Malleable Coding with Fixed Segment Reuse

Julius Kusuma, Lav R. Varshney, and Vivek K Goyal
Research Laboratory of Electronics, Massachusetts Institute of Technology

Abstract—In cloud computing, storage area networks, and remote backup storage, stored data is modified with updates from new versions. It is desirable for the data to not only be compressed but also be easily modified during updates, since representing information and modifying the representation are both expensive. A malleable coding scheme considers both compression efficiency and ease of alteration, promoting codeword reuse. We examine the trade-off between compression efficiency and malleability cost—the difficulty of synchronizing compressed versions—measured as the length of a reused prefix portion. Through a coding theorem, the region of achievable rates and malleability cost—the difficulty of synchronizing compressed versions—is characterized.

I. INTRODUCTION

Many information systems store frequently-updated documents. There is often a separation between hosts used to process information and devices used to store information. Storage costs include not only the average length of the coded document, but also costs in updating, both in communicating and in applying changes. Current technological trends in transmission and storage technologies show that transmission capacity has grown more slowly than disk storage capacity [1]. Hence “new” representation symbols may be more expensive than “old” representation symbols, suggesting that reusing parts of codewords may be more economical than simply reducing their lengths, as in conventional data compression.

Moving toward a tractable abstraction, suppose that after compressing a random source sequence \( X_1 \), it is modified to become a new source sequence \( Y_1 \) according to a memoryless update process \( p_y|x \). A malleable coding scheme preserves a portion of the codeword of \( X_1 \) and modifies the remainder into a new codeword from which \( Y_1 \) may be decoded reliably.

Our notion of preservation requires reusing a fixed part of the codeword for \( X_1 \) in generating a codeword for \( Y_1 \). We call this fixed segment reuse. Without loss of generality, the fixed portion can be taken to be at the beginning, so the new codeword is a fixed prefix followed by a new suffix. A contrasting scenario is for a cost to be incurred when a symbol is changed in value, regardless of its location; we studied this random access problem in [2].

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. 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. Phrased in the language of waste avoidance and resource recovery: classical Shannon theory shows how to optimally reduce; we here study reuse and find these goals to be fundamentally in tension.

II. PROBLEM STATEMENT

Let \( \{(X_i, Y_i)\}_{i=1}^\infty \) be a sequence of independent drawings of a pair of 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 joint distribution determines \( p_X(x) \), \( p_Y(y) \), and \( p_{Y|X}(y|x) \). Denote the finite storage medium alphabet by \( \mathcal{V} \). We measure all rates in numbers of symbols from \( \mathcal{V} \). This is analogous to using base-\(|\mathcal{V}| \) logarithms.

Our interest is in coding of \( X_1^n \) followed by coding of \( Y_1^n \) where the first \( n \) letters of the codewords are exactly the same. As depicted in Fig. 1, \( A_1^{nK} \in \mathcal{V}^{nK} \) is the representation of \( X_1^n \), \( B_1^{nL} \in \mathcal{V}^{nL} \) is the representation of \( Y_1^n \), and \( C_1^{nJ} \in \mathcal{V}^{nJ} \) is the common part. Encoder and decoder mappings are thus defined as follows.

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

\[
\phi_E^{(X)} = \phi_E^{(U)} \times \phi_E^{(L)}
\]

where

\[
\phi_E^{(U)} : \mathcal{W}^n \rightarrow \mathcal{V}^{nJ} \text{ and } \phi_E^{(L)} : \mathcal{W}^n \rightarrow \mathcal{V}^{(K-J)}.
\]

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

\[
\phi_E^{(Y)} = \phi_E^{(U)} \times \phi_E^{(L)}
\]

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

\[
\phi_E^{(L)} : \mathcal{W}^n \times \mathcal{V}^{nJ} \rightarrow \mathcal{V}^{(L-J)}.
\]
Notice that \( f_E^{(Y)} \) is defined so as to have access to the previously stored prefix. Given these encoders, a common decoder with parameter \( n \) is

\[
f_D : \mathcal{Y}^n \to \mathcal{W}^n = \begin{cases} \mathcal{Y}^n K \to \mathcal{W}^n, & \text{first version} \\ \mathcal{Y}^n L \to \mathcal{W}^n, & \text{second version.} \end{cases}
\]

The encoders and decoder define a block code.

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

\[
A_i^{nK} = f_E^{(X)}(X_i^n) = [f_E^{(U)}(X_i^n), f_E^{(X)}(X_i^n)], \quad A_i^{nK} \in \mathcal{V}^{nK},
\]

be the source code for \(X_i^n\), where the first part of the code—which will be reused—is explicitly notated as

\[
C_i^{nJ} \in \mathcal{V}^{nJ} = f_E^{(U)}(X_i^n).
\]

The partial codeword \(C_i^{nJ}\) asymptotically almost surely (a.a.s.) losslessly represents a random variable we call \(U_i^n\). Then the encoding of \(Y_i^n\) is carried out as

\[
B_1^{nL} = f_E^{(Y)}(C_1^{nJ}, Y_1^n) = [C_1^{nJ}, f_E^{(Y)}(C_1^{nJ}, Y_1^n)], \quad B_1^{nL} \in \mathcal{V}^{nL}.
\]

We also let

\[
(\hat{X}_i^n, \hat{Y}_i^n) = (f_D(A_i^{nK}), f_D(B_i^{nL})).
\]

We define the error rate \(\Delta = \max(\Delta_X, \Delta_Y)\), where

\[
\Delta_X = \text{Pr}[X_i^n \neq \hat{X}_i^n] \quad \text{and} \quad \Delta_Y = \text{Pr}[Y_i^n \neq \hat{Y}_i^n].
\]

Note that by construction we insist that the first \(nJ\) symbols are identical: \(A_i^{nJ} = B_i^{nJ} = C_i^{nJ}\).

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

\[
K = \frac{1}{n} \log |\mathcal{V}| |\mathcal{V}|^{nK} \quad \text{and} \quad L = \frac{1}{n} \log |\mathcal{Y}| |\mathcal{V}|^{nL},
\]

and add, as a third performance criterion, the normalized length of the portion of the code that does not overlap

\[
M = L - J = \frac{1}{n} \log |\mathcal{V}| |\mathcal{V}|^{(nL-J)}.
\]

We call \(M\) the malleability rate.

**Definition 1**: Given a source \(p(X,Y)\), a triple \((K_0, \delta_0, M_0)\) is said to be achievable if, for arbitrary \(\epsilon > 0\), there exists (for \(n\) sufficiently large) a block code with error rate \(\Delta \leq \epsilon\), 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, \(\mathcal{M}\).

It follows from definition that \(\mathcal{M}\) is a closed subset of \(\mathbb{R}^3\) and that if \((K_0, \delta_0, M_0) \in \mathcal{M}\), then \((K_0 + \delta_i, L_0 + \delta_i, M_0 + \delta_2) \in \mathcal{M}\) for any \(\delta_i \geq 0, i = 0, 1, 2\). The region \(\mathcal{M}\) is completely defined by its lower boundary, which is itself closed.

The triple \((J, K, L)\) may be used in place of \((K, L, M)\) when convenient. Moreover, \(R = (R_0, R_1, R_2)\) may be used, in place of \((J, K, L)\) as shown in Fig. 2. Using this notation is more consistent with established work in multiterminal information theory, but less natural for the engineering problem. The relation is \((J = R_0, K = R_0 + R_1, L = R_0 + R_2)\).

### III. Time Ordering, Markov Relations, and Two Achievable Points

We begin by considering the effect of time ordering on our problem and give two achievable points.

**A. Simplification**

There is a time ordering in malleable coding. The sources \(X_i^n\) and \(Y_i^n\) come from a joint distribution, however the partial codeword \(C_i^{nJ}\) that represents \(U_i^n\) is generated by encoder \(f_E^{(U)}\) based on \(X_i^n\) prior to the encoding of \(Y_i^n\) by \(f_E^{(Y)}\). Consequently the time ordering of the encoding procedure implies the Markov relation \(U \leftrightarrow X \leftrightarrow Y\).

**Proposition 1**: Taking \(K > H(X)\) provides no advantage in malleable coding with fixed reuse.

**Proof**: Consider the representation of \(X_i^n\), \(A_i^{nK} = [f_E^{(U)}(X_i^n), f_E^{(X)}(X_i^n)]\) and for convenience, let \(A_i^{n(K-J)} = f_E^{(X)}(X_i^n)\) denote the portion that is not reused, so that \(A_i^{nK} = [C_i^{nJ}, A_i^{n(K-J)}]\). Suppose we expand the representation by taking \(K > H(X)\). The extra symbols are either spent in \(C\), in \(A'\), or in both.

From the time-ordering derived Markov structure, \(U \leftrightarrow X \leftrightarrow Y\), \(X\) is a sufficient statistic of \(U\) for \(Y\).

Spending extra symbols in \(A'\) is wasteful since \(A'\) is not used to encode \(Y_i^n\). Extra symbols in \(C_i^{nJ}\) means that \(J > H(f_E^{(Y)}(X_i^n))\); spending extra symbols in \(C_i^{nJ}\) is wasteful since \(X\) is a sufficient statistic of \(U\) for \(Y\).

We focus on expanding \(L\) beyond \(H(Y)\) and analyze the achievable rate region. Moreso, we focus on how \(L\) depends on the size of the portion to be reused, \(J\), thereby establishing the malleability rate \(M\); recall \(M = L - J\). In particular, we fix \(J\) and find the best \(L\); the smallest \(L\) is denoted \(L^*(J)\).

**B. Two Achievable Points**

It is easy to note 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 \(X_i^n\) has to be preserved, \(L^*(H(X)) = H(X, Y)\).

Since the first \(H(X)\) symbols are fixed, losslessly representing the conditionally typical set 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 and creating an entirely new codeword for \(Y_i^n\); unless \(X\) and \(Y\) are independent, this is strictly better.
We cast the fixed reuse malleable coding problem as a single-letter information-theoretic optimization. This is not computable in general, but see Sec. V.

A proof of the Slepian-Wolf theorem uses the method of binning [3], [4], where the codebooks are segmented and codewords are binned. Results are obtained by choosing appropriate bin sizes. However, this approach says nothing about whether or how labels are kept synchronized between the different codebooks and bins. We apply a tree-structured binning approach, but insist on consistent representation to enforce malleability in the representations.

Consider the trade-off between \( L \) and \( J \). For a given malleability, the compression efficiency of \( Y^n \) is determined by the quality of the binning in the codebook for \( X^n \). We insist that \( U \) is a deterministic function of \( X \), i.e., \( U = f(X) \). Then, we can formulate the following information-theoretic optimization problem:

\[
    L^*(J) - J = \min_{U:U = f(X), H(U) = J} H(Y|U). \tag{1}
\]

**Theorem 1:** The optimization problem (1) provides a boundary to the rate region \( \mathcal{M} \) when \( K = H(X) \).

For clarity, before stating the proof to Theorem 1 we describe the dimensions and alphabets of the codebooks used.

1) Numbers \( K \) and \( J \) are given. The first codebook is used to encode a source sequence of length \( n \), \( x^n \).

Let \( \mathcal{C} = \{c_1, c_2, \ldots, c_n\} \) be the prefix-stage codebook of size \( \rho_u = |\mathcal{V}|^{nJ} \), drawn from the alphabet \( \mathcal{V} \). Corresponding to every codeword \( c_i \in \mathcal{C} \), let \( \mathcal{A}(c_i) = \{a_1(c_i), a_2(c_i), \ldots, a_{\rho_x}(c_i)\} \) be the suffix-stage codebook of size \( \rho_x = |\mathcal{V}|^{n(K-J)} \), drawn from the alphabet \( \mathcal{V} \). The whole codebook for \( x^n \) is then \( \mathcal{A} = \bigcup_{i=1}^{n} \mathcal{A}(c_i) \), which is a tree-structured codebook of size \( |\mathcal{V}|^{nK} \).

2) The prefix-stage codebook \( \mathcal{C} \) from above and a number \( L \) are given. The second codebook is used to encode a source sequence of length \( n \), \( y^n \). Corresponding to every codeword \( c_i \in \mathcal{C} \), let \( \mathcal{B}(c_i) = \{b_1(c_i), b_2(c_i), \ldots, b_{\rho_y}(c_i)\} \) be the suffix-stage codebook of size \( \rho_y = |\mathcal{V}|^{n(L-J)} \), drawn from the alphabet \( \mathcal{V} \). The whole codebook for \( y^n \) is then \( \mathcal{B} = \bigcup_{i=1}^{n} \mathcal{B}(c_i) \), which is a tree-structured codebook of size \( |\mathcal{V}|^{nL} \).

The two codebooks share the first level of the tree, but have different second levels.

The proof of Theorem 1 uses a lemma of Körner [5], reproduced as [6, Lemma 1].

**Proof:** Fix a function \( f \) that partitions \( \mathcal{W} \). This function is used to induce a random variable \( U_i = f(X_i) \). The function \( f \) is applied to all \( X^n \) in the same manner to produce the memoryless random variables \( U^n \).

a) **Generating the first codebook:** Choose the prefix part codebook rate as \( J = \frac{1}{n} \log|\mathcal{V}| \rho_u = H(U) + \delta_1(n) \), where \( \delta_1(n) \to 0 \) as \( n \to \infty \). Generate a set of size \( |\mathcal{V}|^{nJ} \) of sequences in \( \mathcal{W}^n \) with elements drawn i.i.d. according to \( p_U \). Now take these sequences in order and create a codebook \( \mathcal{C} \) with codewords from \( \mathcal{V}^{nJ} \) listed in lexicographic order, by making a one-to-one correspondence between the two sets (which are of the same size).

b) **Encoding the first version:** For a source realization \( x^n \), compute \( U^n = f(x^n) \). If \( U^n \) is represented in the codebook \( \mathcal{C} \), then its corresponding codeword is written to the storage medium in the prefix-part position. If \( U^n \) is not represented in the codebook, then a codeword in \( \mathcal{C} \) is chosen uniformly at random from \( \mathcal{C} \) and written to the storage medium in the prefix-part position.

c) **Decoding the first version:** Decoding is performed using lookup in \( \mathcal{A} \) to generate \( \hat{x}^n \) in \( \mathcal{W}^n \).

d) **Error analysis for first version:** The two possible error events are the following:

1) \( \mathcal{E}_1: U^n \) is not represented in \( \mathcal{C} \); and
2) \( \mathcal{E}_2: U^n \) is represented by \( c_{u'n} \in \mathcal{C} \), but \( x^n \) is not represented in \( \mathcal{A}(c_{u'n}) \).

The codebook \( \mathcal{C} \) represents \( |\mathcal{V}|^{n[H(U)+\delta_1(n)]} \) sequences generated i.i.d. according to \( p_U \). The probability that a source sequence \( u^n \) generated i.i.d. according to \( p_U \) is identical to the first codeword of the codebook is bounded as \( |\mathcal{W}|^{-n} \), by memorylessness and the length of the codebook.

Since these identifiability events are independent, for a codebook of size \( |\mathcal{V}|^{n[H(U)+\delta_1(n)]} \), the probability of \( \mathcal{E}_1 \) is therefore bounded as

\[
    Pr[\mathcal{E}_1] \leq 1 - \left[1 - |\mathcal{W}|^{-n}\right]^{|\mathcal{V}|^{n[H(U)+\delta_1(n)]}}
\]

which goes to zero as \( n \to \infty \).

Furthermore, Körner’s lemma guarantees that \( \Pr[\mathcal{E}_2] \to 0 \) as \( n \to \infty \). Thus by the union bound, the total error probability goes to zero asymptotically.

e) **Converse arguments for first version:** By the converse of the source coding theorem, the size of \( \mathcal{C} \) cannot be chosen smaller than \( H(U) \) to drive the error probability to zero as

\[
    K \approx H(U) + H(X|U) = H(U) + H(X,U) - H(X|U) = H(X) - H(U) + H(X|U)
\]

where (a) is due to the chain rule of entropy and (b) is due to the fact that \( f(\cdot) \) is a deterministic function.

The codebook \( \mathcal{A} = [\mathcal{C}, \mathcal{A}'] \) is revealed to both the encoder and decoder.
\( n \to \infty \). By the converse part of Körner’s lemma, the suffix-part of the code cannot be chosen smaller than \( H(X|U) \) to drive the error probability to zero as \( n \to \infty \).

\( f) \) Decoding the prefix for use with the second version: The prefix-part is preserved in its entirety on the storage medium, therefore \( c \) is identical to above. For a given block-length \( n \), it can be used to decode \( u_1^n \) with an error probability \( \Pr[\xi_1^n] = \epsilon, \epsilon(n) \to 0 \) as \( n \to \infty \). The decoded version is called \( \hat{u}_1^n \) : note that \( \hat{U}_1^n \) is a memoryless sequence of random variables because the codebook \( C \) is a random codebook with i.i.d. \( p_U \) entries and since error events lead to a uniformly random choice of codeword within \( C \).

\( g) \) Generating the second codebook: The prefix part is drawn from the same codebook \( C \) as above. For the suffix part, consider generating the codebook according to the memoryless random variable \((Y_1^n, U_1^n)\) when the decoder is assumed to have side information \( \hat{U}_1^n \). Since \( g(Y, \hat{U}) = \hat{U} \) is a function that partitions the space, we can use Körner’s optimal complementary code as the suffix-part code \( B' \). As given in Körner’s lemma, it should have rate \( L - J = \frac{1}{2} \log |\mathcal{V}| \rho_Y = H(Y|U) \).

By the continuity argument [6, Appendix A], \( H(Y|\hat{U}) = H(Y|U) \to 0 \) as \( n \to \infty \), so we can take \( L - J = H(Y|U) \).

The codebook \( B = [C, B'] \) is revealed to both the encoder and decoder.

\( h) \) Encoding the second version: The prefix part is as for the first version, \( b_1^n = c_{\epsilon_1^n} \).

For the suffix-part \( b_{n+1}^L \), let \( \hat{u}_1^n \) be represented by \( c_{\hat{u}_1^n} \in C \). If \( y_1^n \) is represented in the codebook \( B'(c_{\hat{u}_1^n}) \), then its corresponding codeword is written to the storage medium. If \( y_1^n \) is not represented in the codebook \( B'(c_{\hat{u}_1^n}) \), then the all-zeros sequence in \( V^{n(L-J)} \) is written to the suffix-part position of the storage medium.

\( i) \) Decoding the second version: Decoding is performed using lookup in \( B \) to generate \( \hat{y}_1^n \in \mathcal{W}^n \).

\( j) \) Error analysis for second version: There is one possible error event:

1) \( \xi_2^n; y_1^n \) is not represented in \( B'(c_{\hat{u}_1^n}) \).

Körner’s lemma guarantees that \( \Pr[\xi_2^n] \to 0 \) as \( n \to \infty \).

\( k) \) Converse arguments for second version: By the converse part of Körner’s lemma, the suffix-part of the codebook cannot be chosen smaller than \( H(Y|U) \) to drive the error probability to zero as \( n \to \infty \).

V. FURTHER CHARACTERIZATIONS

As in the source coding with side information problem \[7\]–\[9\] and elsewhere, Theorem 1 left us to optimize an auxiliary random variable \( U \) that describes the method of binning. Here we give further characterization in terms of \( W \), a minimal 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. 3. The bounds, marked by dotted lines, are as follows:

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

\( \text{(b)} \) A trivial lower bound by construction is \( L^*(J) \geq J \).

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

A. Convexity of Regime

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 \( U_1^{n-J} \), 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.

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

\[ L^*(J) = H(Y|W). \]

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

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

This is used to draw the portion of the boundary determined by Theorem 2 with a solid line in Fig. 3.

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

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

\[ \text{Proof: Follows from the convexity of conditional entropy, by mixing possible distributions } U. \]
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.

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

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

The following are extreme cases of 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 $\frac{d}{dJ} L^*(J) = 0$.

**VI. Connections**

An alternate method of further analyzing the rate-malleability region is to make connections with solved problems in the literature.

A source coding problem intimately related to the the Gács–Körner common information provides a partial converse. A seemingly related problem of Vasudevan and Perron [10] does not provide too much further insight into the rate-malleability region. Relating their problem statement to our problem statement requires the rate $R_1$ in our problem to be set to 0 and the decoder for $Y$ to decode both $(X, Y)$.

**A. Relation to Gács–Körner Common Information**

The Gács–Körner common information [11], helps characterize the rate-malleability region. It also arises in lossless coding with coded side information [9].

**Definition 2:** 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 3:** 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.

**Lemma 1 ([11]):** Common information $C(X; Y) = 0$ if and only if $X$ and $Y$ have an indecomposable distribution.

**Lemma 2:** Consider the source network [4, Fig. P.28 on p. 403], redrawn as Fig. 4. With or without the dashed line, the largest $R_0$ for which the rate triple $(R_0, R_1 = H(X) - R_0, R_2 = H(Y) - R_0)$ is achievable (with Shannon reliability) is $R_0 = C(X; Y)$.

**Proof:** See [4, P28 on p. 404] without the dashed line. With the dashed line, the result follows additionally from the Markov relation $U \leftrightarrow X \leftrightarrow Y$, so additional knowledge of $U$ provides no benefit to $f_E(Y)$.

The Gács–Körner common information allows a characterization of the malleable coding problem.

**Theorem 5:** The rate triple $(R_0 = C(X; Y), R_1 = H(X) - C(X; Y), R_2 = H(Y) - C(X; Y))$ provides a partial converse to the rate-malleability triple.

**Proof:** In our block-diagrammatic convention, more lines and less noisy channel boxes signify more extensive information patterns. The source network in Fig. 4, has a more extensive information pattern than the malleable coding problem (see Fig. 2). Thus, the result follows from Lemma 2.

An implication is that if want $K = H(X)$ and $L = H(Y)$ for the malleable coding problem, then $M$ must be large: $M \geq H(Y) - C(X; Y)$. In general $C(X; Y) = 0$ by Lemma 1, so in this case the stored symbols cannot be reused at all.

**VII. DISCUSSIONS AND CLOSING REMARKS**

We have formulated an information-theoretic problem motivated by transmission to edit the compressed version of a document after it has been updated. Theorem 1 provides a complete characterization as a single-letter information-theoretic optimization. A relationship with Gács–Körner common information, shows that in general, if the original and modified sources are required to be coded close to entropy, then the reused fraction must asymptotically be negligible.

**ACKNOWLEDGMENT**

The authors thank R. Johari, S. K. Mitter, I. E. Telatar, V. Tarokh, and D. Marco.

**REFERENCES**


