Malleable Coding with Fixed Segment Reuse

Julius Kusuma, *Member, IEEE,* Lav R. Varshney, *Graduate Student Member, IEEE,*
and Vivek K Goyal, *Senior Member, IEEE*

Abstract

Storage area networks, remote backup storage systems, and similar information systems frequently modify stored data with updates from new versions. In these systems, it is desirable for the data to not only be compressed but to also be easily modified during updates. A malleable coding scheme considers both compression efficiency and ease of alteration, promoting some form of reuse or recycling of codewords. Malleability cost is the difficulty of synchronizing compressed versions, and malleable codes are of particular interest when representing information and modifying the representation are both expensive. We examine the trade-off between compression efficiency and malleability cost measured with respect to the length of a reused prefix portion. The region of achievable rates and malleability is formulated as an information-theoretic optimization and a single-letter expression is provided. Relationships to coded side information and common information problems are also established.

Index Terms

common information, concurrency control, data compression, distributed databases, multiterminal source coding, side information

This work was supported in part by the NSF Graduate Research Fellowship, Grant CCR-0325774, and Grant CCF-0729069.

J. Kusuma is with Schlumberger Technology Corporation, Sugar Land, TX 77478 USA (e-mail: kusuma@alum.mit.edu).

L. R. Varshney and V. K. Goyal are with the Department of Electrical Engineering and Computer Science and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: lrv@mit.edu, vgoyal@mit.edu). L. R. Varshney is additionally with the Laboratory for Information and Decision Systems.

I. Introduction

Conventional data compression uses a small number of compressed-domain symbols but otherwise picks the symbols without care. This carelessness renders codewords utterly disposable; little can be salvaged when the source data changes even slightly. Such data compression is concerned only with reducing the length of coded representations. In this paper and a companion paper with a distinct formulation [1], we adopt the mantra of the green age, *reduce, reuse, recycle*. We formulate problems motivated not only by reduction of representation length but also by the reuse or recycling of compressed data when the source sequence to be coded changes.

In Shannon's original formulation of asymptotically lossless block coding,[1] "the high probability group is coded in an *arbitrary* one-to-one way" into an index set of the appropriate size. This arbitrariness may seem to impede reuse, but it also suggests that many codes are equally good for compression, and one may choose amongst them to optimize a reuse criterion.[2] One may also allow suboptimal compression to improve reuse; this trade-off, under a specific model of reuse, is the focus of this paper.

Moving toward formalizing, suppose that after compressing a random source sequence X_1^n, it is modified to become a new source sequence Y_1^n according to a memoryless editing process $p_{Y|X}$. A *malleable coding* scheme preserves some portion of the codeword of X_1^n and modifies the remainder into a new codeword from which Y_1^n may be decoded reliably using the same deterministic codebook.

There are several possible notions of preserving a portion of the codeword of X_1^n. Here we concentrate on a *malleability cost* defined through the reuse of a fixed part of the old codeword in generating a codeword for Y_1^n. We call this *fixed segment reuse* since a segment is cut from the codeword for X_1^n and reused as part of the codeword for Y_1^n. Without loss of generality, the fixed portion can be taken to be the beginning of the codeword, so the new codeword is a fixed prefix followed by a new suffix.

The fixed reuse formulation is suitable for applications where the update information (new suffix) must be transmitted through a rate-limited communication channel. If the locations of changed symbols were arbitrary, the locations would also need to be communicated, communication which may be prohibitively costly. This formulation is also suitable for information storage systems that use linked lists such as the FAT and NTFS systems. A contrasting scenario is for a cost to be incurred when a symbol is changed in value, regardless of its location. We studied this in [1].

[1] From [2] with emphasis added.

[2] The arbitrariness of code mappings have also been exploited in redundancy-free methods for joint source channel coding [3] and for modulation [4], in a manner related to [1].

Our main result is a characterization of achievable rates as a single-letter expression. To the best of our knowledge, this is among the first works connecting problems of information storage—communication across time—with problems in multiterminal information theory. We relate the fixed reuse problem to several previously-studied problems in multiterminal information theory, some of which are exploited in this work. In particular, a connection to the Gács–Körner common information shows that a large malleability cost must be incurred if the rates for the two versions are required to be near entropy.

The remainder of the paper is organized as follows. Motivations from engineering practice in areas such as database management and network information storage are given in Section II. Section III then provides a formal problem statement for malleable coding with fixed segment reuse. The region describing the trade-off between the rates for the original codeword, for the reused portion, and for the new codeword is the main object of study. Section III-B uses an implicit Markov property to simplify the analysis of the rate–malleability region and Section III-C describes two easily achieved points. Using a random coding argument, Theorem 1 in Section IV gives an achievable rate–malleability region in terms of an auxiliary random variable. There is also a matching converse. Section IV-B looks at the auxiliary random variable in detail; Theorem 2 is a partial characterization of the unknown auxiliary random variable when there is a sufficient statistic for the new version based on the old version. Section V connects this malleable coding problem to other problems in multiterminal information theory. Section VI closes the paper, drawing comparison to the problem of designing side information.

II. Background

Our study of malleable coding is motivated by information systems that store frequently-updated documents. In such systems, storage costs include not only the average length of the coded signal, but also costs in updating. We describe these systems and some of their applications.

In information technology infrastructures, there is often a separation between computer hosts used to process information and storage devices used to store information. Storage area networks (SAN) and network-attached storage (NAS) are two technologies that transfer data between hosts and storage elements. SAN and NAS systems comprise a communication infrastructure for physical connections and a management infrastructure for organizing connections, storage elements, and computers for robust and efficient data transfers [5], [6]. Grid computing and distributed storage systems also display similar distributed caching [7], [8]. Even within single computers, updating caches within the memory hierarchy involves data transfers among levels [9].

Data may be dynamic, being updated or edited after some time. Separate data streams may be generated,

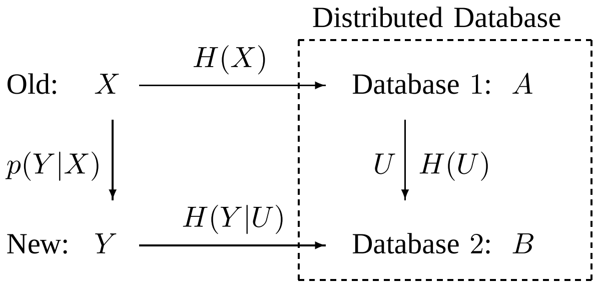

Fig. 1. Distributed database access.

but the contents may differ only slightly [10]–[13]. Moreover, old versions of the stream need not be preserved. Examples include the storage of a computer file backup system after a day's work or graded homework in distance learning [14]. Correlations among versions differentiates malleable coding from write-efficient memories [15], where messages are assumed independent; see [1] for further contrasts.

Storage of communication transcripts in email hosting services such as GMail provides another area where different versions of snippets of text are stored in one common access point. The problem there is made more interesting by the presence of a large number of users who have created different modifications of original shared sources. We do not deal with such problems explicitly.

Systems such as SAN and NAS have complicated interplays between storage and transmission. Current technological trends in transmission and storage technologies show that transmission capacity has grown more slowly than disk storage capacity [7]. Hence "new" representation symbols may be more expensive than "old" representation symbols, suggesting that *reuse* may be more economical than *reduce*.

Recent advances in biotechnology have demonstrated storage of artificial messages in the DNA of living organisms [16]; such systems provide another motivating application. Certain biotechnical editing costs correspond to the malleability costs defined for fixed reuse, as detailed in [1].

Here we describe several scenarios where malleable coding is applicable. Consider the topology given in Fig. 1. The first user has stored a codeword A for document X in database 1. Now the second user, who has a copy of X, modifies it to create Y. The second user wants to save the new version to the information system, but since the users are separated, database 2 rather than database 1 serves this user. Transmission costs for different links may be different. The natural problem is to minimize the total cost needed to create a codeword B at database 2 that losslessly represents Y.

Consider two users who both have access to a distributed database system that stores several copies of the first user's document on different media at different locations. Due to proximity considerations, the users will access the document from different physical stores. Suppose that the first user downloads and edits her document and then wishes to send the new version to the second user. There are different ways to accomplish this. The first user can send the entire new version to the second user or the second user can download the old version from his local store and require that the first user only send the modification. In the former scheme, the cost of transmission is borne entirely by the link between the users, rendering distributed storage pointless. In the latter scheme, there is a trade-off between the rate at which the second user downloads the original version from the database system and the rate at which the first user communicates the modification.

Even in a single user scenario, there may be similar considerations. The first user may simply wish to update the storage device with her edited version. The goal would be to avoid having to create an entirely new version of the stored codeword by taking advantage of the availability of the stored original in the database.

III. Problem Statement and Simplification

We are now ready to give the formal problem statement. Following the formal problem statement, we deduce simplifications to the problem statement and quickly find two achievable points.

A. Formal Problem Statement

Let $\{(X_i, Y_i)\}_{i=1}^{\infty}$ be a sequence of independent drawings of a pair of jointly-distributed random variables (X, Y), $X \in \mathcal{W}$, $Y \in \mathcal{W}$, where $\mathcal{W}$ is a finite set and $p_{X,Y}(x, y) = \Pr[X = x, Y = y]$. The marginal distributions are

$$p_X(x) = \sum_{y \in \mathcal{W}} p(x, y)$$

and

$$p_Y(y) = \sum_{x \in \mathcal{W}} p(x, y),$$

and the conditional distribution

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

describes a *modification channel*. When the random variable is clear from context, we write $p_X(x)$ as $p(x)$ and so on.

Denote the storage medium alphabet by $\mathcal{V}$, which is also a finite set. It is natural to measure all rates in numbers of symbols from $\mathcal{V}$. This is analogous to using base-$|\mathcal{V}|$ logarithms in place of base-2 logarithms, and all logarithms should be interpreted as such.

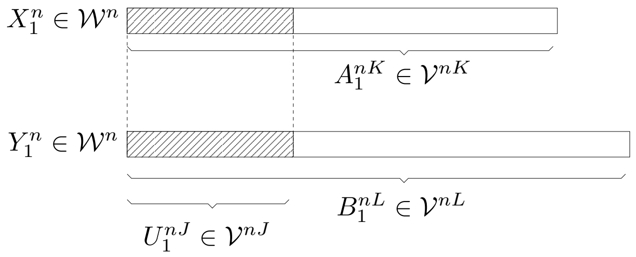
$X_1^n \in \mathcal{W}^n$
$A_1^{nK} \in \mathcal{V}^{nK}$
$Y_1^n \in \mathcal{W}^n$
$B_1^{nL} \in \mathcal{V}^{nL}$
$U_1^{nJ} \in \mathcal{V}^{nJ}$

Fig. 2. In malleable coding with fixed segment reuse, the compressed representations of X_1^n and Y_1^n have the first nJ storage symbols in common.

Our interest is in coding of X_1^n followed by coding of Y_1^n where the first nJ letters of the codes are (asymptotically almost surely) in common. We show this in Fig. 2, where $A_1^{nK} \in \mathcal{V}^{nK}$ is the representation of X_1^n, $B_1^{nL} \in \mathcal{V}^{nL}$ is the representation of Y_1^n, and $U_1^{nJ} \in \mathcal{V}^{nJ}$ is the common initial symbols. We thus define the encoding and decoding mappings as follows.

An encoder for X with parameters (n, J, K) is the concatenation of two mappings:

$$f_E^{(X)} = f_E^{(U)} \times f_E'^{(X)},$$

where

$$f_E^{(U)} : \mathcal{W}^n \to \mathcal{V}^{nJ}$$

and

$$f_E'^{(X)} : \mathcal{W}^n \to \mathcal{V}^{n(K-J)}.$$

An encoder for Y with parameters (n, J, L) is defined as:

$$f_E^{(Y)} = f_E^{(U)} \times f_E'^{(Y)},$$

where we use one of the previous encoders $f_E^{(U)}$ together with

$$f_E'^{(Y)} : \mathcal{W}^n \times \mathcal{V}^{nJ} \to \mathcal{V}^{n(L-J)}.$$

Given these encoders, a common decoder with parameter n is

$$f_D : \mathcal{V}^* \to \mathcal{W}^n.$$

The encoders and decoder define a block code for fixed reuse malleability. Although not strictly required, a common decoder is a convenient way of expressing the requirement of a common deterministic codebook.

A trio $(f_E^{(X)}, f_E^{(Y)}, f_D)$ with parameters (n, J, K, L) is applied as follows. Let

$$A_1^{nK} = f_E^{(X)}(X_1^n)$$

be the source code for X_1^n, where the first part of the code is explicitly notated as

$$U_1^{nJ} = f_E^{(U)}(X_1^n).$$

Then the encoding of Y_1^n is carried out as

$$B_1^{nL} = f_E^{(Y)}(Y_1^n, U_1^{nJ}).$$

We also let

$$(\hat{X}_1^n, \hat{Y}_1^n) = (f_D(A_1^{nK}), f_D(B_1^{nL})).$$

We define the error rate

$$\Delta = \max(\Delta_X, \Delta_Y),$$

where

$$\Delta_X = \Pr[X_1^n \neq \hat{X}_1^n]$$

and

$$\Delta_Y = \Pr[Y_1^n \neq \hat{Y}_1^n],$$

and we define the disagreement rate as

$$\Delta_U = \Pr[A_1^{nJ} \neq B_1^{nJ}].$$

The fact that there is a disagreement rate rather than requiring the first nJ symbols to always be equal introduces the usual slack associated with Shannon reliability. (We will require Δ_U to be arbitrarily small, so the possibility of $\Delta_U \neq 0$ is ignored in Fig. 2.)

We use conventional performance criteria for the code, which are the numbers of storage-medium letters per source letter

$$K = \frac{1}{n} \log_{|\mathcal{V}|} |\mathcal{V}|^{nK}$$

and

$$L = \frac{1}{n} \log_{|\mathcal{V}|} |\mathcal{V}|^{nL},$$

and add as the third performance criterion the normalized length of the portion of the code which does not overlap

$$M = L - J = \frac{1}{n} \log_{|\mathcal{V}|} |\mathcal{V}|^{n(L-J)}.$$

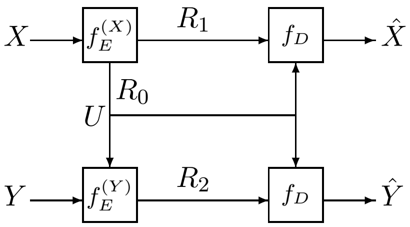

Fig. 3. Block diagram of malleable coding with fixed segment reuse.

We call M the *malleability rate*.

Definition 1: Given a source $p(X, Y)$, a triple (K_0, L_0, M_0) is said to be *achievable* if, for arbitrary $\epsilon > 0$, there exists (for n sufficiently large) a block code for fixed reuse with error rate $\Delta \leq \epsilon$, disagreement rate $\Delta_U \leq \epsilon$, and lengths $K \leq K_0 + \epsilon$, $L \leq L_0 + \epsilon$, and $M \leq M_0 + \epsilon$.

We want to determine the set of achievable rate triples, denoted as $\mathcal{M}$. It follows from the definition that $\mathcal{M}$ is a closed subset of $\mathbb{R}^3$ and has the property that if $(K_0, L_0, M_0) \in \mathcal{M}$, then $(K_0 + \delta_0, L_0 + \delta_1, M_0 + \delta_2) \in \mathcal{M}$ for any $\delta_i \geq 0$, $i = 0, 1, 2$. The rate region $\mathcal{M}$ is thus completely defined by its lower boundary, which is itself closed.

Rather than using (K, L, M), the triple (J, K, L) may be used to characterize the achievable region. Equivalently, we can use (R_0, R_1, R_2) in place of (K, L, M) as shown in Fig. 3. Using this notation is more consistent with established work in multiterminal information theory. The relation is:

1) $J = R_0$,
2) $K = R_0 + R_1$, and
3) $L = R_0 + R_2$.

B. Problem Simplification

A priori, it seems there are two approaches to trading off storage rate for malleability rate in the fixed reuse problem: expending K greater than $H(X)$ might allow a better side information U to be formed; and expending L greater than $H(Y)$ might allow greater flexibility in the design of U. It turns out that expanding the representation of X_1^n provides no advantage, i.e., any extra bits used to encode X will not help in the representation for Y. This is due to the Markov relation $U \leftrightarrow X \leftrightarrow Y$ that holds due to the ordering of the encoding procedure.

For the remainder of this paper, we focus on expending L greater than $H(Y)$ and analyze the achievable rate–malleability. We focus on how L depends on the size of the portion to be reused J, thus establishing the malleability M. When proceeding in this regime, two constraints are imposed:

1) $H(U) = J$, and
2) $H(U, X) = H(X)$.

The second constraint states that U is a subrandom variable of X, which is implicit in the formal problem statement in Section III and the block diagram, Fig. 3.

Rather than characterizing the entire region of achievable triplets $\mathcal{M}$, we consider fixing J and finding the best L (thus fixing $M = L - J$). We want to characterize the achievable rates L as a function of J. The smallest such L is denoted $L^*(J)$.

C. Two Achievable Points

It is easy to note the values of the corner points corresponding to $J = 0$ and $J = H(X)$. For $J = 0$, the lossless source coding theorem yields $L^*(0) = H(Y)$. For $J = H(X)$, since the lossless compression of X_1^n has to be preserved, we will need $L^*(H(X)) = H(X, Y)$. This follows from noting that since the first $H(X)$ symbols have to be fixed, we need to be able to losslessly represent the conditionally typical set, which requires $H(Y|X)$ additional symbols, for a total of $H(X) + H(Y|X) = H(X, Y)$. Since $H(Y|X) \leq H(Y)$, this is better than discarding the old codeword of X_1^n and creating an entirely new codeword for Y_1^n; unless X and Y are independent, this is strictly better.

IV. Main Results

We cast the fixed reuse malleable coding problem as a single letter information theoretic optimization, providing matching achievability and converse statements. Unfortunately, this is not computable in general. Later we will give a computable partial characterization for cases where there exists a sufficient statistic for the estimation of the new version of the source from the reused part of the compressed old version. The basic concepts are also applicable to a lossy formulation with Gaussian sources.

The achievability proof for the boundary of the fixed reuse rate region uses definitions and properties of strongly typical sets (Lemmas 1–4), given in Appendix A.

A. The Fixed Reuse Malleability Region

We consider the trade-off between L and J. From the previous section, it is clear that for a given malleability, the compression efficiency of Y_1^n is determined by the quality of the binning assignment for the typical strings of X_1^n. We capture this assignment by a (probabilistic) function $p(U|X)$. Then, we can formulate the following information theoretic optimization problem:

$$\begin{aligned} L^*(J) - J = \min_{p(U|X)} & H(Y|U) \qquad (1) \\ \text{s.t.} \qquad & H(U) + H(X|U) = H(X), \\ & H(U) = J. \end{aligned}$$

Theorem 1: The optimization problem (1) provides a boundary to the rate region $\mathcal{R} = (R_0, R_1, R_2)$ when $K = R_0 + R_1 = H(X)$.

Proof: **Achievability:** The constraints require $H(U) = J$ and that there is a Markov condition $U \leftrightarrow X \leftrightarrow Y$. Codebooks for X_1^n and Y_1^n are randomly generated according to $p(x)$ and $p(y)$. These codebooks are of size $|\mathcal{V}|^{nK} = |\mathcal{V}|^{nH(X)}$ and $|\mathcal{V}|^{nL}$ respectively. Each codebook is partitioned into $|\mathcal{V}|^{nH(U)}$ bins with a corresponding bin label U_1^{nJ}. Since U_1^{nJ} is a function of X_1^n, it may be written as $U_1^{nJ}(X_1^n)$. Clearly, we can choose $H(U) = J$ and use J symbols to assign the bin label U_1^{nJ}. For the X_1^n codebook, $H(X) - J$ symbols are used to assign labels to members of each bin; the intra-bin label is denoted I_X. Similarly for the Y_1^n codebook, $L - J$ symbols are used to assign labels to members of each bin; the intra-bin label is denoted I_Y.

The encoder for X_1^n, $f_E^{(X)} = f_E^{(U)} \times f_E'^{(X)}$, operates by generating a label $A_1^{nK} = [U_1^{nJ}, I_X]$ according to which x_1^n is realized. The encoder for Y_1^n, $f_E^{(Y)} = f_E^{(U)} \times f_E'^{(Y)}$ generates the same bin label U_1^{nJ} and also generates the intra-bin label I_Y, based on which y_1^n is realized; the resulting encoding is $B_1^{nL} = [U_1^{nJ}, I_Y]$. Since both encoders use the identical bin label u_1^{nJ}, it is clear that the disagreement rate Δ_U can be made arbitrarily small.

The common decoder f_D operates according to strong typicality in the usual way.

By the direct part of Shannon's source coding theorem (see Lemma 1) and the splitting possible due to the entropy chain rule [17], it follows that $\Delta_X = \Pr[X_1^n \neq f_D(A_1^{nK})]$ is arbitrarily small with increasing block length.

Now consider recovering Y_1^n from the codeword $B_1^{nL} = [U_1^{nJ}, I_Y]$, which uses the same prefix but different suffix. The encoder had found the index I_Y such that $(U_1^{nJ}(X_1^n), Y_1^n) \in T^n_{[U_1^{nJ}, Y]\delta}$. The probability of successful encoding is determined by two error events. The first is that (U_1^{nJ}, Y_1^n) does not belong to the typical set; the second is that U_1^{nJ} is jointly typical with X_1^n but not with Y_1^n. The first event has arbitrarily small probability of error by the joint AEP, Lemma 2. The second event has arbitrarily small probability of error by applying Lemma 4 to the $U \leftrightarrow X \leftrightarrow Y$ Markov chain.

Decoding error happens when there is another typical $\tilde{Y}_1^n \neq Y_1^n$ that is jointly typical with U_1^{nJ}. The probability goes to zero almost surely when $L - J > H(Y|U)$ by an AEP argument [18, (14.278)].

Thus Δ_Y and also Δ may be made arbitrarily small, as required for achievability.

Converse: The converse for the encoding and decoding of X_1^n via $[U_1^{nJ}, I_X]$ as a tree-based label follows directly from the converse to Shannon's source coding theorem.

We focus on the encoding of Y_1^n onto $[I_Y]$ and the decoding of $\hat{Y}_1^n$ from $[U_1^{nJ}, I_Y]$. By the encoding strategy, U is a function of X_1^n. We then have a chain of inequalities:

$$\begin{aligned}
n(L-J) = nR_2 &\overset{(a)}{\geq} H(I_Y) \\
&\overset{(b)}{\geq} H(I_Y|U_1^{nJ}) \\
&= I(Y_1^n; I_Y|U_1^{nJ}) + H(I_Y|Y_1^n, U_1^{nJ}) \\
&\overset{(c)}{=} I(Y_1^n; I_Y|U_1^{nJ}) \\
&= H(Y_1^n|U_1^{nJ}) - H(Y_1^n|I_Y, U_1^{nJ}) \\
&\overset{(d)}{\geq} H(Y_1^n|U_1^{nJ}) - n\epsilon \\
&\overset{(e)}{=} nH(Y|U) - n\epsilon.
\end{aligned}$$

Step (a) follows from dimensionality considerations; step (b) from noting that conditioning can only decrease entropy; step (c) from the fact that Y_1^n and U_1^{nJ} determine I_Y; step (d) by applying Fano's inequality; and step (e) from the chain rule of entropy and independence in time. Thus we have obtained the desired inequality. ■

B. Further Characterizations

As in the source coding with side information problem [19]–[21] and several other problems in multiterminal information theory, Theorem 1 left us to optimize an auxiliary random variable U that describes the method of partitioning. Here we will provide simple bounds on $L^*(J)$ and then further characterization in terms of a sufficient statistic of X for Y.

Theorem 1 demonstrated that we require

$$L(J) \geq H(Y|U) + J.$$

The easily achieved corner points discussed previously and a few simple bounds are shown in Fig. 4. The bounds, marked by dotted lines, are as follows:

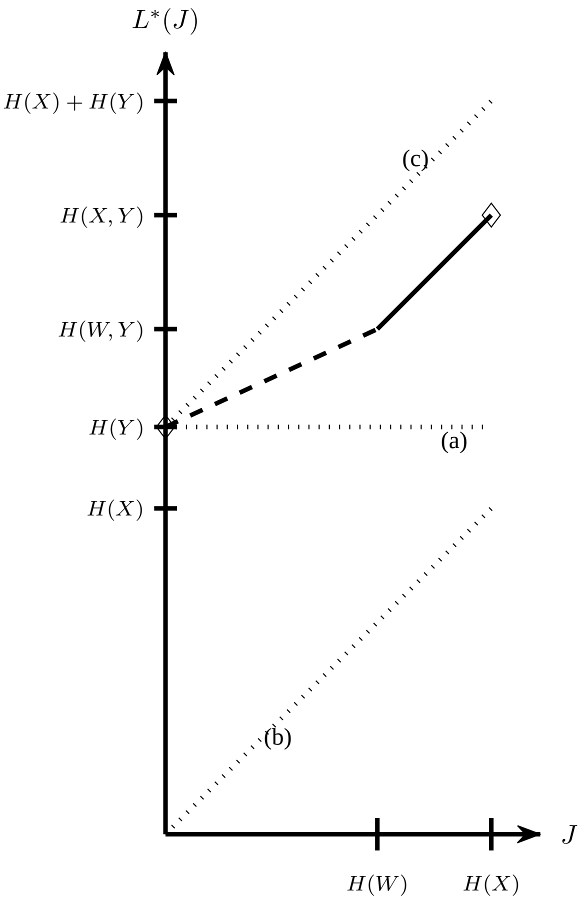

Fig. 4. Characterizations of the fixed reuse malleability region boundary $L^*(J)$. Each $\Diamond$ is a point determined in Section III-C, and the dotted lines are simple bounds from Section IV-B. With W defined as a minimal sufficient statistic of X for Y, the solid line shows the unit-slope boundary determined by Theorem 2. The dashed line represents a portion of boundary that is unknown (but known to be convex by Theorem 3).

(a) The lossless source coding theorem applied to Y alone gives $L^*(J) \geq H(Y)$.

(b) Another trivial lower bound from the construction is $L^*(J) \geq J$.

(c) Since one could encode Y_1^n without trying to take advantage of the J symbols already available, $L^*(J) \leq J + H(Y)$.

In evaluating the properties of $L^*(J)$ further, let W be a minimal sufficient statistic of X for Y. Intuitively, if J is large enough that one can encode W in the shared segment U_1^{nJ}, it is efficient to do so. Thus we obtain regimes based on whether J is larger than $H(W)$.

For the regime of $J \geq H(W)$, the boundary of the region is linear by the following theorem:

Theorem 2: Consider the problem of (1). Let W be a minimal sufficient statistic of X for Y. For $J > H(W)$, the solution is given by:

$$L^*(J) - J = H(Y|W). \tag{2}$$

Proof: By definition, a sufficient statistic contains all information in X about Y. Therefore any rate beyond the rate required to transmit the sufficient statistic is not useful. Beyond $H(W)$, the solution is linear. ■

A rearrangement of (2) is

$$L^*(J) = H(Y, W) + [J - H(W)].$$

This is used to draw the portion of the boundary determined by Theorem 2 with a solid line in Fig. 4.

For the regime of $J < H(W)$, we have not determined the boundary but we can show that $L^*(J)$ is convex.

Theorem 3: Consider the problem of (1). Let W be a minimal sufficient statistic of X for Y. For $J < H(W)$, the solution $L^*(J)$ is convex.

Proof: Follows from the convexity of conditional entropy, by mixing possible distributions U. ■

The convexity from Theorem 3 and the unit slope of $L^*(J)$ for $J > H(W)$ from Theorem 2 yield the following theorem by contradiction. An alternative proof is given in Appendix B.

Theorem 4: The slope of $L^*(J)$ is bounded below and above:

$$0 \leq \frac{d}{dJ} L^*(J) \leq 1.$$

The following can be seen as extremal cases for the theorem: when X and Y are independent, $L^*(J) = J + H(Y)$ and so $\frac{d}{dJ} L^*(J) = 1$. When $X = Y$, $L^*(J) = H(Y)$ for any J, and so $\frac{d}{dJ} L^*(J) = 0$.

Without regard to the constraint on J, it is known that the sufficient statistic for Y upon the observation $X = x$ is $p(Y|X = x)$. Therefore for the regime where $J > H(p(Y|X = x))$, this is the best knowledge of Y we can endow to the decoder for decoding Y.

The challenge lies when $J \leq H(p(Y|X = x))$: this is an estimation problem with limited communication budget. In a lossy setting, for the special case of jointly Gaussian X and Y this problem may be entirely solved by casting it as a linear least-squares estimation problem.

In fact, (1) can be stated as follows:

$$\begin{aligned} \max_{f(X)} \quad & H(Y|H(f(X))) \\ \text{s.t.} \quad & H(f(X)) = m \end{aligned} \tag{3}$$

It is clear that (1) and (3) are equivalent. In this problem the design of the label is cast as the problem of designing a sufficient statistic of Y given X, consistent with our previous discussion. The fact that in this statement U equals $f(X)$ ensures that U is a subrandom variable of X.

V. Connections

An alternate method of further analyzing the rate–malleability region for fixed segment reuse is to make connections with solved problems in the literature. Here we connect our problem and the lossless source coding with coded side information problem [19]–[21]. Source coding with coded side information problems provide achievable rate regions for fixed reuse malleability. We also discuss relations to a common information problem [22]. If $K = H(X)$ and $L = H(Y)$ are required, then the length of the common portion of the source code is less than or equal to $C(X;Y)$, the Gács–Körner common information.

A. Relation to the Coded Side Information Problem

In this section, we show that rate regions for the coded side information problem (also called the helper problem) are achievable rate regions for the malleability problem. Results are expressed in terms of the rate triple $\mathcal{R}$ rather than the rate–malleability triple $\mathcal{M}$.

Definition 2: Let

$$\mathcal{R}_{\text{help}_1} = \left\{ \begin{array}{rrcl} (R_0, R_1, R_2): & R_0 & \geq & H(U) \\ & R_0 + R_1 & \geq & H(X) \\ & R_2 & \geq & H(Y|U) \end{array} \right\},$$

where U is any auxiliary random variable.

Theorem 5: The rate region for the coded side information problem $\mathcal{R}_{\text{help}_1}$ is an achievable rate region for the fixed reuse malleability problem, i.e. $\mathcal{R}_{\text{help}_1} \subseteq \mathcal{R}$.

Proof: The result follows simply by noting that the malleability problem has a more extensive information pattern than the coded side information problem (see Fig. 5) and by the achievability result for the coded side information problem [20, Theorem 2.1]. Wyner's rate region in the case where the side information need not be compressed satisfies $R_0 \geq H(U)$, $R_1 \geq H(X|U)$, and $R_2 \geq H(Y|U)$, which implies $\mathcal{R}_{\text{help}_1}$. ■

For the malleable coding problem, the auxiliary random variable U may be generated from X and will be given to the encoder for Y. Lossless source coding is always successively refinable [17], but it is unclear how to split off some of the information from X into U.

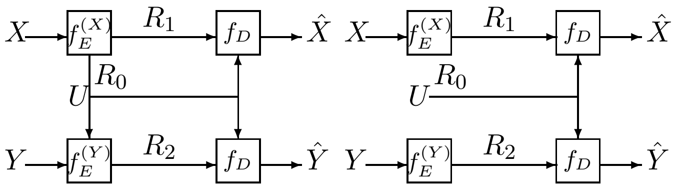

Fig. 5. The fixed reuse malleable coding problem (left, Fig. 3) has a more extensive information pattern than the coded side information problem (right). For fixed reuse, the side information may be designed from X and this side information is available at the encoder for Y.

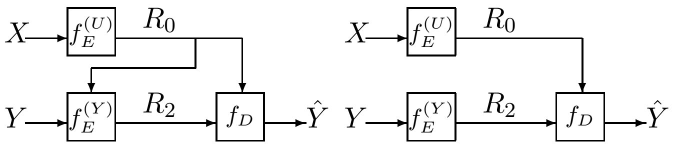

Fig. 6. The fixed reuse malleable coding problem (left) has a more extensive information pattern than the coded side information problem (right). For fixed reuse, the coded side information is available at the encoder for Y.

In the result just given, the side information was not compressed and so the rate region was actually a Slepian–Wolf region [23] rather than a true coded side information rate region, even though the coded side information theorem was invoked in the proof. An alternative comparison leads to the side information actually being compressed. In particular, consider the coded side information problem where X is side information to be compressed, and Y is the source to be compressed. There is a decoder that takes these two things and tries to reproduce Y. This describes only the lower branch of the fixed reuse system. The upper branch would produce a code to allow lossless reconstruction of X at total rate $R_0 + R_1 \approx H(X)$.

We focus on the lower branch, studying the trade-off between R_0 and R_2. This is equivalent to looking at $L^*(J)$, as in previous sections. In order to cast an equivalence to the coded side information problem, assume that the side information code is not available to the Y encoder. Since the malleable coding problem has a more extensive information pattern, this implies that the derived rate region will be an achievable region. The lower branch as described, is now exactly the coded side information problem [19], [20].

Definition 3: Let

$$\mathcal{R}_{\text{help}_2} = \left\{ \begin{array}{lccc} (R_0, R_2): & R_0 & \geq & H(Y|U) \\ & R_2 & \geq & I(X;U) \end{array} \right\},$$

where U is any auxiliary random variable that satisfies the Markov condition $U \leftrightarrow X \leftrightarrow Y$.

Theorem 6: The rate region for the coded side information problem $\mathcal{R}_{\text{help}_2}$ is an achievable rate region for the lower branch of the malleable coding with fixed reuse problem, i.e. $\mathcal{R}_{\text{help}_2} \subseteq \text{proj}_{(R_0,R_2)}\mathcal{R}$.

Proof: The result follows simply by noting that the malleable coding problem has a more extensive information pattern than the coded side information problem (see Fig. 6) and by the achievability result for the coded side information problem [19, Theorem 2]. ■

Since we are interested in the lower boundary of the rate region, finding $\mathcal{R}_{\text{help}_2}$ may be reduced to optimizing the auxiliary random variable U for the coded side information problem, which is also the reused segment of the source code for the malleable coding problem. This is usually difficult, but see [21], [24]. The optimization problem for R_0 as a function of R_2 is

$$F(R_2) = \min_{p(U|X)} H(Y|U) \tag{4}$$
$$\text{s.t. } I(U;X) \leq R_2.$$

Interestingly, a problem in machine learning called the information bottleneck problem formulates a similar optimization function and provides an alternative operational interpretation of $\mathcal{R}_{\text{help}_2}$ [25], [26].[3] The optimization problem is

$$B(R_2) = \max_{p(U|X)} I(Y;U) \tag{5}$$
$$\text{s.t. } I(U;X) \leq R_2,$$

which clearly satisfies $F(R_2) = H(Y) - B(R_2)$, since $H(Y)$ is not open to optimization [26].

One can notice that the optimization problem (1) is closely related to the optimization problems that arise for the coded side information problem and the information bottleneck problem. In particular, it can be noted that the constraint is a subset of the constraint for the coded side information problems. Since $I(U;X) = H(U) - H(U|X)$, it follows that $\{p(U|X) : I(U;X) \leq R_0\} \supseteq \{p(U|X) : H(U) \leq R_0 \text{ and } H(U|X) = 0\}$.

B. Relation to Gács–Körner Common Information

We have found that rate regions for lossless coding with coded side information are achievable for malleable coding, however computing these regions involves optimizing auxiliary random variables. It turns out that for particular ranges of rates, the rate region is actually known in closed form [21]; the

[3] New developments in computing the rate region for the coded side information problem [21], [24] also have implications for computing the information bottleneck function [25], [26], though these do not appear to have been exploited.

range is partially delimited by the common information functional of Gács and Körner [22], [27, pp. 402–404]. The Gács–Körner common information also yields a characterization of malleable coding with fixed segment reuse.

Definition 4: For random variables X and Y, let $U = f(X) = g(Y)$ where f is a function of X and g is a function of Y such that $f(X) = g(Y)$ almost surely and the number of values taken by f (or g) with positive probability is the largest possible. Then the Gács–Körner common information, denoted $C(X;Y)$, is $H(U)$.

Definition 5: The joint distribution $p(x, y)$ is *indecomposable* if there are no functions f and g each with respect to the domain $\mathcal{W}$ so that

- $\Pr[f(X) = g(Y)] = 1$, and
- $f(X)$ takes at least two values with non-zero probability.

It can be shown that $C(X;Y) = 0$ if X and Y have an indecomposable joint distribution. Further properties of indecomposable joint distributions are given in [27, p. 350] and [21]. In particular, an auxiliary random variable U that satisfies the Markov relation $U \leftrightarrow X \leftrightarrow Y$ is used for characterization.

Gács and Körner show that the maximal length of the common beginning portion of entropy-achieving source codes for X and for Y, the operational definition of common information, coincides with the informational definition of common information. The basic result, [22, Theorem 1], is that it is not possible in general to code two sources so that the resulting codes have some common fixed length of order n. This is because in general, $p(x, y)$ is indecomposable and so the common information is zero. Such a negative result also carries over to the fixed reuse problem.

Consider the block diagram for the coding problem that involves the common information in its solution [27, Fig. P.28 on p. 403], Fig. 7. If it is required that $R_1 = H(X) - R_0$ and that $R_2 = H(Y) - R_0$, then the largest possible R_0 is $C(X;Y)$. Since entropy is being achieved, it follows that $R_2 = H(Y|U)$ through Slepian–Wolf or conditional entropy means at $f_E^{\prime(Y)}$ [17]. Since the distributed system does as well as a centralized system, even if U is given to $f_E^{\prime(Y)}$, this will not improve things. In particular, the system shown in Fig. 8 will have the same relationship to the common information. Showing this rigorously involves modifying the converse of the common information proof and seeing that the arguments follow through. Now one can observe that this block diagram is an enhanced version of the fixed reuse malleable coding block diagram, redrawn as Fig. 9.

Theorem 7: The Gács–Körner common information rate triple provides a partial converse to the rate–malleability triple.

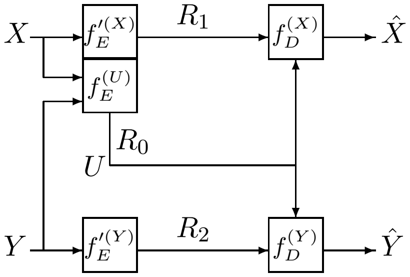

Fig. 7. Block diagram for the Gács–Körner common information problem.

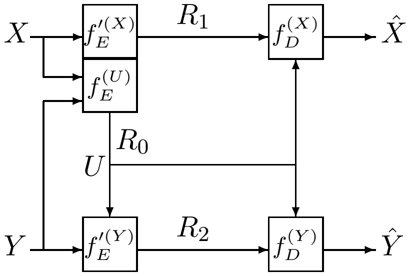

Fig. 8. Block diagram for the Gács–Körner common information problem when giving U to $f_E'^{(Y)}$. This additional information does not help in coding.

Proof: The result follows from the fact that the common information problem has a more extensive information pattern than the fixed reuse malleable coding problem (see Fig. 9) and the converse for the enhanced common information problem [22]. ■

This theorem gives an outer bound to go with the achievable region defined in Definition 2. Thus for the malleable coding problem, if we want $K = H(X)$ and $L = H(Y)$, then M must be bad: $M \geq H(Y) - C(X;Y)$, where $C(X;Y)$ is often zero. Since there is almost no overlap possible when requiring $L = H(Y)$, allowing larger L in Section III-B was a good approach.

VI. Discussions and Closing Remarks

Phrased in the language of waste avoidance and resource recovery: classical Shannon theory shows how to optimally *reduce*; we have here studied *reuse* and in [1] studied *recycling* and have found these goals to be fundamentally in tension.

We have formulated an information-theoretic problem motivated by the transmission of data to up-

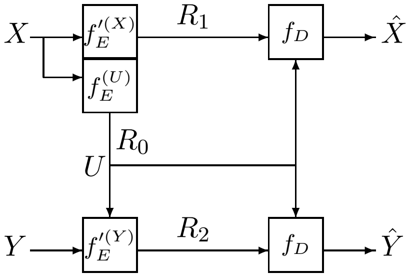

Fig. 9. Block diagram for malleable coding with fixed segment reuse. This has a reduced information pattern as compared to the Gács–Körner common information problem when giving U to $f_E'^{(Y)}$.

date the compressed version of a document after it has been edited. Any technique akin to optimally compressing the difference between the documents would require the receiver to uncompress, apply the changes, and recompress. We instead require *reuse* of a fixed portion of the compressed version of the original document; this segment cut from the compressed version of the original document is pasted into the compressed version of the new document. This requirement creates a trade-off between the amount of reuse and the efficiency in compressing the new document. Theorem 1 provides a complete characterization as a single-letter information-theoretic optimization.

We established relationships to several previously-studied multiterminal information theory problems. Perhaps the most interesting is with the Gács–Körner common information problem. Through that relationship one can see that if the original and modified sources have an indecomposable joint distribution and are required to be coded close to their entropies, then the reused fraction must asymptotically be negligible. We also showed through a Markovianity argument that there is no benefit from coding the original source above its entropy. Our focus was therefore on cases where the modified source is encoded with excess rate.

A. On the Effectiveness of Binning

We informally describe the ineffectiveness of independent, uniform binning. Place the codewords of $T^n_{[X]\delta}$ that have the same first nJ symbols into the same bin. There are $|\mathcal{V}|^{nJ}$ bins, each of which has $|\mathcal{V}|^{n(H(X)-J)}$ elements. Let the bins be labeled by $U_1^{nJ} = 1, \ldots, |\mathcal{V}|^{nJ}$. For each of the bins u_1^{nJ} containing some sequences of x_1^n, create a corresponding bin to contain the conditionally typical sequences y_1^n, given that $x_1^n \in u_1^{nJ}$. This gives the smallest sized bins for y_1^n given that the first nJ symbols of the representation of x_1^n are the same as the first nJ symbols used to represent y_1^n. It is clear that the representation of y_1^n is not unique, as the same y_1^n may be represented in more than one bin.

For each $x_1^n \in T^n_{[X]\delta}$ there are about $|\mathcal{V}|^{nH(Y|X)}$ conditionally typical members of $T^n_{[Y|X]\delta}(x_1^n)$ by Lemma 3. Through the union bound (Boole's inequality) we obtain:

$$\begin{aligned} |T^n_{[Y|X]\delta}(x_1^n \in u_1^{nJ})| &\le |u_1^{nJ}||T^n_{[Y|X]\delta}(x_1^n)| \\ &= |\mathcal{V}|^{n(H(X)-J)}|\mathcal{V}|^{nH(Y|X)}; \end{aligned}$$

note that there are $|\mathcal{V}|^{nJ}$ such bins u_1^{nJ}. Although this may suggest that the compression of Y_1^n may require up to $nH(X,Y) = n(H(X)+H(Y|X))$ regardless of the value of J, this is not the case.

The union bound is tight if and only if it consists of independent events, but it is difficult to examine the tightness or to find a tighter bound. One might believe that the union bound is tight for any $J > 0$, implying a rate requirement of $H(X) + H(Y|X) = H(X,Y)$ symbols for the compression of Y_1^n to have any nontrivial malleability. With the upper bound of Section IV-B, we have shown that this belief is false. Thus the union bound is not tight, and independent, uniform binning [23] fails.

B. Designing Side Information

Even after characterization by a coding theorem, rate regions in multiterminal information theory are notoriously difficult to examine because of optimizations involving auxiliary random variables. For several source coding problems with coded side information, achievable rates are characterized by product-space characterizations with implicit optimizations over infinite-letter mappings. One can think of these optimizations as problems of designing useful side information. For malleable coding problems, the design of side information takes central importance.

For the Slepian–Wolf problem [23], side information formed through random binning is good. For point-to-point problems, (side) information formed through quantization binning is good. For other problems, however, there is no intuition about optimal auxiliary random variables and the nature of good binning. Recent work on the source coding with coded side information problem [19], [20] provides some insight into regimes where side information generated through codes like random-binning works and where it does not [21], however there is no general theory.

One fundamental difference between coding with side information problems and the malleable coding problem is the time ordering of when codes are applied. Here, the first source is compressed and then the second source is compressed with access to a portion of the actual realization of the compressed version of the first source, not just a statistical description.

Appendix A
Strong Typicality

Definition 6: The strongly typical set $T^n_{[X]\delta}$ with respect to $p(x)$ is

$$T^n_{[X]\delta} = \left\{ x_1^n \in \mathcal{W}^n \mid \sum_x \left| \frac{N(x;x_1^n)}{n} - p(x) \right| \leq \delta \right\},$$

where $N(x;x_1^n)$ is the number of occurrences of x in x_1^n and $\delta > 0$.

Definition 7: The strongly jointly typical set $T^n_{[XY]\delta}$ with respect to $p(x,y)$ is

$$T^n_{[XY]\delta} = \left\{ (x_1^n, y_1^n) \in \mathcal{W}^{n \times n} \mid \sum_{x,y} \left| \frac{N(x,y;x_1^n,y_1^n)}{n} - p(x,y) \right| \leq \delta \right\}.$$

Definition 8: For any $x_1^n \in T^n_{[X]\delta}$, define a strongly conditionally typical set

$$T^n_{[Y|X]\delta}(x_1^n) = \left\{ y_1^n \in T^n_{[Y]\delta} \mid (x_1^n, y_1^n) \in T^n_{[XY]\delta} \right\}.$$

Now that we have definitions of typical sets, we put forth some lemmas.

Lemma 1 (Strong AEP): Let η be a small positive number such that $\eta \to 0$ as $\delta \to 0$. Then for sufficiently large n,

$$\left| T^n_{[X]\delta} \right| \leq |\mathcal{V}|^{n(H(X)+\eta)}.$$

Proof: See [28, Theorem 5.2]. ■

Lemma 2 (Strong JAEP): Let λ be a small positive number such that $\lambda \to 0$ as $\delta \to 0$. Then for sufficiently large n,

$$\Pr[(X_1^n, Y_1^n) \in T^n_{[XY]\delta}] > 1 - \delta$$

and

$$(1-\delta)|\mathcal{V}|^{n(H(X,Y)-\lambda)} \leq \left| T^n_{[XY]\delta} \right| \leq |\mathcal{V}|^{n(H(X,Y)+\lambda)}.$$

Proof: See [28, Theorem 5.8]. ■

Lemma 3: If $\left| T^n_{[Y|X]\delta}(x_1^n) \right| \geq 1$, then

$$|\mathcal{V}|^{n(H(Y|X)-\nu)} \leq \left| T^n_{[Y|X]\delta}(x_1^n) \right| \leq |\mathcal{V}|^{n(H(Y|X)+\nu)},$$

where $\nu \to 0$ as $n \to \infty$ and $\delta \to 0$.

Proof: See [28, Theorem 5.9]. ■

Lemma 4 (Berger's Markov Lemma): Let (X,Y,Z) form a Markov chain $X \leftrightarrow Y \leftrightarrow Z$. Then for sufficiently large n,

$$\Pr[(X_1^n, z_1^n) \in T^n_{[XZ]|\mathcal{X}|\delta} | (Y_1^n, z_1^n) \in T^n_{[YZ]\delta}] > 1 - \delta$$

for any $\delta > 0$ and any realization z_1^n.

Proof: See [29, Lemma 4.1]. ■

APPENDIX B
ALTERNATE PROOF OF THEOREM 4

Proof of upper bound: Let $J_1 > J_2$ be any two values of J. Let V_1 and V_2 be the corresponding auxiliary random variables U that solve the optimization problem (1). Then by the successive refinability of lossless coding [17], it follows that V_1 and V_2 will satisfy the Markov chain $V_2 \leftrightarrow V_1 \leftrightarrow X \leftrightarrow Y$.

By the data processing inequality,

$$I(Y; V_2) \leq I(Y; V_1)$$
$$H(V_1|Y) - H(V_2|Y) \leq H(V_1) - H(V_2).$$

By definition,

$$\begin{aligned} L^*(J_1) - L^*(J_2) &= H(Y|V_1) + H(V_1) - H(Y|V_2) - H(V_2) \\ &= H(V_1|Y) - H(V_2|Y). \end{aligned}$$

Therefore,

$$L^*(J_1) - L^*(J_2) \leq H(V_1) - H(V_2) = J_1 - J_2$$

which implies

$$\frac{L^*(J_1) - L^*(J_2)}{J_1 - J_2} \leq 1.$$

Proof of lower bound: We want to show that $H(V_1|Y) - H(V_2|Y) \geq 0$. This property may be verified using Yeung's ITIP [28] after invoking the Markov chain $V_2 \leftrightarrow V_1 \leftrightarrow X \leftrightarrow Y$ and the subrandomness conditions, $H(V_1|X) = H(V_2|X) = 0$.

ACKNOWLEDGMENT

The authors thank Ramesh Johari for discussions on the problem statement; Sanjoy K. Mitter for useful discussions; İ. Emre Telatar for making them aware of [22]; Daniel Marco for providing a preprint of [21] and Vahid Tarokh for discussions on storage area networks.

REFERENCES

[1] L. R. Varshney, J. Kusuma, and V. K. Goyal, "Malleable coding: Compressed palimpsests," arXiv:0806.4722v1 [cs.IT]., June 2008.

[2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July/Oct. 1948.

[3] K. Zeger and A. Gersho, "Pseudo-Gray coding," *IEEE Trans. Commun.*, vol. 38, no. 12, pp. 2147–2158, Dec. 1990.

[4] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[5] T. K. Lala, "Storage area networking," *IEEE Commun. Mag.*, vol. 41, no. 8, pp. 70–71, Aug. 2003.

[6] T. C. Jepsen, "The basics of reliable distributed storage networks," *IEEE IT Prof.*, vol. 6, no. 3, pp. 18–24, May-June 2004.

[7] F. Z. Wang, S. Wu, N. Helian, M. A. Parker, Y. Guo, Y. Deng, and V. R. Khare, "Grid-oriented storage: A single-image, cross-domain, high-bandwidth architecture," *IEEE Trans. Comput.*, vol. 56, no. 4, pp. 474–487, Apr. 2007.

[8] A. G. Dimakis and K. Ramchandran, "Network coding for distributed storage in wireless networks," in *Networked Sensing Information and Control*, V. Saligrama, Ed. New York: Springer, 2008, pp. 115–136.

[9] D. A. Patterson and J. L. Hennessy, *Computer Organization & Design: The Hardware/Software Interface*, 2nd ed. San Francisco: Morgan Kaufmann Publishers, Inc., 1998.

[10] D. R. Bobbarjung, S. Jagannathan, and C. Dubnicki, "Improving duplicate elimination in storage systems," *ACM Trans. Storage*, vol. 2, no. 4, pp. 424–448, Nov. 2006.

[11] C. Policroniades and I. Pratt, "Alternatives for detecting redundancy in storage systems data," in *Proc. 2004 USENIX Annu. Tech. Conf.*, June 2004, pp. 73–86.

[12] R. Burns, L. Stockmeyer, and D. D. E. Long, "In-place reconstruction of version differences," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 4, pp. 973–984, July-Aug. 2003.

[13] T. Suel and N. Memon, "Algorithms for delta compression and remote file synchronization," in *Lossless Compression Handbook*, K. Sayood, Ed. London: Academic Press, 2003, pp. 269–290.

[14] N. Garg, S. Sobti, J. Lai, F. Zheng, K. Li, R. Y. Wang, and A. Krishnamurthy, "Bridging the digital divide: Storage media + postal network = generic high-bandwidth communication," *ACM Trans. Storage*, vol. 1, no. 2, pp. 246–275, May 2005.

[15] R. Ahlswede and Z. Zhang, "Coding for write-efficient memory," *Inf. Comput.*, vol. 83, no. 1, pp. 80–97, Oct. 1989.

[16] P. C. Wong, K.-K. Wong, and H. Foote, "Organic data memory using the DNA approach," *Commun. ACM*, vol. 46, no. 1, pp. 95–98, Jan. 2003.

[17] J. Körner, "A property of conditional entropy," *Stud. Sci. Math. Hung.*, vol. 6, pp. 355–359, 1971.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[19] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.

[20] A. D. Wyner, "On source coding at the decoder with side information," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.

[21] D. Marco and M. Effros, "On lossless coding with coded side information," *IEEE Trans. Inf. Theory*, submitted.

[22] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Control Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[23] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 2, pp. 226–228, Mar. 1975.

[24] W. Gu, R. Koetter, M. Effros, and T. Ho, "On source coding with coded side information for a binary source with binary side information," in *Proc. 2007 IEEE Int. Symp. Inf. Theory*, June 2007, pp. 1456–1460.

[25] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Annu. Allerton Conf. Commun. Control Comput.*, Sept. 1999, pp. 368–377.

[26] R. Gilad-Bachrach, A. Navot, and N. Tishby, "An information theoretic tradeoff between complexity and accuracy," in *Learning Theory and Kernel Machines*, B. Schölkopf and M. K. Warmuth, Eds. Berlin: Springer, 2003, pp. 595–609.

[27] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 3rd ed. Budapest: Akadémiai Kiadó, 1997.

[28] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer Academic/Plenum Publishers, 2002.

[29] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1977, pp. 172–231.