2^{-\Theta(b)} = O(1) \) probability of error. The compressed bits at higher levels \( \ell \geq 1 \) can be viewed as additional refinement bits that are probed only when we desire a lower probability of error. By probing blocks corresponding to higher levels, we obtain a more reliable estimate of \( (X_i, Y_i) \). The compressed blocks at level \( \ell \) are obtained by using a random binning scheme applied to blocks of size \( n_{\ell} \), where \( n_{\ell} \) is growing superexponentially with \( \ell \). However, the rates for higher levels is chosen to decay exponentially with \( \ell \). The key challenge is to choose the parameters carefully so that the additional bits corresponding to higher levels provide a negligible contribution to the overall compression rates.

1) Parameters: We choose\(^5\) a sufficiently small \( \varepsilon_0 > 0 \), positive integers 
\[
  b_0 = n_0 \tag{9}
\]
which are constants independent of \( n \), and 
\[
  k_0^{(1)} = \lceil (R_1 + \varepsilon_0) b_0 \rceil \quad \text{and} \quad k_0^{(2)} = \lceil (R_2 + \varepsilon_1) b_1 \rceil \tag{10}
\]
such that the probability of error of a Slepian-Wolf code [38] for sequences of length \( b_0 \) satisfies 
\[
  \Pr[(\hat{X}^{b_0}, \hat{Y}^{b_0}) \neq (X^{b_0}, Y^{b_0})] \leq 2^{-\beta n_{b_0}} \leq \delta^\ell
\]
where \( \beta > 0 \) depends on \( p_{XY}, R_1, R_2 \) only and \( 0 < \delta^\ell < 1 \) is a parameter that determines an upper bound on the probability of local decoding error that can be achieved. 

For each \( \ell = 1, 2, \ldots, \ell_{\text{max}} \), define 
\[
  \varepsilon_\ell = \varepsilon_{\ell-1}/2 = \varepsilon_0/2^\ell \\
  b_\ell = 16b_{\ell-1} = 16^\ell b_0 \\
  n_\ell = b_\ell n_{\ell-1} = 4^{\ell(\ell+1)} b_{\ell-1}^{\ell+1} \\
  k^{(1)}_\ell = \varepsilon_\ell n_\ell \left( \beta + |\mathcal{X}| + b_\ell n_\ell \log \frac{2^\ell}{\varepsilon_0} \right) \\
  k^{(2)}_\ell = \varepsilon_\ell n_\ell \left( \beta + |\mathcal{Y}| + b_\ell n_\ell \log \frac{2^\ell}{\varepsilon_0} \right) \tag{11}
\]

2) Codes for Various Levels: At the heart of our construction is a multilevel random binning argument that can be described by a sequence of random codes \( C_0, C_1, \ldots, C_{\ell_{\text{max}}} \).

The code \( C_\ell \) at level \( \ell \) consists of two encoders \( \Phi_\ell : X^{n_\ell} \rightarrow \{0, 1\}^{k_\ell^{(1)}} \) and \( \Psi_\ell : Y^{n_\ell} \rightarrow \{0, 1\}^{k_\ell^{(2)}} \), where \( k_\ell^{(1)} \) and \( k_\ell^{(2)} \) are as defined previously. For each \( X^{n_\ell} \in \mathcal{X}^{n_\ell} \), we assign a codeword \( \Phi_\ell(X^{n_\ell}) \) drawn uniformly at random from \( \{0, 1\}^{k_\ell^{(1)}} \). Similarly, for each \( Y^{n_\ell} \in \mathcal{Y}^{n_\ell} \), we assign a codeword \( \Psi_\ell(Y^{n_\ell}) \) drawn uniformly at random from \( \{0, 1\}^{k_\ell^{(2)}} \). For any \( u_\ell^{(i)} \in \{0, 1\}^{k_\ell^{(i)}} \) and \( v_\ell^{(i)} \in \{0, 1\}^{k_\ell^{(i)}} \), we have 
\[
  \Pr[\Phi_\ell(u_\ell^{(1)}) = v_\ell^{(1)}] = 2^{-k_\ell^{(1)}} \quad \text{and} \quad \Pr[\Psi_\ell(u_\ell^{(2)}) = v_\ell^{(2)}] = 2^{-k_\ell^{(2)}}. 
\]

The codes are known to the decoder and the respective encoders.

3) Encoder: We now describe the encoding of the sequences \( X^n \) and \( Y^n \). Let us suppose that user 1 has \( X^n \) and user 2 has \( Y^n. \)

The codeword generated by each user comprises of various blocks spread over multiple levels, and the encoding is done independently at each level. Consider any level \( 0 \leq \ell \leq \ell_{\text{max}} \). Each user partitions its source sequence into blocks of \( n_\ell \) symbols each. For \( i = 1, 2, \ldots, n/n_\ell \), define the \( i \)-th (source) block at level \( \ell \) to be \( X^{n_\ell}(\ell, i) = X^{n_\ell}_{(i-1)n_\ell+1} \) and \( Y^{n_\ell}(\ell, i) = Y^{n_\ell}_{(i-1)n_\ell+1} \). Let \( U^{(0)}_\ell(\ell, i) \equiv \Phi^\ell(X^{n_\ell}(\ell, i)) \) be the \( i \)-th level-\( \ell \) codeword for user 1, and \( U^{(1)}_\ell(\ell, i) \equiv \Psi^\ell(Y^{n_\ell}(\ell, i)) \) be the \( i \)-th level-\( \ell \) codeword for user 2.

The codeword for \( X^n \) is obtained by taking the concatenation of all level \( \ell \) codewords for \( 0 \leq \ell \leq \ell_{\text{max}} \). This is equal to \( (U^{(0)}_\ell(\ell, i) : 0 \leq \ell \leq i, 1 \leq i \leq n/n_\ell) \). Similarly, the codeword for \( Y^n \) is equal to \( (U^{(1)}_\ell(\ell, i) : 0 \leq \ell \leq i, 1 \leq i \leq n/n_\ell) \).

An illustration of the encoding process is provided in Fig. 1. The level-0 codewords correspond to the concatenation scheme, and most of the entropy of the compressed sequence lies in the level-0 codewords. The level \( \ell \geq 1 \) codewords give extra information that allow us to reduce the probability of local decoding error. The rates \( k_\ell^{(1)}/n_\ell \) and \( k_\ell^{(2)}/n_\ell \) are exponentially decaying functions of \( \ell \), and the overall sum rates of all the \( \ell = 1, 2, \ldots, \ell_{\text{max}} \) codewords is negligible.

4) Local Decoder: The local decoder takes two parameters as input: a location \( i \in [n] \), and \( \ell_d \in [0, 1, 2, \ldots, \ell_{\text{max}}] \). The first parameter specifies which \( (X_i, Y_i) \) the decoder wishes to recover. The second parameter specifies the number of bits to probe (which decides the probability of error). For a specified \( \ell_d \), the local decoder probes \( 2^{O(\ell_d^2)} \) compressed bits, and the probability of error is \( 2^{-\Theta(\ell_d^2)} \). This statement will be made more precise shortly.

The decoder works by probing compressed bits up to level \( \ell_d \) as follows:

- The decoder first finds which level-\( \ell_d \) chunk the desired location \( i \) lies in. In other words, it sets \( i_d = \lceil i/n_{\ell_d} \rceil \). It then reads all the compressed chunks up to level \( \ell_d \) corresponding to the symbols \( X^{n_{\ell_d}}(i_d) \).
- The decoder now iteratively improves its estimate by processing the compressed bits from level 0 to level \( \ell_d \) as follows:
  - At level 0, the decoder uses the Slepian-Wolf decoder to obtain the level-0 estimates of \( X^{n_{\ell_d}}(i_d) \). Call this estimate \( \hat{X}^{n_{\ell_d}}(i_d, 0) \).

\(^5\)Since we only aim to get order-optimal results, we have not attempted to optimize over the various parameters.
Fig. 1. A depiction of the encoding structure for user 1. For ease of illustration, we have chosen $\ell_{max} = 2$, $b_2 = 4b_1$, and $b_1 = 2b_0$. To obtain the codewords at level $\ell$, the source $X^n$ is partitioned into blocks of $n_\ell$ symbols each, and the $j$th level-$\ell$ codeword $U^{(j)}_\ell (\ell, j) = \Phi_j(X^{n_\ell}(\ell, j))$. The overall codeword is the concatenation of all the level $\ell = 0, 1, \ldots, \ell_{max}$ codewords. A similar encoding process is performed by user 2 but using $\Psi_j$. Considering decoding $X^{n_0}(0, 2)$, the local decoder may choose to probe $U^{(j)}_0 (0, 2)$, $V^{(k)}_0 (0, 2)$ and use the standard Slepian-Wolf (joint typicality) decoder. If a lower probability of error is required, then it additionally probes $U^{(l)}_\ell (\ell, 0)$, $V^{(k)}_\ell (\ell, b)$ for $\ell = 1$, or $\ell = 1, 2$ depending on the target probability of error. The additional bits are then used to refine the estimate of $X^{n_0}(0, 2)$.

For all subsequent levels $\ell \in \{1, 2, \ldots, \ell_d\}$, the decoder does the following. Suppose that for some $i$, we want to estimate $(X^{n_i}(\ell, i), Y^{n_i}(\ell, i))$ assuming that we already have the level $\ell - 1$ estimate. The decoder outputs $(X^{n_i}(\ell, i), Y^{n_i}(\ell, i))$ if $(X^{n_i}, Y^{n_i})$ is the unique pair of sequences which match $U^{(j)}_{\ell-1} (\ell, i), V^{(k)}_{\ell-1} (\ell, i)$ and also match at least $(1 - \varepsilon_1) b_\ell$ of the level-$(\ell - 1)$ estimated blocks. If there is no such unique sequence, then the level-$\ell$ decoder outputs the zero sequence.

In Lemma 4, we derive an upper bound on the local decoding error probability. We then derive an upper bound on the probability of decoding error in Lemma 5. Combining the two gives us Theorem 4.

**Lemma 4.** For any given parameters $(i, \ell_d)$, the number of bits probed by the local decoder is

$$\bar{d}(\ell_d) \leq n_d + 1 + 4d(\ell_d + 1)(R_1 + R_2 + \gamma_1 \varepsilon_0) \leq 2\gamma_2 \ell_d$$

where $\gamma_1$ is a constant that only depends on $\rho_{XY}$, $R_1$, $R_2$, while $\gamma_2$ may depend on $\rho_{XY}$, $R_1$, $R_2$, $\varepsilon_0$, $b_0$.\n
**Proof:** The total number of compressed bits probed is equal to

$$\bar{d}(\ell_d) = \sum_{\ell=0}^{\ell_d} \frac{n_\ell}{n_\ell} (k^{(1)}_\ell + k^{(2)}_\ell)$$

$$\leq n_d R_1 + R_2 + \gamma_1 \varepsilon_0 + b_0^{(\ell_d + 1)}(R_1 + R_2 + \gamma_1 \varepsilon_0)$$

where $\gamma_1$ is a constant that only depends on $\rho_{XY}, R_1, R_2$.

**Lemma 5.** For any given parameters $(i, \ell_d)$, the probability of error of decoding $X^{n_d}, Y^{n_d}$ after decoding up to level $\ell_d$ is upper bounded as follows

$$p_{\ell_d} \leq 2^{-\beta(\varepsilon_0 b_0)^{\frac{1}{\ell_d + 1} + \frac{1}{2}}} = 2^{-2^c \ell_d \log(d)}$$

for a suitable constant $\beta > 0$.

**Proof:** We will derive the bound by obtaining an upper bound on $P_{\ell_d}$ in terms of $P_{\ell_{d-1}}$. For $\ell = 0$, we know that

$$p_{\ell_d} \leq 2^{-\beta(\varepsilon_0 b_0)}$$

for a suitable constant $\beta > 0$.

For decoding at level $\ell > 1$, there are two possible error events:

1) Event $E_1$: More than $\varepsilon_1 b_\ell$ blocks were decoded incorrectly at level $\ell - 1$.

2) Event $E_2$: There is an incorrect pair of sequences $(\tilde{X}^{n_\ell}, \tilde{Y}^{n_\ell})$ that has the same level-$\ell$ hash/codeword as the true sequence and matches the $(\ell - 1)$-level decoded sequence on at least $(1 - \varepsilon_1) b_\ell$ blocks.
The overall probability of error is then
\[ P_e^{(k)} \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2 | \mathcal{E}_1^c]. \]

We will bound the two terms separately. For the first term, observe that
\[ \Pr[\mathcal{E}_1] \leq \left( \frac{b_{k}}{\epsilon_{0}b_{k}} \right) \left( \frac{p_{X}^{(k-1)}}{\epsilon_{0}} \right)^{b_{k}} \]
\[ \leq \left( \frac{e^{\beta(\epsilon_{0}b_{k})^{2}(1-\epsilon_{0})^{2}}}{\epsilon_{0}/2} \right)^{\epsilon_{0}b_{k} \delta_{k}} \]

where in the last step, we have assumed that \( P_{e}^{(k-1)} \leq 2^{-\beta(\epsilon_{0}b_{k})^{2}(1-\epsilon_{0})^{2}} \). Rewriting the right-hand side, we get
\[ \Pr[\mathcal{E}_1] \leq \exp \left( -\beta(\epsilon_{0}b_{k})^{\ell} + \epsilon_{0}b_{k} \delta_{k} \log \left( \frac{e^{\delta_{k}}}{\epsilon_{0}} \right) \right). \]

Since \( \epsilon_{0}b_{k} \) is large enough, the absolute value of the first term in the exponent is at least twice that of the second. Therefore,
\[ \Pr[\mathcal{E}_1] \leq \exp \left( -\beta(\epsilon_{0}b_{k})^{\ell} + \epsilon_{0}b_{k} \delta_{k} \log \left( \frac{e^{\delta_{k}}}{\epsilon_{0}} \right) \right) \leq 2^{-2\beta(\epsilon_{0}b_{k})^{\ell} + \epsilon_{0}b_{k} \delta_{k}}. \]

To compute the probability of the second error event, let us define \( \mathcal{E}_2^{A} \) (resp. \( \mathcal{E}_2^{B} \)) to be the event that there is an incorrect sequence \( \hat{x}^m \) (resp. \( \hat{x}^m \)) that has the same hash as the true sequence and matches the \((\ell-1)\)-level decoded sequence on at least \((1-\epsilon_{0})b_{k}\) blocks. We have,
\[ \Pr[\mathcal{E}_2^{A} | \mathcal{E}_1^c] \leq \left( \frac{b_{k}}{\epsilon_{0}b_{k}} \right) \left| \hat{x}_{1}^{m} \right| n_{i}^{-k_{(1)}} \]
\[ \leq \left( \frac{e^{\beta(\epsilon_{0}b_{k})^{\ell}}}{\epsilon_{0}/2} \right) \left| \hat{x}_{1}^{m} \right| n_{i}^{-k_{(1)}}. \]

Substituting for \( k_{(1)} \) in the above and simplifying, we get
\[ \Pr[\mathcal{E}_2^{A} | \mathcal{E}_1^c] \leq 2^{-\beta \epsilon_{0} n_{i}} \leq 2^{-\beta \epsilon_{0} b_{k}^{\ell} + 1/2} \]

Similarly,
\[ \Pr[\mathcal{E}_2^{B} | \mathcal{E}_1^c] \leq 2^{-\beta \epsilon_{0} n_{i}} \leq 2^{-\beta \epsilon_{0} b_{k}^{\ell} + 1/2}. \]

Combining (12), (13) and (14), we get
\[ P_{e}^{(k)} \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2^{A} | \mathcal{E}_1^c] + \Pr[\mathcal{E}_2^{B} | \mathcal{E}_1^c] \]
\[ \leq 2^{-\beta \epsilon_{0} b_{k}^{\ell} + 1/2}, \]
which completes the proof.

VII. EXTENSION TO \( k > 2 \) SOURCES

We first extend Theorem 2 to a \( k \)-source distribution \( p_{X_{1},...,X_{k}} \) defined over alphabet \( X_{1} \times \cdots \times X_{k} \). Let the source be \( X_{1} \)-confusable if for every \( x_{1}, x'_{1} \in X_{1} \), there exist \( (x_{2}, x_{3}, \ldots, x_{k}) \) for which \( p_{X_{1},...,X_{k}}(x_{1}, \ldots, x_{k}) > 0 \) and \( p_{X_{1},...,X_{k}}(x'_{1}, \ldots, x'_{k}) > 0 \). Observe that this condition holds if and only if for every \( x_{1}, x'_{1} \in X_{1} \), there exist an index \( i \geq 2 \) and \( x_{i} \in X_{i} \)

for which \( p_{X_{1},X_{i}}(x_{1}, x_{i}) > 0 \) and \( p_{X_{1},X_{i}}(x'_{1}, x_{i}) > 0 \). Now if we repeat the same line of arguments as for the proof of Theorem 2, but with the side information \( Y_{n} \) replaced by all sources except \( X_{1}^{n} \), that is \( X_{2}^{n}, \ldots, X_{k}^{n} \), we get:

**Theorem 5 (Confusable, \( k \geq 2 \) Sources):** Suppose source \( p_{X_{1},...,X_{k}} \) is \( X_{1} \)-confusable. If \( X_{1}^{n} \) is compressed at rate \( R_{1} < H(X_{1}) \), then
\[ \max_{1 \leq i \leq n} \Pr[\hat{X}_{1}^{i}(C_{T_{i}}(X_{1}^{n}), X_{2}^{n}, \ldots, X_{k}^{n}) \neq X_{1}^{i}] \geq 2^{-\Omega(\delta)}, \]

where \( \hat{X}_{1}^{i}(C_{T_{i}}(X_{1}^{n}), X_{2}^{n}, \ldots, X_{k}^{n}) \) is any estimator of the \( i \)-th symbol of source \( X_{1}^{n} \) given at most \( \delta \) components \( C_{T_{i}}(X_{1}^{n}) \) of \( C_{T_{i}}(X_{2}^{n}) \) and \( (X_{2}^{n}, \ldots, X_{k}^{n}) \).

Similarly, Theorem 3 immediately generalizes to

**Theorem 6 (Non-Confusable, \( k \geq 2 \) Sources):** Suppose source \( p_{X_{1},...,X_{k}} \) is not \( X_{1} \)-confusable. Then, it is possible to achieve strong locality at some \( R_{1} < H(X_{1}) \) and \( R_{i} = H(X_{i}), \ i \in \{2, \ldots, k\} \).

The coding scheme of Section VI easily extends to more than two sources, with the same encoding scheme for each source, and an identical local decoder:

**Theorem 7 (Hierarchical Coding Scheme, \( k \geq 2 \) Sources):** For any \( (R_{1}, R_{2}, \ldots, R_{k}) \) in the interior of the Slepian-Wolf rate region, there exists a rate \( (R_{1}, R_{2}, \ldots, R_{k}) \) distributed compression scheme such that for every \( 1 > \eta > 2^{-\Omega(\log n)} \), the local decoder achieves \( \hat{d} = \text{poly}(\log(1/\eta)) \) and \( P_{e}^{\text{loc}} \leq \eta \).

VIII. CONCLUDING REMARKS

In contrast with the single source set up, we showed that for multiple sources lossless compression and strong locality can generally not be accommodated. For confusable sources, for strong locality to hold all sources must be compressed at rates above their respective entropies. On the other hand, if the distribution is not confusable, an arguably peculiar situation, strong locality may hold even if compression rates are below individual entropies. For this case, the characterization of all rate pairs for which strong locality can be achieved remains an open problem.

Our compression scheme is able to achieve \( \hat{d} = \text{poly}((1/\eta)) \) for any target probability of local decoding error \( \eta \) specified at the decoder. Note that from our lower bound, \( \hat{d} = \Omega((1/\eta)) \) and our scheme is suboptimal by a polynomial factor. Designing an improved scheme that achieves this lower bound is left as future work.

In this paper, we only considered the problem of local decidability in the context of distributed compression. One may also require provisioning of local substitutions/insertions/deletions of source symbols in the compressed domain. This is an interesting problem that warrants more attention.

---

**REFERENCES**

[1] S. Vatedka, V. Chandar, and A. Tchamkerten, “Locally decodable Slepian-Wolf compression,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2022, pp. 1430–1435.

[2] D. S. Pavlović, T. Weissman, and G. Yona, “The human genome contracts again,” Bioinformatics, vol. 29, no. 17, pp. 2199–2202, 2013.
[3] D. Pavličin, T. Weissman, and G. Mahly, “The quest to save genomics: Unless researchers solve the looming data compression problem, biomedical science could stagnate,” IEEE Spectr., vol. 55, no. 9, pp. 27–31, Sep. 2018.

[4] C. L. P. Chen and C.-Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on big data,” Inf. Sci., vol. 275, pp. 314–347, Aug. 2014.

[5] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, “The rise of ‘big data’ on cloud computing: Review and open research issues,” Inf. Syst., vol. 47, pp. 98–115, Jan. 2015.

[6] M. P. Ball et al., “A public resource facilitating clinical use of genomes,” Proc. Nat. Acad. Sci., vol. 109, no. 30, pp. 11920–11927, 2012.

[7] K. Walter et al., “The UK10K project identifies rare variants in health and disease,” Nature, vol. 526, no. 7571, pp. 82–90, 2015.

[8] E. E. Schadt, M. D. Linderman, J. Sorenson, L. Lee, and G. P. Nolan, “Computational solutions to large-scale data management and analysis,” Nat. Rev. Genet., vol. 11, no. 9, pp. 647–657, 2010.

[9] J. Ziv and A. Lempel, “A universal algorithm for sequential data compression,” IEEE Trans. Inf. Theory, vol. 23, no. 3, pp. 337–343, May 1977.

[10] J. Ziv and A. Lempel, “Compression of individual sequences via variable-rate coding,” IEEE Trans. Inf. Theory, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

[11] A. Mazumdar, V. Chaudhary, and G. W. Wornell, “Local recovery in data compression for general sources,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2015, pp. 2984–2988.

[12] K. Tatwawadi, S. Bidokhti, and T. Weissman, “On universal compression with constant random access,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2018, pp. 891–895.

[13] M. Pătraşcu, “Succinct,” in Proc. 49th Annu. IEEE Symp. Found. Comput. Sci., 2008, pp. 305–313.

[14] Y. Dodis, M. Pătraşcu, and M. Thorup, “Changing base without losing space,” in Proc. 42nd ACM Symp. Theory Comput. (STOC), 2010, pp. 593–602.

[15] J. I. Munro and Y. Nekrich, “Compressed data structures for dynamic sequences,” in Proc. Algorithms-ESA, 2015, pp. 891–902.

[16] R. Raman and S. S. Rao, “Succinct dynamic dictionaries and trees,” in Proc. Int. Colloq. Automata Lang. Program., 2003, pp. 357–368.

[17] V. Chaudhary, D. Shah, and G. W. Wornell, “A locally encodable and decodable compressed data structure,” in Proc. 47th Annu. Allerton Conf. Commun. Control Comput., 2009, pp. 613–619.

[18] V. B. Chaudhary, “Sparse graph codes for compression, sensing and secrecy,” Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 2010.

[19] S. Kreft and G. Navarro, “LZ77-like compression with fast random access,” in Proc. Data Compression Conf. (DCC), 2010, pp. 239–248.

[20] A. Dutta, R. Levi, D. Ron, and R. Rubinfeld, “A simple online competitive adaptation of Lempel-Ziv compression with efficient random access support,” in Proc. Data Compression Conf. (DCC), 2013, pp. 113–122.

[21] P. Bille, G. M. Landau, R. Raman, K. Sadakane, S. R. Satti, and O. Weimann, “Random access to grammar-compressed strings,” in Proc. 22nd Annu. ACM-SIAM Symp. Discrete Algorithms (SODA), 2011, pp. 373–389.

[22] E. Viola, O. Weinstein, and H. Yu, “How to store a random walk,” in Proc. 31st Annu. ACM-SIAM Symp. Discrete Algorithms (SODA), 2020, pp. 426–445.

[23] K. Sadakane and R. Grossi, “Squeezing succinct data structures into entropy bounds,” in Proc. 17th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA), 2006, pp. 1230–1239.

[24] R. González and G. Navarro, “Statistical encoding of succinct data structures,” in Proc. Annu. Symp. Comb. Pattern Matching, 2006, pp. 294–305.

[25] P. Ferragina and R. Venturini, “A simple storage scheme for strings achieving entropy bounds,” in Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA), 2007, pp. 690–696.

[26] A. Makhdoumi, S.-L. Huang, M. Medard, and Y. Polyanskiy, “On locally decodable source coding,” in Proc. IEEE Int. Conf. Commun. (ICC), London, U.K., Jun. 2015, pp. 4394–4399.

[27] P. K. Nicholson, V. Raman, and S. S. Rao, “A survey of data structures in the bitprobe model,” in Space-Efficient Data Structures, Streams, and Algorithms. Berlin, Germany: Springer, 2013, pp. 303–318.

[28] A. Makhdoumi, S.-L. Huang, M. Medard, and Y. Polyanskiy, “On locally decodable source coding,” 2013, arXiv:1308.5239.

[29] H. Buhrman, P. B. Miltersen, J. Radhakrishnan, and S. Venkatesh, “Are bitvectors optimal?” SIAM J. Comput., vol. 31, no. 6, pp. 1723–1744, 2002.

[30] A. Pananjady and T. A. Courtade, “The effect of local decodability constraints on variable-length compression,” IEEE Trans. Inf. Theory, vol. 64, no. 4, pp. 2593–2608, Apr. 2018.

[31] S. Vatedka and A. Tchamkerten, “Local decode and update for big data compression,” IEEE Trans. Inf. Theory, vol. 66, no. 9, pp. 5790–5805, Sep. 2020.

[32] S. Vatedka, V. Chaudhary, and A. Tchamkerten, “O (log log n) worst-case local decoding and update efficiency for data compression,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 2371–2376.

[33] R. Vestergaard, Q. Zhang, and D. E. Lucani, “Enabling random access in universal compressors,” in Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), 2021, pp. 1–6.

[34] S. Kamparaju, S. Mastan, and S. Vatedka, “Low-complexity compression with random access,” in Proc. IEEE Int. Conf. Signal Process. Commun. (SPCOM), 2022, pp. 1–5.

[35] D. Slepian and J. Wolf, “Noiseless coding of correlated information sources,” IEEE Trans. Inf. Theory, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[36] I. Csiszár and J. Körner, “Towards a general theory of source networks,” IEEE Trans. Inf. Theory, vol. IT-26, no. 2, pp. 155–165, Mar. 1980.

[37] S. Kamath, “Computational solutions to large-scale data management and analysis,” IEEE Trans. Inf. Theory, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

[38] V. V. Vazirani and M. Yannakakis, “Space-Efficient Data Structures, Streams, and Algorithms. Berlin, Germany: Springer, 2013, pp. 303–318.

[39] A. Makhdoumi, S.-L. Huang, M. Medard, and Y. Polyanskiy, “On locally decodable source coding,” 2013, arXiv:1308.5239.

[40] H. Buhrman, P. B. Miltersen, J. Radhakrishnan, and S. Venkatesh, “Are bitvectors optimal?” SIAM J. Comput., vol. 31, no. 6, pp. 1723–1744, 2002.

[41] A. Pananjady and T. A. Courtade, “The effect of local decodability constraints on variable-length compression,” IEEE Trans. Inf. Theory, vol. 64, no. 4, pp. 2593–2608, Apr. 2018.