Forty-Eighth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 1, 2010

Lossy Communication of Correlated Sources over Multiple Access Channels

Sung Hoon Lim
Department of EE
KAIST
Daejeon, Korea
sunghlim@kaist.ac.kr

Paolo Minero
Department of ECE
UCSD
La Jolla, CA 92093, USA
pminero@ucsd.edu

Young-Han Kim
Department of ECE
UCSD
La Jolla, CA 92093, USA
yhk@ucsd.edu

***Abstract*—A new approach to joint source–channel coding is presented in the context of communicating correlated sources over multiple access channels. Similar to the separation architecture, the joint source–channel coding system architecture in this approach is modular, whereby the source encoding and channel decoding operations are decoupled. However, unlike the separation architecture, the same codeword is used for both source coding and channel coding, which allows the resulting coding scheme to achieve the performance of the best known schemes despite its simplicity. In particular, it recovers as special cases previous results on lossless communication of correlated sources over multiple access channels by Cover, El Gamal, and Salehi, distributed lossy source coding by Berger and Tung, and lossy communication of the bivariate Gaussian source over the Gaussian multiple access channel by Lapidoth and Tinguely. The proof of achievability involves a new technique for analyzing the probability of decoding error when the message index depends on the codebook itself. Applications of the new joint source–channel coding system architecture in other settings are also discussed.**

I. Problem Statement and the Main Result

Consider the problem of communicating a pair of correlated discrete memoryless sources (2-DMS) (S_1, S_2) over a discrete memoryless multiple access channel (DM-MAC) $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$ as depicted in Fig. 1. Here each sender $j = 1, 2$ wishes to communicate its source S_j to a common receiver so the sources can be reconstructed with desired distortions. We will consider the block coding setting in which the source sequences $S_1^n = (S_{11}, \ldots, S_{1n})$ and $S_2^n = (S_{21}, \ldots, S_{2n})$ are communicated by n transmissions over the channel.

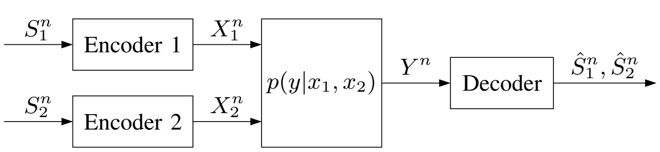


Fig. 1. Communication of a 2-DMS over a DM-MAC.

Formally, a $(|\mathcal{S}_1|^n, |\mathcal{S}_2|^n, n)$ joint source–channel code consists of

- two encoders, where encoder $j = 1, 2$ assigns a sequence $x_j^n(s_j^n) \in \mathcal{X}_j^n$ to each sequence $s_j^n \in \mathcal{S}_j^n$, and
- a decoder that assigns an estimate $(\hat{s}_1^n, \hat{s}_2^n) \in \hat{\mathcal{S}}_1^n \times \hat{\mathcal{S}}_2^n$ to each sequence $y^n \in \mathcal{Y}^n$.

Let $d_1(s_1, \hat{s}_1)$ and $d_2(s_2, \hat{s}_2)$ be two distortions measures. The average per-letter distortion $d_j(s_j^n, \hat{s}_j^n)$, $j = 1, 2$, is defined as $d(s_j^n, \hat{s}_j^n) = (1/n)\sum_{i=1}^n d(s_{ji}, \hat{s}_{ji})$. A distortion pair (D_1, D_2) is said to be achievable for communication of the 2-DMS (S_1, S_2) over the DM-MAC $p(y|x_1, x_2)$ if there exists a sequence of $(|\mathcal{S}_1|^n, |\mathcal{S}_2|^n, n)$ joint source–channel codes such that

$$\limsup_{n\to\infty} \mathsf{E}(d_j(S_j^n, \hat{S}_j^n)) \le D_j, \quad j = 1, 2.$$

The optimal distortion region $\mathscr{D}^*$ is the closure of the set of all achievable distortion pairs (D_1, D_2).

A computable characterization of the optimal distortion region is not known in general. This paper establishes the following inner bound on the optimal distortion region. For simplicity, we will assume that the sources S_1 and S_2 have no common part in the sense of Gács–Körner [1] and Witsenhausen [2].

Theorem 1: A distortion pair (D_1, D_2) is achievable for communication of the 2-DMS (S_1, S_2) without common part over a DM-MAC $p(y|x_1, x_2)$ if

$$\begin{aligned} I(U_1; S_1|Q) &< I(U_1; Y, U_2|Q), \\ I(U_2; S_2|Q) &< I(U_2; Y, U_1|Q), \\ I(U_1; S_1|Q) + I(U_2; S_2|Q) &< I(U_1, U_2; Y|Q) + I(U_1; U_2|Q) \end{aligned}$$

for some pmf $p(s_1, s_2)p(q)p(u_1, x_1|s_1, q)p(u_2, x_2|s_2, q)$ and functions $\hat{s}_1(u_1, u_2, y, q)$ and $\hat{s}_2(u_1, u_2, y, q)$ such that $\mathsf{E}(d_j(S_j, \hat{S}_j)) \le D_j$, $j = 1, 2$.

Here and throughout, we use notation in [3].

As we will see in Section II, Theorem 1 includes previous results on lossless communication of a 2-DMS over a DM-MAC by Cover, El Gamal, and Salehi [4], distributed lossy source coding of a 2-DMS by Berger [5] and Tung [6], and lossy communication of a bivariate Gaussian source over a Gaussian MAC by Lapidoth and Tinguely [7]. The main contribution of the paper, however, lies not with the generality of Theorem 1 that unifies these results, but with a simple joint source–channel coding system architecture that is used in the proof of achievability. The new joint source–channel coding scheme is very similar to *separate* source and channel coding, except that a single codeword is used for both source and channel coding.

In the next section, we digress a bit and show how Theorem 1 recovers the aforementioned prior results as special cases. The new joint source–channel coding system architecture is described first in the simple point-to-point communication setting in Section III and then in the multiple access setting for Theorem 1 in Section IV. Potential applications of this new joint source–channel coding system architecture are discussed in Section V.

II. Special Cases

A. Lossless Communication

When specialized to the lossless case, wherein d_1, d_2 are Hamming distortion measures and $D_1 = D_2 = 0$, Theorem 1 reduces to the following sufficient condition for lossless communication of a 2-DMS over a DM-MAC.

Corollary 1 (Cover, El Gamal, and Salehi [4]): A 2-DMS (S_1, S_2) can be communicated losslessly over a DM-MAC $p(y|x_1, x_2)$ if

$$\begin{aligned} H(S_1|S_2) &< I(X_1; Y|X_2, S_2, Q), \\ H(S_2|S_1) &< I(X_2; Y|X_1, S_1, Q), \\ H(S_1, S_2) &< I(X_1, X_2; Y, Q) \end{aligned}$$

for some pmf $p(q, x_1, x_2|s_1, s_2) = p(q)p(x_1|s_1, q)p(x_2|s_2, q)$.

The proof follows by setting $U_j = (X_j, S_j)$ and $\hat{S}_j = S_j$, $j = 1, 2$, in Theorem 1. The details are given in Appendix A.

B. Distributed Lossy Source Coding

When specialized to a noiseless MAC $Y = (X_1, X_2)$ with $\log|\mathcal{X}_1| = R_1$ and $\log|\mathcal{X}_2| = R_2$, Theorem 1 reduces to the following inner bound on the rate–distortion region for distributed lossy source coding.

Corollary 2 (Berger [5] and Tung [6]): A distortion pair (D_1, D_2) is achievable for distributed lossy source coding of a 2-DMS (S_1, S_2) with rate pair (R_1, R_2) if

$$\begin{aligned} R_1 &> I(S_1; V_1|V_2, Q), \\ R_2 &> I(S_2; V_2|V_1, Q), \\ R_1 + R_2 &> I(S_1, S_2; V_1, V_2|Q) \end{aligned}$$

for some pmf $p(q)p(v_1|s_1, q)p(v_2|s_2, q)$ and functions $\hat{s}_1(v_1, v_2, q)$ and $\hat{s}_2(v_1, v_2, q)$ such that $\mathsf{E}(d_j(S_j, \hat{S}_j)) \le D_j$, $j = 1, 2$.

The proof follows by setting $U_j = (X_j, V_j)$, $j = 1, 2$. The details are given in Appendix B.

C. Bivariate Gaussian Source over a Gaussian MAC

Suppose the sources are bivariate Gaussian with $(S_1, S_2) \sim \mathrm{N}(0, K_{\mathbf{S}})$, where

$$K_{\mathbf{S}} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix},$$

and is to be reconstructed under the quadratic distortion measure $d_j(s_j, \hat{s}_j) = (s_j - \hat{s}_j)^2$, $j = 1, 2$. Further assume the channel is the Gaussian MAC $Y = X_1 + X_2 + Z$ with $Z \sim \mathrm{N}(0, N)$ and input power constraints $\sum_{i=1}^n \mathsf{E}(x_{ji}^2(S_j^n)) \le nP_j$, $j = 1, 2$. Theorem 1 can be adapted to this case via the standard discretization method [3, Lecture Note 3].

Given $\alpha_j \in [0, \sqrt{P_j/\sigma^2}]$ and $R_j > 0$, $j = 1, 2$, let $Q = \emptyset$, $U_j = (1 - 2^{-2R_j})S_j + \hat{Z}_j$, $j = 1, 2$, and $X_j = \alpha_j S_j + \beta_j U_j$, $j = 1, 2$, where $\hat{Z}_j$ are independent Gaussian random variables with zero mean and variance $\sigma^2 2^{-2R_j}(1 - 2^{-2R_j})$, and

$$\beta_j = \sqrt{\frac{P_j - \alpha_j^2\sigma^2 2^{-2R_j}}{\sigma^2(1 - 2^{-2R_j})}} - \alpha_j. \tag{1}$$

Let

$$K(\alpha_1, \alpha_2, R_1, R_2) = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{12} & k_{22} & k_{23} \\ k_{13} & k_{23} & k_{33} \end{bmatrix}$$

denote the covariance matrix of (U_1, U_2, Y), where

$$\begin{aligned} k_{jj} &= \sigma^2(1 - 2^{-2R_j}), \quad j = 1, 2, \\ k_{12} &= \sigma^2\rho(1 - 2^{-2R_1})(1 - 2^{-2R_2}), \\ k_{13} &= (\alpha_1 + \beta_1 + \alpha_2\rho)k_{11} + \beta_2 k_{12}, \\ k_{23} &= (\alpha_2 + \beta_2 + \alpha_1\rho)k_{22} + \beta_1 k_{12}, \\ k_{33} &= (\alpha_1^2 + 2\alpha_1\alpha_2\rho + \alpha_2^2)\sigma^2 + (2\alpha_1\beta_1 + \beta_1^2 + 2\beta_1\alpha_2\rho)k_{11} \\ &\quad + (2\alpha_1\beta_2\rho + 2\alpha_2\beta_2 + \beta_2^2)k_{22} + 2\beta_1\beta_2 k_{12} + N. \end{aligned}$$

Then Theorem 1 reduces to the following sufficient condition for lossy communication.

Corollary 3 (Lapidoth and Tinguely [7]): The distortion pair (D_1, D_2) is achievable if

$$D_j > \sigma^2 - \gamma_{j1}c_{j1} - \gamma_{j2}c_{j2} - \gamma_{j3}c_{j3}, \quad j = 1, 2,$$

for some $\alpha_j \in [0, \sqrt{P_j/\sigma^2}]$ and $R_j > 0$, $j = 1, 2$, such that

$$\begin{aligned} R_1 &< \frac{1}{2}\log\left(\frac{{\beta_1'}^2 k_{11}(1 - \tilde{\rho}^2) + N'}{N'(1 - \tilde{\rho}^2)}\right), \\ R_2 &< \frac{1}{2}\log\left(\frac{{\beta_2'}^2 k_{22}(1 - \tilde{\rho}^2) + N'}{N'(1 - \tilde{\rho}^2)}\right), \\ R_1 + R_2 &< \frac{1}{2}\log\left(\frac{{\beta_1'}^2 k_{11} + {\beta_2'}^2 k_{22} + 2\tilde{\rho}{\beta_1'}^2{\beta_1'}^2\sqrt{k_{11}k_{22}} + N'}{N'(1 - \tilde{\rho}^2)}\right), \end{aligned}$$

where $c_{11} = k_{11}$, $c_{12} = \rho k_{22}$, $c_{21} = \rho k_{11}$, $c_{22} = k_{22}$, $c_{13} = (\alpha_1 + \alpha_2\rho)\sigma^2 + \beta_1 k_{11} + \beta_2\rho k_{22}$, $c_{23} = (\alpha_2 + \alpha_1\rho)\sigma^2 + \beta_1\rho k_{11} + \beta_2 k_{22}$,

$$\begin{bmatrix} \gamma_{j1} \\ \gamma_{j2} \\ \gamma_{j3} \end{bmatrix} = K^{-1}(\alpha_1, \alpha_2, R_1, R_2)\begin{bmatrix} c_{j1} \\ c_{j2} \\ c_{j3} \end{bmatrix},$$

and

$$\begin{aligned}
\tilde{\rho} &= \rho\sqrt{(1-2^{-2R_1})(1-2^{-2R_2})},\\
N' &= \alpha_1^2\nu_1 + \alpha_2^2\nu_2 + 2\alpha_1\alpha_2\nu_3 + N,\\
\nu_1 &= \sigma^2 - (1 - a_1\rho(1-2^{-2R_2}))^2 k_{11}\\
&\quad - 2(1 - a_1\rho(1-2^{-2R_2}))a_1 k_{12} - a_1^2 k_{22},\\
\nu_2 &= \sigma^2 - (1 - a_2\rho(1-2^{-2R_1}))^2 k_{22}\\
&\quad - 2(1 - a_2\rho(1-2^{-2R_1}))a_2 k_{12} - a_2^2 k_{11},\\
\nu_3 &= \rho\sigma^2 - (1 - a_1\rho(1-2^{-2R_2}))(1 - a_2\rho(1-2^{-2R_1})k_{12}\\
&\quad - a_1 a_2 k_{12} - (1 - a_1\rho(1-2^{-2R_2}))a_2 k_{11}\\
&\quad - (1 - a_2\rho(1-2^{-2R_1}))a_1 k_{22},\\
\beta_1' &= \alpha_1\left(1 - \frac{\rho^2 2^{-2R_1}(1-2^{-2R_2})}{1-\tilde{\rho}^2}\right) + \beta_1 + \frac{\alpha_2\rho 2^{-2R_2}}{1-\tilde{\rho}^2},\\
\beta_2' &= \alpha_2\left(1 - \frac{\rho^2 2^{-2R_2}(1-2^{-2R_1})}{1-\tilde{\rho}^2}\right) + \beta_2 + \frac{\alpha_1\rho 2^{-2R_1}}{1-\tilde{\rho}^2}.
\end{aligned}$$

III. A New Joint Source–Channel Coding System Architecture

Shannon [8] studied the general point-to-point joint source–channel coding problem in Fig. 2 and established the sufficient and necessary condition for the discrete memoryless source S to be communicated over the discrete memoryless channel $p(y|x)$ with prescribed distortion D.

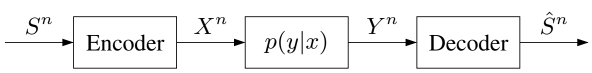


Fig. 2. Point-to-point joint source–channel coding.

When translated to our problem setting, Shannon's sufficient condition states that a distortion D is achievable if

$$R(D) < C, \tag{2}$$

where

$$R(D) = \min_{p(\hat{s}|s):\, \mathsf{E}(d(S,\hat{S}))\le D} I(S;\hat{S})$$

is the rate–distortion function for the source S and distortion measure $d(s,\hat{s})$ and

$$C = \max_{p(x)} I(X;Y)$$

is the capacity of the channel $p(y|x)$. The proof of this result uses separate source and channel coding, as illustrated in Fig. 3(a). Under this separate source and channel coding architecture, the source sequence is mapped into one of 2^{nR} indices M and then this index is mapped into a channel codeword X^n, which is transmitted over the channel. Upon receiving Y^n, the decoder finds an estimate $\hat{M}$ of M and reconstructs $\hat{S}^n$ from $\hat{M}$. The index M provides a digital interface between the source code and the channel code, which can be designed and operated separately. By the lossy source coding theorem and the channel coding theorem, the desired distortion D can be achieved, provided that the index rate R satisfies $R > R(D)$ and $R < C$.

We now propose the joint source–channel coding system architecture in Fig. 3(b), which closely resembles the above source–channel separation architecture. Under this new architecture, the source sequence S^n is mapped to one of 2^{nR} sequences $U^n(M)$ and then this sequence $U^n(M)$ (along with S^n) is mapped to X^n *symbol-by-symbol*, which is transmitted over the channel. Upon receiving Y^n, the decoder finds an estimate $U^n(\hat{M})$ of $U^n(M)$ and reconstructs $\hat{S}^n$ from U^n (and Y^n) again by a *symbol-by-symbol* mapping. Thus, the codeword $U^n(M)$ plays the roles of both the source codeword $\hat{S}^n(M)$ and the channel codeword $X^n(M)$ simultaneously. This dual role of $U^n(M)$ allows simple symbol-by-symbol interfaces $x(u,s)$ and $\hat{s}(u,y)$ that replace the channel encoder and the source decoder in the separation architecture. Moreover, the source encoder and the channel decoder can be operated separately. Roughly speaking, again by the lossy source coding theorem, the condition $R > I(U;S)$ guarantees a reliable source encoding operation and by the channel coding theorem, the condition $R < I(U;Y)$ guarantees a reliable channel decoding operation (over the channel $p(y|u) = \sum_s p(y|x(u,s))p(s)$). Thus, a distortion D is achievable if

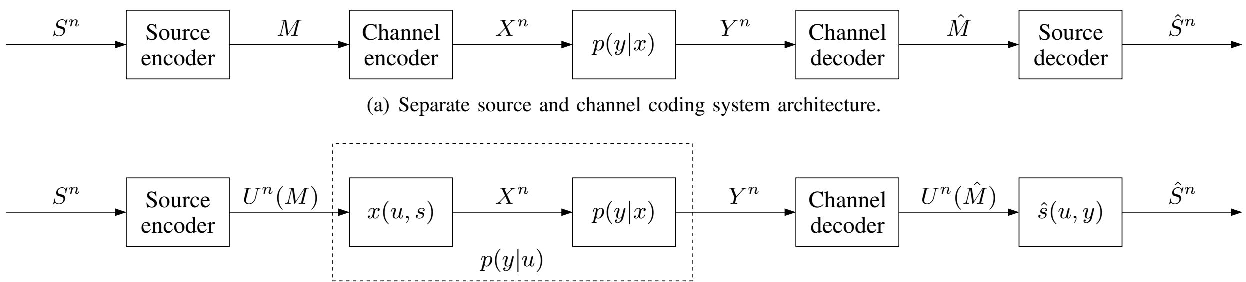
$$I(S;U) < I(U;Y) \tag{3}$$

for some pmf $p(u|s)$ and functions $x(u,s)$ and $\hat{s}(u,y)$ such that $\mathsf{E}(d(S,\hat{S})) \le D$. By taking $U = (X,S)$, where $X \sim p(x)$ is independent of S, and using the memoryless property of the channel, it can be easily shown that this condition simplifies to (2).

Conceptually speaking, this new coding scheme is as simple as the separation scheme and hence will be used as a basic building block for the joint source–channel coding system architecture for communicating a 2-DMS over a DM-MAC in the next section. The precise analysis of its performance involves a technical subtlety, however. In particular, because $U^n(M)$ is used as a source codeword, the index M depends on the entire codebook $\mathcal{C} = \{U^n(M) : M \in [1:2^{nR}]\}$. But the conventional random coding proof technique for a channel codeword $U^n(M)$ is developed for situations for which the index M and the (random) codebook $\mathcal{C}$ are independent of each other. The dependency issue for joint source–channel coding has been well noted by Lapidoth and Tinguely [7, Proof of Proposition D.1], who developed a geometric approach for the Gaussian setup discussed in Subsection II-C to avoid this difficulty. In the following, we provide a formal proof of the sufficient condition (3) along with a new analysis technique that handles this subtle point. The standard proof steps are skipped, as these can be found in [3, Lecture Note 3].

Codebook generation: Fix $p(x,u|s)$ and $\hat{s}(u,y)$. Randomly and independently generate 2^{nR} sequences $u^n(m)$, $m \in [1:2^{nR}]$, each according to $\prod_{i=1}^n p_U(u_i)$. The codebook $\mathcal{C} = \{u^n(m) : m \in [1:2^{nR}]\}$ is revealed to both the encoder and the decoder.

Encoding: Fix $\epsilon' > 0$. We use joint typicality encoding. Upon observing a sequence s^n, the encoder finds an index m such



(a) Separate source and channel coding system architecture.

(b) A new joint source–channel coding system architecture.

Fig. 3. Two system architectures for the problem of lossy transmission of a source over a point to point channel.

that $(u^n(m), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one such index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR}]$. The encoder then transmits $x_i = x(u_i(m), s_i)$ for $i \in [1 : n]$.

Decoding: Upon receiving y^n, the decoder finds the unique index $\hat{m}$ such that $(u^n(\hat{m}), y^n) \in \mathcal{T}_\epsilon^{(n)}$. If there is none or more than one, it chooses an arbitrary index. The decoder then sets the reproduction sequence as $\hat{s}_i = \hat{s}(u_i(\hat{m}), y_i)$ for $i \in [1 : n]$.

Analysis of the expected distortion: Let $\epsilon' < \epsilon$. We bound the distortion averaged over S^n and the random choice of the codebook $\mathcal{C}$. Let M be the random variable denoting the chosen index at the encoder. Define the "error" event

$$\mathcal{E} = \{(S^n, U^n(\hat{M}), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\}$$

and partition it into

$$\begin{aligned}
\mathcal{E}_1 &= \{(U^n(m), S^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m\},\\
\mathcal{E}_2 &= \{(S^n, U^n(M), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\},\\
\mathcal{E}_3 &= \{(U^n(\tilde{m}), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } \tilde{m} \neq M\}.
\end{aligned}$$

Then by the union of events bound,

$$\mathsf{P}(\mathcal{E}) \le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2 \cap \mathcal{E}_1^c) + \mathsf{P}(\mathcal{E}_3).$$

We show that all three terms tend to zero as $n \to \infty$. This implies that the probability of "error" tends to zero as $n \to \infty$, which, in turn, implies that, by the law of total expectation and the typical average lemma [3, Lecture Note 2],

$$\begin{aligned}
&\limsup_{n\to\infty} \mathsf{E}(d(S^n, \hat{S}^n))\\
&\le \limsup_{n\to\infty} \big(\mathsf{P}(\mathcal{E})\, \mathsf{E}(d(S^n, \hat{S}^n)|\mathcal{E}) + \mathsf{P}(\mathcal{E}^c)\, \mathsf{E}(d(S^n, \hat{S}^c)|\mathcal{E}^c)\big)\\
&\le (1+\epsilon)\, \mathsf{E}(d(S, \hat{S})),
\end{aligned}$$

and hence the desired distortion is achieved.

By the covering lemma and the conditional typicality lemma [3, Lecture Notes 2 and 3], it can be easily shown that the first two terms tend to zero as $n \to \infty$ if $R > I(U;S) + \delta(\epsilon)$. The third term requires special attention. Consider

$$\begin{aligned}
&\mathsf{P}\{(U^n(\tilde{m}), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{m} \neq M\}\\
&\overset{(a)}{\le} \sum_{\tilde{m}=1}^{2^{nR}} \mathsf{P}\{(U^n(\tilde{m}), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M \neq \tilde{m}\}\\
&= \sum_{\tilde{m}=1}^{2^{nR}} \sum_{s^n} p(s^n)\, \mathsf{P}\{(U^n(\tilde{m}), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M \neq \tilde{m} \,|\, S^n = s^n\}\\
&\overset{(b)}{=} 2^{nR} \sum_{s^n} p(s^n)\, \mathsf{P}\{(U^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M \neq 1 \,|\, S^n = s^n\},
\end{aligned}$$

where (a) follows by the union of events bound and (b) follows by the symmetry of the codebook generation and encoding. Note that unlike in the conventional proof of the channel coding theorem [3, Lecture Note 1] where the event is analyzed conditioned on the event $M = 1$, here the event of interest is $M \neq 1$. Let $\bar{\mathcal{C}} = \mathcal{C} \backslash \{U^n(1)\}$. Then, for n sufficiently large,

$$\begin{aligned}
&\mathsf{P}\{(U^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M \neq 1 \,|\, S^n = s^n\}\\
&\le \mathsf{P}\{(U^n(1), Y^n) \in \mathcal{T}_\epsilon^{(n)} \,|\, M \neq 1, S^n = s^n\}\\
&= \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} \mathsf{P}\{U^n(1) = u^n, Y^n = y^n \,|\, M \neq 1, S^n = s^n\}\\
&= \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} \sum_{\bar{c}} \mathsf{P}\{U^n(1) = u^n, Y^n = y^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\
&\qquad\qquad \cdot \mathsf{P}\{\bar{\mathcal{C}} = \bar{c} \,|\, M \neq 1, S^n = s^n\}\\
&\overset{(a)}{=} \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} \sum_{\bar{c}} \mathsf{P}\{U^n(1) = u^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\
&\qquad\qquad \cdot \mathsf{P}\{Y^n = y^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\
&\qquad\qquad \cdot \mathsf{P}\{\bar{\mathcal{C}} = \bar{c} \,|\, M \neq 1, S^n = s^n\}\\
&\overset{(b)}{\le} \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} \sum_{\bar{c}} 2\, \mathsf{P}\{U^n(1) = u^n\}\\
&\qquad\qquad \cdot \mathsf{P}\{Y^n = y^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\
&\qquad\qquad \cdot \mathsf{P}\{\bar{\mathcal{C}} = \bar{c} \,|\, M \neq 1, S^n = s^n\}\\
&= \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} 2\, \mathsf{P}\{U^n(1) = u^n\}\\
&\qquad\qquad \cdot \mathsf{P}\{Y^n = y^n \,|\, M \neq 1, S^n = s^n\}\\
&\overset{(c)}{\le} \sum_{(u^n,y^n)\in\mathcal{T}_\epsilon^{(n)}} 4\, \mathsf{P}\{U^n(1) = u^n\}\, \mathsf{P}\{Y^n = y^n \,|\, S^n = s^n\}
\end{aligned} \tag{4}$$

where (a) follows since, given $M \neq 1$, $U^n(1) \to (\bar{\mathcal{C}}, S^n) \to Y^n$ form a Markov chain.

To justify step (b), we prove the following.

Lemma 1: For n sufficiently large,

$$\mathsf{P}\{U^n(1) = u^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\} \leq 2\,\mathsf{P}\{U^n(1) = u^n\}.$$

Proof: We first show that

$$\mathsf{P}\{M = 1 \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\} \leq \frac{1}{2} \tag{5}$$

for n sufficiently large. Let $k = k(\bar{c}, s^n) = |\{u^n(m) \in \bar{c} : (u^n(m), s^n(\in \mathcal{T}_{\epsilon'}^{(n)}\}|$. Then, by the symmetry of the encoding procedure, if $k \geq 1$,

$$\begin{aligned}\mathsf{P}\{M = 1 \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\ = \frac{1}{k+1}\,\mathsf{P}\{(U^n(1), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}\} \leq \frac{1}{2},\end{aligned}$$

and if $k = 0$, for n sufficiently large,

$$\begin{aligned}&\mathsf{P}\{M = 1 \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\ &\leq \mathsf{P}\{(U^n(1), s^n) \in \mathcal{T}_{\epsilon'}^{(n)}\} + \frac{1}{2^{nR}}\,\mathsf{P}\{(U^n(1), s^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\}\\ &\leq 2^{-n(I(U;S)-\delta(\epsilon'))} + \frac{1}{2^{nR}}\\ &\leq \frac{1}{2}.\end{aligned}$$

Thus

$$\begin{aligned}&\mathsf{P}\{U^n(1) = u^n \,|\, M \neq 1, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\ &= \mathsf{P}\{U^n(1) = u^n \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}\\ &\quad\cdot \frac{\mathsf{P}\{M \neq 1 \,|\, U^n(1) = u^n, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}}{\mathsf{P}\{M \neq 1 \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}}\\ &\overset{(d)}{=} p(u^n)\frac{\mathsf{P}\{M \neq 1 \,|\, U^n(1) = u^n, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}}{1 - \mathsf{P}\{M = 1 \,|\, S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}}\\ &\overset{(e)}{\leq} 2\,\mathsf{P}\{U^n(1) = u^n\},\end{aligned}$$

where (d) follows from the independence of $U^n(1)$ and $(S^n, \bar{\mathcal{C}})$, and (e) follows from (5). ■

For step (c), we prove the following.

Lemma 2: For n sufficiently large,

$$\mathsf{P}\{Y^n = y^n \,|\, M \neq 1, S^n = s^n\} \leq 2p(y^n|s^n).$$

Proof: By symmetry, $\mathsf{P}\{M \neq 1 | S^n = s^n\} = (2^{nR} - 1)/2^{nR} \leq 1/2$ for n sufficiently large. Hence,

$$\begin{aligned}&\mathsf{P}\{Y^n = y^n \,|\, M \neq 1, S^n = s^n\}\\ &\quad= p(y^n|s^n)\frac{\mathsf{P}\{M \neq 1 | S^n = s^n, Y = y^n\}}{\mathsf{P}\{M \neq 1 | S^n = s^n\}}\\ &\quad\leq 2p(y^n|s^n).\end{aligned}$$

■

Continuing the upper bound on $\mathsf{P}(\mathcal{E}_3)$, by the joint typicality lemma and (4), we have for n sufficiently large,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_3) &= \mathsf{P}\{(U^n(1), Y^n) \in \mathcal{T}_{\epsilon}^{(n)} \text{ for some } 1 \neq M\}\\ &\leq 4 \cdot 2^{nR} \sum_{s^n} p(s^n)\\ &\quad\cdot \sum_{(u^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}} \mathsf{P}\{U^n(1) = u^n\} p(y^n|s^n)\\ &= 4 \cdot 2^{nR} \sum_{(u^n, y^n) \in \mathcal{T}_{\epsilon}^{(n)}} \prod_{i=1}^{n} p_U(u_i) p(y^n)\\ &\leq 4 \cdot 2^{n(R - I(U;Y) + \delta(\epsilon))},\end{aligned}$$

which tends to zero as $n \to \infty$, if $R < I(U;Y) - \delta(\epsilon)$. Therefore, the probability of "error" tends to zero as $n \to \infty$ and the average distortion over the random codebook is bounded as desired. Thus, there exists at least one sequence of codes achieving the desired distortion. This establishes the sufficient condition (3).

IV. Proof of Achievability for Theorem 1

We generalize the joint source–channel coding system architecture for point-to-point communication in Section III to the multiple access channel, as depicted in Fig. 4. As before, $U_1^n(M_1)$ and $U_2^n(M_2)$ play the dual role of codewords for source coding (joint typicality encoding of the sources S_1^n and S_2^n) and for channel coding (joint typicality decoding from the channel output Y^n).

At a high level, the proof of the achievability for the sufficient condition is rather elementary. Following the same argument as in the point-to-point case (i.e., by the covering lemma), the source encoding operation is successful if

$$\begin{aligned}R_1 &> I(U_1; S_1|Q),\\ R_2 &> I(U_2; S_2|Q).\end{aligned}$$

On the other hand, once we ignore the issue of the dependence between the indices and the codebook, by the packing lemma [3, Lecture Note 3], the channel decoding operation is successful if

$$\begin{aligned}R_1 &< I(U_1; Y, U_2|Q),\\ R_2 &> I(U_2; Y, U_1|Q),\\ R_1 + R_2 &> I(U_1, U_2; Y|Q) + I(U_1; U_2|Q).\end{aligned}$$

Hence, by eliminating the intermediate rate pair (R_1, R_2), the sufficient condition in Theorem 1 can be established.

In the following, we provide a formal proof, focusing on the steps to justify the sufficient condition for channel decoding. For simplicity, we consider the case $Q = \emptyset$. Achievability for an arbitrary Q can be proved using coded time sharing technique [3, Lecture Note 4].

Codebook generation: Fix $p(x_1, u_1|s_1)p(x_2, u_2|s_2)$ and two reconstruction functions $\hat{s}_1(u_1, u_2, y)$ and $\hat{s}_2(u_1, u_2, y)$. For $j = 1, 2$, randomly and independently generate 2^{nR_j} sequences $u_j^n(m_j)$, $m_j \in [1 : 2^{nR_j}]$, each according to $\prod_{i=1}^{n} p_{U_j}(u_{ji})$.

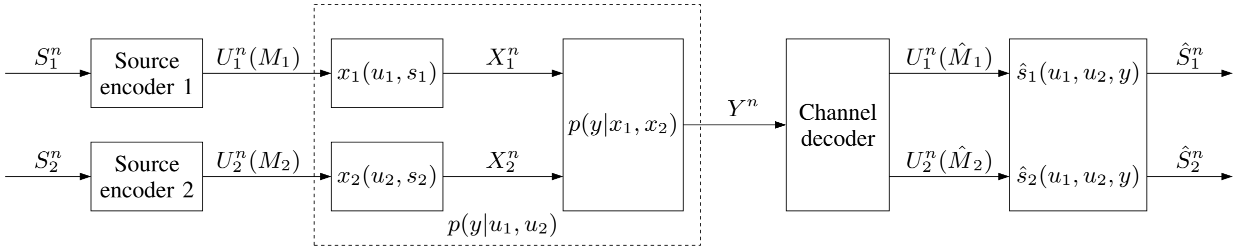


Fig. 4. Joint source–channel coding system architecture for communicating a 2-DMS over a DM-MAC.

Encoding: Fix $\epsilon' > 0$. Upon observing s_j^n, encoder $j = 1, 2$ finds an index $m_j \in [1 : 2^{nR_j}]$ such that $(s_j^n, u_j^n(m_j)) \in \mathcal{T}_{\epsilon'}^{(n)}$. If there is more than one such index, it chooses one of them at random. If there is no such index, it chooses an arbitrary index at random from $[1 : 2^{nR_j}]$. Encoder j then transmits $x_{ji}(m_j, s_{ji})$ for $i \in [1 : n]$.

Decoding: Upon receiving y^n, the decoder finds the unique index index pair $(\hat{m}_1, \hat{m}_2)$ such that $(u_1^n(m_1), u_2^n(m_2), y^n) \in \mathcal{T}_\epsilon^{(n)}$ and sets the reproduction sequence as $\hat{s}_{ji} = \hat{s}_j(u_{1i}(m_1), u_{2i}(m_2), y_i)$, $j = 1, 2$, for $i \in [1 : n]$.

Analysis of the expected distortion: Let $\epsilon' < \epsilon$. We bound the distortion averaged over (S_1^n, S_2^n) and the random codebook. Let M_1 and M_2 be random variables denoting the chosen indices at the encoders. Define the "error" events

$$\mathcal{E} = \{(S_1^n, S_2^n, U_1^n(\hat{M}_1), U_2^n(\hat{M}_2), Y^n) \notin \mathcal{T}_\epsilon^{(n)}\}$$

and partition it into

$$\begin{aligned}
\mathcal{E}_1 &= \{(U_1^n(m_1), S_1^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m_1\},\\
\mathcal{E}_2 &= \{(U_2^n(m_2), S_2^n) \notin \mathcal{T}_{\epsilon'}^{(n)} \text{ for all } m_2\},\\
\mathcal{E}_3 &= \{(S_1^n, S_2^n, U_1^n(M_1), U_2^n(M_2), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\},\\
\mathcal{E}_4 &= \{(U_1^n(\tilde{m}_1), U_2^n(M_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{m}_1 \neq M_1\},\\
\mathcal{E}_5 &= \{(U_1^n(M_1), U_2^n(\tilde{m}_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{m}_2 \neq M_2\},\\
\mathcal{E}_6 &= \{(U_1^n(\tilde{m}_1), U_2^n(\tilde{m}_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}\\
&\qquad\qquad \text{for some } \tilde{m}_1 \neq M_1, \tilde{m}_2 \neq M_2\}.
\end{aligned}$$

Then by the union of events bound,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}) &\le \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_2) + \mathsf{P}(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c)\\
&\quad + \mathsf{P}(\mathcal{E}_4) + \mathsf{P}(\mathcal{E}_5) + \mathsf{P}(\mathcal{E}_6).
\end{aligned}$$

As before, the desired distortion pair is achieved if $\mathsf{P}(\mathcal{E})$ tends to zero as $n \to \infty$. By the covering lemma, $\mathsf{P}(\mathcal{E}_1)$ and $\mathsf{P}(\mathcal{E}_2)$ tend to zero as $n \to \infty$, if

$$R_1 > I(U_1; S_1) + \delta(\epsilon'), \tag{6}$$
$$R_2 > I(U_2; S_2) + \delta(\epsilon'). \tag{7}$$

By the Markov lemma [3, Lecture Note 13], the third term tends to zero as $n \to \infty$.

To bound $\mathsf{P}(\mathcal{E}_4)$, let $S^n = (S_1^n, S_2^n)$ to simplify the notation and consider (8) at the top of the next page. Here step (a) is justified as in the point-to-point case, with $U_1^n(1)$ in place of $U^n(1)$ and $(U_2^n(M_2), Y^n)$ in place of Y^n. Hence, $\mathsf{P}(\mathcal{E}_4)$ tends to zero as $n \to \infty$, if $R_1 < I(U_1; Y, U_2) - \delta(\epsilon)$. Similarly, $\mathsf{P}(\mathcal{E}_5)$ tends to zero as $n \to \infty$, if $R_2 < I(U_2; Y, U_1) - \delta(\epsilon)$.

Finally, to bound $\mathsf{P}(\mathcal{E}_6)$, we use the similar steps to the above with the following two lemmas replacing Lemmas 1 and 2.

Lemma 3: Let $\bar{\mathcal{C}} = \{(U_1^n(m_1), U_2^n(m_2) : m_1 \neq 1, m_2 \neq 1\}$. Then, for n sufficiently large,

$$\begin{aligned}
\mathsf{P}\{U_1^n(1) = u_1^n, U_2^n(1) = u_2^n \,|\, M_1 \neq 1, M_2 \neq 1,&\\
S^n = s^n, \bar{\mathcal{C}} = \bar{c}\}&\\
\le 4\,\mathsf{P}\{U_1^n(1) = u_1^n\}\,\mathsf{P}\{U_2^n(1) = u_2^n\}.&
\end{aligned}$$

Proof: Let $\bar{\mathcal{C}}_1 = \{(U_1^n(m_1) : m_1 \neq 1\}$ and $\bar{\mathcal{C}}_2 = \{(U_2^n(m_2) : m_2 \neq 1\}$. Then, by the Markovity

$$(U_1^n(1), \bar{\mathcal{C}}_1, M_1) \to S_1^n \to S_2^n \to (U_2^n(1), \bar{\mathcal{C}}_2, M_2)$$

and Lemma 1,

$$\begin{aligned}
\mathsf{P}\{U_1^n(1) = u_1^n, U_2^n(1) = u_2^n \,|\, M_1 \neq 1, M_2 \neq 1, S_1^n = s_1^n,&\\
S_2^n = s_2^n, \bar{\mathcal{C}}_1 = \bar{c}_1, \bar{\mathcal{C}}_2 = \bar{c}_2\}&\\
= \mathsf{P}\{U_1^n(1) = u_1^n \,|\, M_1 \neq 1, S_1^n = s_1^n, \bar{\mathcal{C}}_1 = \bar{c}_1\}&\\
\cdot\, \mathsf{P}\{U_2^n(1) = u_2^n \,|\, M_2 \neq 2, S_2^n = s_2^n, \bar{\mathcal{C}}_2 = \bar{c}_2\}&\\
\le (2\,\mathsf{P}\{U_1^n(1) = u_1^n\}) \cdot (2\,\mathsf{P}\{U_2^n(1) = u_2^n\}).&
\end{aligned}$$

■

Lemma 4: For n sufficiently large,

$$\mathsf{P}\{Y^n = y^n \,|\, M_1 \neq 1, M_2 \neq 1, S^n = s^n\} \le 2p(y^n|s^n).$$

Proof: The proof is essentially identical to that of Lemma 2. ■

Therefore, $\mathsf{P}(\mathcal{E}_6)$ tends to zero as $n \to \infty$ if $R_1 + R_2 < I(U_1, U_2; Y) + I(U_1; U_2) + \delta(\epsilon)$. Finally, by eliminating R_1 and R_2, we have shown that $\mathsf{P}(\mathcal{E})$ tends to zero as $n \to \infty$, if

$$\begin{aligned}
I(U_1; S_1) &< I(U_1; Y, U_2) - \delta'(\epsilon),\\
I(U_2; S_2) &< I(U_2; Y, U_1) - \delta'(\epsilon),\\
I(U_1; S_1) + I(U_2; S_2) &< I(U_1, U_2; Y) + I(U_1; U_2) - \delta'(\epsilon).
\end{aligned}$$

V. Concluding Remarks

The great appeal of Shannon's source–channel separation architecture is the universal binary interface that completely decouples source coding and channel coding. The cost of this modular design, however, is suboptimal performance when communicating multiple sources over a multi-user channel. In

$$
\begin{aligned}
&\mathsf{P}\{(U_1^n(\tilde{m}_1), U_2^n(M_2), Y^n) \in \mathcal{T}_\epsilon^{(n)} \text{ for some } \tilde{m}_1 \neq M_1\} \\
&\quad \le \sum_{\tilde{m}_1=1}^{2^{nR_1}} \mathsf{P}\{(U_1^n(\tilde{m}_1), U_2^n(M_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M_1 \neq \tilde{m}_1\} \\
&\quad \le \sum_{\tilde{m}_1=1}^{2^{nR_1}} \sum_{s^n} p(s^n)\, \mathsf{P}\{(U_1^n(\tilde{m}_1), U_2^n(M_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M_1 \neq \tilde{m}_1 \,|\, S^n = s^n\} \\
&\quad = 2^{nR_1} \sum_{s^n} p(s^n)\, \mathsf{P}\{(U_1^n(1), U_2^n(M_2), Y^n) \in \mathcal{T}_\epsilon^{(n)}, M_1 \neq 1 \,|\, S^n = s^n\} \\
&\quad \overset{(a)}{\le} 4 \cdot 2^{nR_1} \sum_{s^n} p(s^n) \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \mathsf{P}\{U_1^n(1) = u_1^n\}\, \mathsf{P}\{U_2^n(M_2) = u_2^n, Y^n = y^n \,|\, S^n = s^n)\} \\
&\quad = 4 \cdot 2^{nR_1} \sum_{(u_1^n, u_2^n, y^n) \in \mathcal{T}_\epsilon^{(n)}} \prod_{i=1}^{n} p_{U_1}(u_{1i})\, \mathsf{P}\{U_2^n(M_2) = u_2^n, Y^n = y^n\} \\
&\quad \le 4 \cdot 2^{n(R_1 - I(U_1; Y, U_2) + \delta(\epsilon))}
\end{aligned} \tag{8}
$$

this paper we have presented a new approach to joint source–channel coding, which "almost" decouples source and channel coding operations yet achieves the best known performance. Matching the semi-modular system architecture, the first-order analysis of the underlying coding scheme is also deceptively simple.

While we have focused on communication of a 2-DMS without common part over a DM-MAC, the proposed architecture can be readily adapted to many joint source–channel coding problems for which separate source coding and channel coding have matching index structures, such as

- communication of a 2-DMS with common part over a DM-MAC (Berger–Tung coding with common part [9], [10] matched to Slepian–Wolf coding for a MAC with common message [11]),
- communication of a 2-DMS over a DM-BC (lossy Gray–Wyner system [12] matched to Marton's coding for a broadcast channel [13]),
- communcation of a bivariate Gaussian source over a Gaussian BC [14], and
- communication of a 2-DMS over a DM-IC (extension of Berger–Tung coding for a 2-by-2 source network matched to Han–Kobayashi coding for an interference channel [15]).

In all these cases, the new architecture, despite its simplicity, performs as well as (and sometimes better than) the existing coding schemes. These findings will be reported elsewhere [16].

Appendix A
Proof of Corollary 1

Fix a pmf $p(q)p(x_1|s_1,q)p(x_2|s_2,q)$ and set $U_j = (X_j, S_j)$ and $\hat{S}_j = S_j$ for $j = 1, 2$. Then,

$$
\begin{gathered}
(S_1, S_2, Q) \to (X_1, X_2) \to Y, \\
X_1 \to (S_1, Q) \to (S_2, X_2), \\
(X_1, S_1) \to (S_2, Q) \to X_2.
\end{gathered}
$$

Now $I(U_1; S_1|Q) = H(S_1)$ and

$$
\begin{aligned}
&I(U_1; Y, U_2|Q) \\
&\quad = I(X_1, S_1; Y, X_2, S_2|Q), \\
&\quad = I(X_1; Y, X_2, S_2|Q) + I(S_1; Y, X_2, S_2|X_1, Q), \\
&\quad = I(X_1; Y|X_2, S_2, Q) + I(X_1; X_2, S_2|Q) \\
&\qquad + I(S_1; X_2, S_2|X_1, Q), \\
&\quad = I(X_1; Y|X_2, S_2, Q) + I(S_1; S_2).
\end{aligned}
$$

Hence, the first inequality in Theorem 1 simplifies to

$$H(S_1|S_2) < I(X_1; Y|X_2, S_2, Q).$$

Similarly, the second inequality in Theorem 1 simplifies to

$$H(S_2|S_1) < I(X_2; Y|X_1, S_1, Q).$$

Finally, since

$$
\begin{aligned}
&I(U_1, U_2; Y|Q) + I(U_1; U_2|Q) \\
&\quad = I(X_1, X_2, S_1, S_2; Y|Q) + I(X_1, S_1; X_2, S_2|Q) \\
&\quad = I(X_1, X_2; Y|Q) + I(S_1; S_2)
\end{aligned}
$$

the last inequality of Theorem 1 simplifies to

$$H(S_1, S_2) < I(X_1, X_2; Y).$$

This shows that the distortion pair $(0, 0)$ is achievable for Hamming distortion measures d_1 and d_2. By properties of typical sequences [3, Lecture Notes 2 and 3], this implies that $\mathsf{P}\{(\hat{S}_1^n, \hat{S}_2^n) \neq (S_1^n, S_2^n)\}$ tends to zero as $n \to \infty$, establishing achievability for lossless communication under the condition in Corollary 1.

Appendix B
Proof of Corollary 2

Fix a pmf $p(q)p(x_1)p(x_2)p(v_1|s_1,q)p(v_2|s_2,q)$ in Corollary 2, where $X_j \sim \text{Unif}(|\mathcal{X}_j|)$, $j = 1, 2$. By setting

$U_1 = (X_1, V_1)$ and $U_2 = (X_2, V_2)$, the first inequality in Theorem 1 simplifies to

$$\begin{aligned}
0 &< I(U_1; Y, U_2|Q) - I(U_1; S_1|Q) \\
&= I(U_1; Y|Q) + I(U_1; U_2|Y, Q) - I(U_1; S_1|Q) \\
&= I(X_1, V_1; Y_1, Y_2|Q) + I(X_1, V_1; X_2, V_2|X_1, X_2, Q) \\
&\quad - I(V_1, X_1; S_1|Q) \\
&= I(X_1; Y_1) + I(V_1; V_2|Q) - I(V_1; S_1|Q) \\
&\stackrel{(a)}{=} I(X_1; Y_1) + I(V_1; V_2|Q) - I(V_1; S_1, V_2|Q) \\
&= R_1 - I(V_1; S_1|V_2, Q)
\end{aligned}$$

where (a) follows since $V_1 \to (S_1, Q) \to V_2$. Similarly, the second inequality in Theorem 1 simplifies to

$$0 < R_2 - I(V_2; S_2|V_1, Q).$$

Finally, the last inequality in Theorem 1 simplifies to

$$\begin{aligned}
0 &< I(U_1, U_2; Y|Q) + I(U_1; U_2|Q) - I(U_1; S_1|Q) \\
&\quad - I(U_2; S_2|Q) \\
&\stackrel{(a)}{=} I(U_1, U_2; Y|Q) + I(U_1; U_2|Q) - I(U_1; S_1, U_2|Q) \\
&\quad - I(U_2; S_2|Q) \\
&= I(U_1, U_2; Y|Q) - I(U_1; S_1|U_2, Q) - I(U_2; S_2|Q) \\
&\stackrel{(b)}{=} I(U_1, U_2; Y|Q) - I(U_1; S_1, S_2|U_2, Q) - I(U_2; S_1, S_2|Q) \\
&= I(U_1, U_2; Y|Q) - I(U_1, U_2; S_1, S_2|Q) \\
&= I(X_1, X_2, V_1, V_2; Y|Q) - I(X_1, X_2, V_1, V_2; S_1, S_2|Q) \\
&= I(X_1; Y_1) + I(X_2; Y_2) - I(V_1, V_2; S_1, S_2|Q), \\
&= R_1 + R_2 - I(V_1, V_2; S_1, S_2|Q),
\end{aligned}$$

where (a) follows since $U_1 \to (S_1, Q) \to S_2$ and (b) follows since $U_2 \to (S_2, Q) \to S_1$.

REFERENCES

[1] 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.

[2] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, 1975.

[3] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," 2010. [Online]. Available: http://arxiv.org/abs/1001.3404

[4] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

[5] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.

[6] S.-Y. Tung, "Multiterminal source coding," Ph.D. Thesis, Cornell University, Ithaca, NY, 1978.

[7] A. Lapidoth and S. Tinguely, "Sending a bivariate Gaussian over a Gaussian MAC," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2714–2752, 2010.

[8] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Int. Conv. Rec., part 4*, 1959, vol. 7, pp. 142–163, reprinted with changes in *Information and Decision Processes,* R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93-126.

[9] A. H. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 828–840, 1982.

[10] A. B. Wagner, B. G. Kelly, and Y. Altuğ, "Distributed rate–distortion with common components," 2009, submitted to *IEEE Trans. Inf. Theory*.

[11] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.

[12] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell System Tech. J.*, vol. 53, pp. 1681–1721, 1974.

[13] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, 1979.

[14] C. Tian, S. N. Diggavi, and S. Shamai, "The achievable distortion region bivariate Gaussian source on Gaussian broadcast channel," in *Proc. IEEE International Symposium on Information Theory*, Austin, TX, June 2010, pp. 146–150.

[15] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, 1981.

[16] S. H. Lim, P. Minero, and Y.-H. Kim, "A new approach to joint source–channel coding," 2010, in preparation.