Information and Inference: A Journal of the IMA (2022) **00**, 1–59
https://doi.org/10.1093/imaiai/iaab021

Conditional and Relevant Common Information

Robert Graczyk† and Amos Lapidoth
ETH Zurich
†Corresponding author: graczyk@isi.ee.ethz.ch

and

Michèle Wigger
LTCI, Telecom Paris, IP Paris

[Received on 14 September 2020; revised on 13 April 2021; accepted on 3 August 2021]

Two variations on Wyner's common information are proposed: *conditional common information* and *relevant common information*. These are shown to have operational meanings analogous to those of Wyner's common information in appropriately defined distributed problems of compression, simulation and channel synthesis. For relevant common information, an additional operational meaning is identified: on a multiple-access channel with private and common messages, it is the minimal common-message rate that enables communication at the maximum sum-rate under a weak coordination constraint on the inputs and output. En route, the weak-coordination problem over a Gray-Wyner network is solved under the no-excess-rate constraint.

Keywords: Common Information; Channel Synthesis; Distributed Source Coding; Coordination; Simulation.

1. Introduction

Inspired by Wyner's *common information*, which he introduced to quantify the information that is shared by two chance variables [33], we propose two notions of shared information: *conditional common information* and *relevant common information*.[1] The former can be viewed as a conditional version of Wyner's common information, whereas the latter measures the amount of information that—in addition to being shared by two chance variables—is also relevant to a third. In the simplest setting where the chance variables T_1 and T_2 are tuples of the form

$$\begin{aligned} T_1 &= (X_1, Y, A) \\ T_2 &= (X_2, Y, A), \end{aligned}$$

where X_1, X_2, Y, and (A, S) are independent, Wyner's common information $\mathrm{C}(T_1; T_2)$ between T_1 and T_2 is $\mathrm{H}(Y) + \mathrm{H}(A)$ (where $\mathrm{H}(\cdot)$ denotes entropy); the conditional common information $\mathrm{C}(T_1; T_2 \mid Y)$ between T_1 and T_2 given Y is $\mathrm{H}(A)$; and the relevant common information $\mathrm{C}(T_1; T_2 \to S)$ between T_1 and T_2 of relevance to S is $\mathrm{I}(A; S)$ (where $I(\cdot;\cdot)$ denotes mutual information).

[1] These notions were first defined in [16], which contains a subset of the present results and proofs.

The definitions of the different common informations apply, of course, to general chance variables that are not necessarily tuples of this form. Indeed, Wyner [33] defined the common information $\mathrm{C}(T_1;T_2)$ between two discrete chance variables T_1 and T_2 of joint probability mass function (PMF) $Q_{T_1T_2}$ as

$$\mathrm{C}(T_1;T_2) \triangleq \min_{W\colon T_1 \to W \to T_2} \mathrm{I}(T_1,T_2;W), \tag{1.1}$$

where the minimization is over all auxiliary chance variables W satisfying $T_1 \to W \to T_2$, i.e., conditionally on which T_1 and T_2 are independent. (Throughout this paper we write $X \to Y \to Z$ to indicate that X and Z are conditionally independent given Y.) When the alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ in which T_1 and T_2 take values are finite, W can be restricted to take values in a finite set of cardinality $|\mathcal{T}_1||\mathcal{T}_2|$ [33]. Strictly speaking, $\mathrm{C}(T_1;T_2)$ is not a function of the chance variables but of their joint distribution. Nevertheless, following common practice in Information Theory, it is denoted $\mathrm{C}(T_1;T_2)$ as though it were.

Now known as *Wyner's common information*, $\mathrm{C}(T_1;T_2)$ was shown by Wyner to have two operational meanings. The first is related to a source-encoding network—the *Gray-Wyner network*—which was studied by Gray and Wyner [12] and which is similar to the one depicted in Fig. 1 but without Y^n. In this network an encoder is presented with an n-length sequence of tuples $\{(T_{1,i},T_{2,i})\}$ that are independent and identically distributed (IID) according to some given joint distribution $Q_{T_1T_2}$. The encoder produces three descriptions of the sequence: a rate-R_1 description, which is provided to Decoder 1 whose task is to reproduce T_1^n; a rate-R_2 description, which is presented to Decoder 2 whose task is to reproduce T_2^n; and a rate-R_0 description, which is presented to both. (We use A^n to denote the n-length sequence $A_1,\ldots,A_n$.) The common information $\mathrm{C}(T_1;T_2)$ indicates the smallest common rate R_0 that is required to achieve (almost) lossless compression by both decoders under the no-excess-rate condition that the sum $R_0+R_1+R_2$ be at its minimum, i.e., at $\mathrm{H}(T_1,T_2)$.

The second operational meaning Wyner provided is related to the simulation of n-length sequences T_1^n and T_2^n in a setting similar to the one in Fig. 2 but without Y^n. Here the common randomness J is used in order to ensure that the joint distribution of $\{(T_{1,i},T_{2,i})\}_{i=1}^n$ resembles $Q_{T_1T_2}^{\otimes n}$, where the latter denotes the n-fold product of $Q_{T_1T_2}$. (Wyner used the normalized Kullback-Leibler (KL) divergence, a.k.a. *relative entropy*, to measure the resemblance, but similar results hold under Total Variation [7, 13, 35] or Rényi divergence [37, 38].)

The conditional common information $\mathrm{C}(T_1;T_2 \mid Y)$ that we define in Definition 1 ahead extends Wyner's by accounting for the side-information sequence Y^n in Figs 1 and 2. For the relevant common information the corresponding figures are Figs 4 and 6. They correspond to source-driven weak coordination and to remote simulation over a multiple-access channel (MAC).

Over the years, additional operational meanings for Wyner's common information were presented. Cuff [7] considered a distributed channel synthesis network similar to the one depicted in Fig. 3 but without Y^n. Here we are presented with a sequence $T_1^n \sim Q_{T_1}^{\otimes n}$, and we wish to simulate the result of feeding it to a discrete memoryless channel (DMC) whose law is the conditional distribution of T_2 given T_1. Aiding us in this task is the equiprobably drawn rate-R_K common randomness K. The common randomness and the sequence T_1^n are mapped to a codeword in a communication codebook of rate R. Based on this codeword and the common randomness, a sequence T_2^n is generated, and it is required that the distribution of the sequence $\{(T_{1,i},T_{2,i})\}$ resemble $Q_{T_1T_2}^{\otimes n}$. In this setting $\mathrm{C}(T_1;T_2)$ is the minimum of the sum R_k+R that makes this possible. A similar result holds for the conditional common information in the presence of Y^n (Corollary 2.3 ahead).

Wyner's common information has found applications, inter alia, in problems pertaining to distributed secret key generation [28] and to the nonnegative rank of matrices [3], [2]. Other operational meanings to Wyner's common information, related to caching problems, were presented in [18, 19, 26, 30]. For example, [18, 19, 26] consider a two-phase caching scenario with a single transmitter observing IID tuples $\{(T_{1,i}, T_{2,i})\}$ and a single receiver wishing to learn either the sequence T_1^n or T_2^n (but not both). Prior to being revealed which, the transmitter uses the first phase, the *placement phase*, to map (prefetch) $\{(T_{1,i}, T_{2,i})\}$ to a rate-C message, which is placed in the receiver's cache memory. In the second phase, the *delivery phase*, the receiver reveals to the transmitter which of the two sequences it seeks. The transmitter—knowing the message it placed in the receiver's cache and now also which sequence the receiver seeks—completes the delivery phase by sending the receiver a message that enables the receiver to losslessly reconstruct the desired sequence. This message is of rate R_1, if the desired sequence is T_1^n, and of rate R_2 if it is T_2^n. Success must be guaranteed irrespective of which of the two sequences the receiver desires. The common information $\mathrm{C}(T_1; T_2)$ is the smallest 'cache capacity' C for which success can be guaranteed with delivery-phase rates R_1 and R_2 satisfying $R_1 + R_2 + C = \mathrm{H}(T_1, T_2)$. (The rate-sum $R_1 + R_2 + C$ must be at least $\mathrm{H}(T_1, T_2)$ because, with the aid of the rate-C cache message in the placement phase and of the two possible rate R_1 and R_2 messages in the delivery-phase, the receiver can reconstruct *both* T_1^n and T_2^n.)

1.1 *Other Extensions of Wyner's Common Information*

Wyner's common information was extended in a number of directions. Liu et al. [17] proposed an extension that measures the information that is common to more than two, say N, chance variables and that maintains Wyner's operational meanings. This extension also maintains the channel synthesis meaning (for an $(N-1)$-receivers broadcast channel) [7] and the caching meaning (for an N-files single-user caching system) [26].

A different direction was followed by Sula and Gastpar [23, 24] who defined *relaxed common information*. It is parameterized by $\gamma \geq 0$ and is defined as

$$\mathrm{C}_\gamma(T_1; T_2) \triangleq \min_{W:\, \mathrm{I}(T_1;T_2|W)\leq\gamma} \mathrm{I}(T_1; T_2). \tag{1.2}$$

When γ is zero, the constraint in the minimization reduces to the constraint $T_1 \to W \to T_2$, and $\mathrm{C}_0(T_1; T_2)$ reduces to Wyner's common information $\mathrm{C}(T_1; T_2)$.

A lossy version of Wyner's common information, the *lossy common information*, was introduced independently in [29] and [36]. Given a pair of distortion functions $d_1(\cdot,\cdot), d_2(\cdot,\cdot)$ and maximum allowed expected distortions D_1, D_2, it is defined as

$$\mathrm{C}_{D_1,D_2}(T_1; T_2) \triangleq \min_{\substack{W,\hat{T}_1,\hat{T}_2:\ \hat{T}_1\to W\to\hat{T}_2 \\ W\to(\hat{T}_1,\hat{T}_2)\to(T_1,T_2) \\ \mathbb{E}[d_1(T_1,\hat{T}_1)]\leq D_1 \\ \mathbb{E}[d_2(T_2,\hat{T}_2)]\leq D_2}} \mathrm{I}(\hat{T}_1, \hat{T}_2; W). \tag{1.3}$$

It reduces to Wyner's common information when the distortion functions are Hamming distortions and $D_1 = D_2 = 0$. It too is related to Gray-Wyner networks: it is the smallest common rate R_0 required in a Gray-Wyner lossy source coding problem when the two decoders have to reconstruct the two source components to within distortions D_1 and D_2 under the no-excess-rate condition that the sum-rate $R_0 + R_1 + R_2$ is at its minimum, i.e., coincides with the joint rate-distortion function for the two sources

[29, 36]. It has an operational meaning similar to Wyner's common information in single-user caching systems where the user is content with a lossy version of the file it seeks [26]. A relaxed version of lossy common information, *relaxed lossy common information*, was proposed in [23].

The Gray-Wyner source-coding network, which motivated Wyner's definition of common information also serves as the motivation for the recently defined *Rényi common information* [11]. The key is to replace the almost-lossless recovery criterion with the requirement that the ρ-th moment of the number of guesses needed by the decoders to guess the source sequence be exponentially small.

Other notions of common information have been proposed and used in the past. A measure of a more combinatorial nature than Wyner's is the *Gács-Körner common information* $\mathrm{K}(T_1;T_2)$ [10], which characterizes the largest normalized entropy of the random variables that can be agreed upon by terminals that observe T_1^n and T_2^n, respectively, when $\{(T_{1,i},T_{i,2})\} \sim Q_{T_1T_2}^{\otimes n}$. This quantity—which is zero unless $T_1 = (X_1,A)$ and $T_2 = (X_2,A)$ with $\mathrm{H}(A)$ positive [10], [31]—never exceeds Wyner's common information, and

$$\mathrm{K}(T_1;T_2) \le \mathrm{I}(T_1;T_2) \le \mathrm{C}(T_1;T_2). \tag{1.4}$$

1.2 *Organization and Sneak Preview*

The conditional common information $\mathrm{C}(T_1;T_2 \mid Y)$ is defined in Section 2. After studying some of its basic properties, we provide three operational meanings for it in Sections 2.1 through 2.3:

1. In the Gray-Wyner source-coding network with side information of Fig. 1, $\mathrm{C}(T_1;T_2 \mid Y)$ is the smallest common rate R_0 that allows the two decoders to reproduce the individual source sequences (almost) losslessly when the encoder and both decoders observe the side information (SI) sequence Y^n, and $R_0+R_1+R_2$ must not exceed $\mathrm{H}(T_1,T_2 \mid Y)$ (Corollary 2.3).
2. In the simulation problem with side information of Fig. 2, $\mathrm{C}(T_1;T_2 \mid Y)$ is the smallest randomness rate allowing the two simulators to produce sequences T_1^n, T_2^n that, together with Y^n, have a joint distribution that closely resembles $Q_{T_1T_2Y}^{\otimes n}$ (Theorem 2.7).
3. In the distributed channel synthesis problem with side information of Fig. 3, where $(T_1^n, Y^n) \sim Q_{T_1Y}^{\otimes n}$, it corresponds to the smallest sum $R_K + R$ of the common randomness rate R_K and the communication rate R that allows the decoder to produce a sequence T_2^n that, together with (T_1^n, Y^n), has a joint distribution that closely resembles $Q_{T_1T_2Y}^{\otimes n}$ (Corollary 2.3).

The relevant common information $\mathrm{C}(T_1;T_2 \to S)$ is defined in Section 3. After studying some of its basic properties, we provide the following operational meanings in Sections 3.1 through 3.3. Section 3.4 addresses a problem (depicted in Fig. 8) to which the relevant common information is often the answer, but not always.

1. In the Gray-Wyner network of Fig. 4 with $S^n \sim Q_S^{\otimes n}$, the quantity $\mathrm{C}(T_1;T_2 \to S)$ is the minimal common rate R_0 that allows encoders of no excess-rate—i.e., of rates satisfying the condition that $R_0+R_1+R_2$ equals $\mathrm{I}(T_1,T_2;S)$ (with the latter computed w.r.t. $Q_{T_1T_2S}$)—to produce sequences T_1^n and T_2^n that are weakly coordinated with S^n in the sense that their joint empirical distribution with S^n approaches $Q_{T_1T_2S}$ in probability as $n \to \infty$ (Corollary 2.2).[2]

[2] We often use the adjective 'weakly' to indicate that the requirement is related to the empirical distribution of sequences. We use 'strongly' when the requirement is that the distribution of n-length sequences be close to some product distribution.

2. On the discrete memoryless multiple-access channel (MAC) of inputs T_1, T_2 and output S depicted in Fig. 5, $\mathrm{C}(T_1; T_2 \to S)$ is the smallest common rate required to reliably transmit common and private messages, when the joint empirical distribution of the inputs and output must be approximately $Q_{T_1 T_2 S}$, where the conditional law of S given (T_1, T_2) under the latter is the channel law (Corollary 3.2).

3. In the network of Fig. 6, where the input sequences T_1^n and T_2^n to the MAC must result in its output sequence S^n being approximately $Q_S^{\otimes n}$ distributed, the least required rate of common randomness is the minimum of $\mathrm{C}(T_1; T_2 \to S)$ over all joint PMFs whose S-marginal is Q_S and under which the conditional distribution of S given (T_1, T_2) coincides with the MAC's law (Theorem 3.9).

The theorem behind the first operational meaning of relevant common information (Item 1. above) solves the Gray-Wyner weak coordination problem under the no-excess-rate condition. It generalizes Ahlswede's result on the rate-distortion region for multiple descriptions without excess rate [1], and Ahlswede's techniques are used heavily in the converse part of its proof in Section 4. Many of the other proofs are provided in appendices.

1.3 *Notation and Conventions*

Unless otherwise specified, all the sets in this paper are finite, and all the chance variables take values in finite sets. Chance variables are typically denoted using upper-case letters such as X, and their realizations using lower-case letters such as x. Sets are typically denoted using the calligraphic font as in $\mathcal{X}$, and the random variable X usually takes value in the set $\mathcal{X}$. The cardinality of the set $\mathcal{X}$ is denoted $|\mathcal{X}|$. The family of PMFs on the set $\mathcal{X}$ is denoted $\mathcal{P}(X)$. We write $X \sim P$ to indicate that X is distributed according to $P \in \mathcal{P}(X)$. In this vein, $X \sim \mathrm{Unif}(\mathcal{X})$ indicates that X is equiprobably distributed over $\mathcal{X}$, and $X \sim \mathrm{Ber}(p)$ indicates that X has a Bernoulli-p distribution, i.e., takes on the values 1 and 0 with probabilities p and $1-p$. If X and Y are independent, we write $X \perp\!\!\!\perp Y$. Expectations are denoted $\mathbb{E}[\cdot]$ or $\mathbb{E}_A[\cdot]$, with the latter indicating that the expectation is over the chance variable A.

We use $[1:n]$ to denote the set $\{1, \ldots, n\}$, and $\mathbb{1}\{\cdot\}$ for the indicator function that equals 1 if the argument is true and 0 otherwise. The joint PMF of an n-tuple $(X_1, \ldots, X_n)$ is denoted P_{X^n}, and, for $k \in [1:n]$, we write X^k for $X_1, \ldots, X_k$ and X_k^n for $X_k, \ldots, X_n$. The n-fold product distribution of Q is denoted $Q^{\otimes n}$: if $X_1, \ldots, X_n$ are IID according to $Q \in \mathcal{P}(X)$, then $P_{X^n} = Q^{\otimes n}$.

The entropy of a chance variable X of PMF Q is denoted $\mathrm{H}(X)$, $\mathrm{H}(Q)$ or $\mathrm{H}_Q(X)$. The mutual information between X and Y is denoted $\mathrm{I}(X;Y)$, and the conditional mutual information between X and Y given a third chance variable Z is denoted $\mathrm{I}(X;Y \mid Z)$. All entropies and mutual informations in this paper are in nats and all logarithms natural.

The empirical distribution of a sequence $x^n \in \mathcal{X}^n$ is denoted π_{x^n}. It is a PMF in $\mathcal{P}(\mathcal{X})$ with $\pi_{x^n}(a)$ being the frequency of occurrence of the symbol $a \in \mathcal{X}$ in the sequence x^n. If X^n is a random sequence, then π_{X^n} is a chance variable taking values in $\mathcal{P}(\mathcal{X})$.

1.4 *Total Variation Distance*

To measure the distance between two PMFs $P, Q \in \mathcal{P}(\mathcal{X})$, we use the Total Variation distance

$$\mathsf{d}_{\mathrm{TV}}(P;Q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x) - Q(x)| = \frac{1}{2}\|P-Q\|_1, \tag{1.5}$$

where $\|\cdot\|_1$ denotes the $\mathbb{L}_1$-norm.

Information measures such as entropy, mutual information, and conditional mutual information are continuous with respect to (w.r.t.) the Total Variation metric. Consequently, since conditional independence can be expressed in terms of conditional mutual information, the following holds:

PROPOSITION 1.1 (Preservation of Markovity). Let $\{P^{(n)}_{XYZ}\}$ be a sequence of PMFs on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ under each of which $X \to Y \to Z$. If the sequence converges in Total Variation to P_{XYZ}, then $X \to Y \to Z$ must also form a Markov chain under P_{XYZ}.

The Triangle inequality for the $\mathbb{L}_1$-norm implies that the distance between two PMFs upper-bounds the distance between the corresponding marginals:

PROPOSITION 1.2 (Total Variation Distance between Marginals). Let P_{XY} and Q_{XY} be two joint distributions on $\mathcal{X} \times \mathcal{Y}$ of X-marginals P_X and Q_X. Then,

$$\mathsf{d}_{\mathrm{TV}}\left(P_X; Q_X\right) \leq \mathsf{d}_{\mathrm{TV}}\left(P_{XY}; Q_{XY}\right). \tag{1.6}$$

COROLLARY 1.1 (Convergence of the Marginals). If $\{P^{(n)}_{XY}\} \subset \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ converges in Total Variation to P_{XY}, then the X-marginals of $P^{(n)}_{XY}$ converge in Total Variation to the X-marginal of P_{XY}.

The following result on Total Variation and DMCs follows directly from (1.5):

PROPOSITION 1.3 (Total Variation Distance and DMCs). Let P_{XY} have the form $P_X(x)\, w(y \mid x)$, where P_X is the X-marginal of P_{XY} and $w(y \mid x)$ is a channel law. Likewise, let Q_{XY} have the form $Q_X(x)\, w(y \mid x)$. Then,

$$\mathsf{d}_{\mathrm{TV}}\left(P_{XY}; Q_{XY}\right) = \mathsf{d}_{\mathrm{TV}}\left(P_X; Q_X\right). \tag{1.7}$$

COROLLARY 1.2 (Converging Sequence of Joint Input-Output PMFs). If each of the elements of a sequence $\{P^{(n)}_{XY}\} \subset \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ converging to P_{XY} has the form $P^{(n)}_X(x)\, w(y \mid x)$, then so does the limit: $P_{XY}(x, y) = P_X(x)\, w(y \mid x)$, where P_X is the X-marginal of P_{XY}.

REMARK 1.1 Proposition 1.2 and Proposition 1.3 imply a Data Processing inequality for Total Variation: the Total Variation between two input distributions to a channel upper-bounds the distance between the corresponding output distributions.

The following bounds on the Total Variation distance follow from its coupling characterization.

PROPOSITION 1.4 (Total Variation Distance between Product PMFs). The Total Variation distance between two product measures is upper-bounded by the sum of the Total Variation distances between their components

$$\mathsf{d}_{\mathrm{TV}}\left(P_1 \times \cdots \times P_m; Q_1 \times \cdots \times Q_m\right) \leq \sum_{k=1}^{m} \mathsf{d}_{\mathrm{TV}}\left(P_k; Q_k\right). \tag{1.8}$$

PROPOSITION 1.5 (Total Variation Distance and Random Indices). Let X^n and Y^n have PMFs P_{X^n} and P_{Y^n}, and let U take values in $[1:n]$ independently of (X^n, Y^n). Let P_{X_U} and P_{Y_U} be the PMFs of X_U and

Y_U. Then,

$$\mathsf{d}_{\mathrm{TV}}\left(P_{X_U}; P_{Y_U}\right) \leq \mathsf{d}_{\mathrm{TV}}\left(P_{X^n}; P_{Y^n}\right). \tag{1.9}$$

2. Conditional Common Information

DEFINITION 2.1 (Conditional Common Information). Given a triple of chance variables (T_1, T_2, Y) of some joint PMF $P_{T_1T_2Y} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y})$, the *conditional common information* between T_1 and T_2 *given* Y is

$$\mathrm{C}(T_1; T_2 \mid Y) \triangleq \min_{W\colon T_1 \to (W,Y) \to T_2} \mathrm{I}(T_1, T_2; W \mid Y), \tag{2.1}$$

where the minimization is over all finite sets $\mathcal{W}$, all joint PMFs $P_{T_1T_2YW} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y} \times \mathcal{W})$ whose (T_1, T_2, Y)-marginal is the given $P_{T_1T_2Y}$ and under which $T_1 \to (W, Y) \to T_2$, and where the conditional mutual information is calculated w.r.t. $P_{T_1T_2YW}$.

Denoting the Y-marginal of $P_{T_1T_2Y}$ by P_Y, we can express the minimum as being over all joint PMFs of the form

$$P_Y(y)\, P_{W|Y}(w \mid y)\, P_{T_1|W,Y}(t_1 \mid w, y)\, P_{T_2|W,Y}(t_2 \mid w, y).$$

For each $Y = y$ it thus entails a minimization over $P_{W|Y=y}$, $P_{T_1|W,Y=y}$ and $P_{T_2|W,Y=y}$. This can be used to represent $\mathrm{C}(T_1; T_2 \mid Y)$ as the expectation over Y of $\mathrm{C}(T_1; T_2 \mid Y = y)$:

PROPOSITION 2.2 The conditional common information $\mathrm{C}(T_1; T_2 \mid Y)$ can be expressed as

$$\mathrm{C}(T_1; T_2 \mid Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, \mathrm{C}(T_1; T_2 \mid Y = y), \tag{2.2}$$

where $\mathrm{C}(T_1; T_2 \mid Y = y)$ is Wyner's common information between T_1 and T_2 when their joint distribution is $P_{T_1T_2|Y=y}$.

Proof. By the definition of conditional mutual information, and using $T_1 \to (W, Y = y) \to T_2$ to indicate that T_1 and T_2 are conditionally independent given W and the event $\{Y = y\}$,

$$\begin{aligned}
\mathrm{C}(T_1; T_2 \mid Y) &= \min_{P_{W|Y=y}, P_{T_1|W,Y=y}, P_{T_2|W,Y=y}} \sum_{y \in \mathcal{Y}} P_Y(y) \mathrm{I}(T_1, T_2; W \mid Y = y) && (2.3)\\
&= \sum_{y \in \mathcal{Y}} P_Y(y) \min_{W\colon T_1 \to (W,Y=y) \to T_2} \mathrm{I}(T_1, T_2; W \mid Y = y) && (2.4)\\
&= \sum_{y \in \mathcal{Y}} P_Y(y) \mathrm{C}(T_1; T_2 \mid Y = y). && (2.5)
\end{aligned}$$

□

The auxiliary chance variable in the optimization defining Wyner's common information can be restricted to take values in a set of cardinality $|\mathcal{T}_1||\mathcal{T}_2|$ [33], and the optimization defining the conditional common information can be broken up into $|\mathcal{Y}|$ separate such optimizations (2.2). Hence, COROLLARY 2.1 The auxiliary chance variable W in the definition of the conditional common information $\mathrm{C}(T_1;T_2 \mid Y)$ may be restricted to take values in a set of cardinality $|\mathcal{T}_1||\mathcal{T}_2|$.

Using (2.2) and known properties of Wyner's common information such as (1.4), we obtain:

REMARK 2.1

1. If T_1 and T_2 are conditionally independent given Y, then $\mathrm{C}(T_1;T_2 \mid Y)$ is zero.
2. Conditional common information is no smaller than conditional mutual information:

$$\mathrm{C}(T_1;T_2 \mid Y) \geq \mathrm{I}(T_1;T_2 \mid Y). \tag{2.6}$$

3. If Y is independent of the pair (T_1,T_2), then conditional common information reduces to Wyner's common information:

$$\mathrm{C}(T_1;T_2 \mid Y) = \mathrm{C}(T_1;T_2), \quad Y \perp\!\!\!\perp (T_1,T_2). \tag{2.7}$$

4. Conditional common information is continuous in the joint distribution $P_{T_1T_2Y}$ w.r.t. the Total Variation topology. (*c.f.* [32, Theorem 1 (v)]).

EXAMPLE 2.3 ($\mathrm{C}(T_1;T_2)$ can exceed $\mathrm{C}(T_1;T_2 \mid Y)$. Suppose $T_1 = (A_1,Y)$ and $T_2 = (A_2,Y)$, with the tuple (A_1,A_2) being independent of Y. Using (2.2), we obtain that

$$\mathrm{C}(T_1;T_2 \mid Y) = \mathrm{C}(A_1;A_2) \tag{2.8}$$

but, as we next argue,

$$\mathrm{C}(T_1;T_2) = \mathrm{H}(Y) + \mathrm{C}(A_1;A_2). \tag{2.9}$$

Indeed, since Y is a component of both T_1 and T_2, the Markov condition $T_1 \to W \to T_2$ implies that Y is conditionally deterministic given W. Consequently, whenever $T_1 \to W \to T_2$

$$\begin{aligned}
\mathrm{I}(T_1,T_2;W) &= \mathrm{I}(T_1,T_2;W,Y) && (2.10)\\
&= \mathrm{I}(T_1,T_2;Y) + \mathrm{I}(T_1,T_2;W \mid Y) && (2.11)\\
&= \mathrm{H}(Y) + \mathrm{H}(A_1,A_2) - \mathrm{H}(A_1,A_2 \mid W,Y) && (2.12)\\
&= \mathrm{H}(Y) + \mathrm{I}(A_1,A_2;W) && (2.13)\\
&\geq \mathrm{H}(Y) + \mathrm{C}(A_1;A_2), && (2.14)
\end{aligned}$$

where the second equality holds by the chain rule for mutual information; the third by the independence between (A_1,A_2) and Y; the fourth because Y is determined by W; and the last inequality holds because $T_1 \to W \to T_2$ implies $A_1 \to W \to A_2$. Minimizing over the choice of W (subject to the Markov condition) establishes that $\mathrm{C}(T_1;T_2) \geq \mathrm{H}(Y) + \mathrm{C}(A_1;A_2)$. Equality is established by considering $W = (\tilde{W},Y)$ with $\tilde{W}$ achieving $\mathrm{C}(A_1;A_2)$.

EXAMPLE 2.4 ($\mathrm{C}(T_1;T_2 \mid Y)$ can exceed $\mathrm{C}(T_1;T_2)$). Let T_1 and T_2 be IID $\sim \mathrm{Ber}(1/2)$, so

$$\mathrm{C}(T_1;T_2) = 0, \tag{2.15}$$

and let $Y = T_1 \oplus T_2$ be their mod-2 sum (exclusive or). Conditional on $Y = y$, the random variables T_1 and T_2 determine each other, so $\mathrm{C}(T_1;T_2 \mid Y = y) = \mathrm{H}(T_1 \mid Y = y) = \mathrm{H}(T_1) = \log 2$. Thus,

$$\mathrm{C}(T_1;T_2 \mid Y) = \log 2. \tag{2.16}$$

REMARK 2.2 Our definition of conditional common information (2.1) is reminiscent of that of Braun and Pokutta [3, Definition 3.1], [2] who defined it as

$$\mathrm{C}_{\mathrm{BP}}(T_1;T_2 \mid Y) \triangleq \min_{\substack{W:\, T_1 \to W \to T_2 \\ W \to (T_1,T_2) \to Y}} \mathrm{I}(T_1,T_2;W \mid Y). \tag{2.17}$$

As the following example shows, the two definitions are not equivalent. In fact—unlike $\mathrm{C}(T_1;T_2 \mid Y)$—$\mathrm{C}_{\mathrm{BP}}(T_1;T_2 \mid Y)$ is zero whenever T_1 and T_2 are independent.

EXAMPLE 2.5 ($\mathrm{C}(T_1;T_2 \mid Y)$ and $\mathrm{C}_{\mathrm{BP}}(T_1;T_2 \mid Y)$ may differ). In the setting of Example 2.4 above,

$$\mathrm{C}_{\mathrm{BP}}(T_1;T_2 \mid Y) = 0 \tag{2.18}$$

because $\mathrm{C}_{\mathrm{BP}}(T_1;T_2 \mid Y)$ is always nonnegative and because, since T_1 and T_2 are independent, a deterministic W satisfies the two constraints in (2.17). Comparing (2.16) with (2.18), we conclude that the two notions of conditional common information differ.

In the following subsections we present three different operational meanings of conditional common information. When the SI $\{Y_i\}$ is absent or deterministic, they reduce to the known operational interpretations of common information: the Gray-Wyner source coding and the simulation interpretations presented in Wyner's original paper [33] and the channel synthesis interpretation presented by Cuff [7].

2.1 *Source-Coding Interpretation*

The first interpretation is related to (almost) lossless source coding over the Gray-Wyner network with side information of Fig. 1. Here a sequence of source and SI triples $\{(T_{1,i}, T_{2,i}, Y_i)\}$ is drawn IID according to some given joint PMF $Q_{T_1T_2Y} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y})$.

For a given blocklength n, the encoder $\phi_{\mathrm{SI}}^{(n)}$ observes all three sequences T_1^n, T_2^n, Y^n and produces the index tuple $(J_0, J_1, J_2) \in \mathcal{J}_{0,n} \times \mathcal{J}_{1,n} \times \mathcal{J}_{2,n}$ so

$$(J_0, J_1, J_2) = \phi_{\mathrm{SI}}^{(n)}(T_1^n, T_2^n, Y^n), \tag{2.19}$$

where

$$\phi_{\mathrm{SI}}^{(n)} : \mathcal{T}_1^n \times \mathcal{T}_2^n \times \mathcal{Y}^n \to \mathcal{J}_{0,n} \times \mathcal{J}_{1,n} \times \mathcal{J}_{2,n} \tag{2.20}$$

is the encoding function, and $\mathcal{J}_{0,n}$, $\mathcal{J}_{1,n}$ and $\mathcal{J}_{2,n}$ are the (nonempty) index sets.

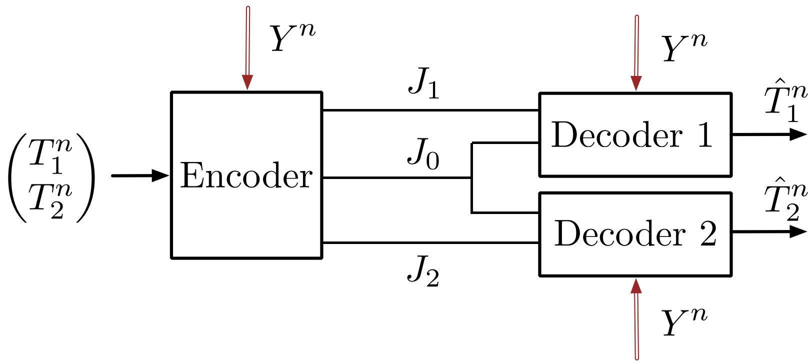

FIG. 1. Lossless Gray-Wyner source coding with side information Y^n.

Indices J_0 and J_1 are fed to Decoder 1 and Indices J_0 and J_2 to Decoder 2. The two decoders also observe the side information Y^n and produce the reconstruction sequences

$$\hat{T}_1^n = \psi_{\mathrm{SI},1}^{(n)}(J_0, J_1, Y^n) \tag{2.21}$$

$$\hat{T}_2^n = \psi_{\mathrm{SI},2}^{(n)}(J_0, J_2, Y^n), \tag{2.22}$$

where $\psi_{\mathrm{SI},1}^{(n)}$ and $\psi_{\mathrm{SI},2}^{(n)}$ are their corresponding decoding functions.

A rate-triple (R_0, R_1, R_2) is said to be achievable on the Gray-Wyner network with SI if, for each blocklength n, there exist index sets $\mathcal{J}_{0,n}$, $\mathcal{J}_{1,n}$ and $\mathcal{J}_{2,n}$; an encoding function $\phi_{\mathrm{SI}}^{(n)}$ as in (2.20); and decoding functions $\psi_{\mathrm{SI},1}^{(n)}$ and $\psi_{\mathrm{SI},2}^{(n)}$ such that:

$$\lim_{n\to\infty} \Pr\big((T_1^n, T_2^n) \neq (\hat{T}_1^n, \hat{T}_2^n)\big) = 0 \tag{2.23}$$

and

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log |\mathcal{J}_{\kappa,n}| \le R_\kappa, \quad \kappa \in \{0, 1, 2\}. \tag{2.24}$$

By the classical (single-user) Source Coding theorem, $\big(\mathrm{H}(T_1, T_2 \mid Y), 0, 0\big)$ is achievable, and every achievable tuple must satisfy

$$R_0 + R_1 + R_2 \ge \mathrm{H}(T_1, T_2 \mid Y).$$

A tuple that is achievable and also satisfies this condition with equality, i.e., for which

$$R_0 + R_1 + R_2 = \mathrm{H}(T_1, T_2 \mid Y), \tag{2.25}$$

is said to be a *no-excess-rate tuple*.

The achievable rate-triples in the absence of SI were characterized in [12] and in its presence in [25, Thm. 1 and Rem. 2):

THEOREM 2.6 (Gray-Wyner Network with Side Information [25]). Given a PMF $Q_{T_1T_2Y}$, a rate-tuple (R_0, R_1, R_2) is achievable on the Gray-Wyner network with SI if and only if there exists an auxiliary chance variable W and a joint PMF $Q_{T_1T_2YW}$ of (T_1T_2Y)-marginal equal to the given $Q_{T_1T_2Y}$ such that

$$R_0 \geq \mathrm{I}(W; T_1, T_2 \mid Y) \tag{2.26a}$$

$$R_1 \geq \mathrm{H}(T_1 \mid W, Y) \tag{2.26b}$$

$$R_2 \geq \mathrm{H}(T_2 \mid W, Y). \tag{2.26c}$$

The following corollary characterizes $\mathrm{C}(T_1; T_2 \mid Y)$ as the minimal common rate R_0 enabling no-excess-rate encoding:

COROLLARY 2.2 A necessary condition for (R_0, R_1, R_2) to be a no-excess-rate tuple is

$$R_0 \geq \mathrm{C}(T_1; T_2 \mid Y). \tag{2.27}$$

Conversely, to each R_0 satisfying (2.27) there correspond private rates R_1, R_2 for which (R_0, R_1, R_2) is a no-excess-rate tuple.

Proof of Corollary. Expressing the mutual information in (2.26a) as $\mathrm{H}(T_1, T_2 \mid Y) - \mathrm{H}(T_1, T_2 \mid Y, W)$ and summing the three inequalities establishes that every achievable rate tuple must satisfy

$$R_0 + R_1 + R_2 \geq \mathrm{H}(T_1, T_2 \mid Y) - \mathrm{H}(T_1, T_2 \mid Y, W) + \mathrm{H}(T_1 \mid W, Y) + \mathrm{H}(T_2 \mid W, Y). \tag{2.28}$$

For a no-excess-rate tuple the left-hand side (LHS) of (2.28) equals $\mathrm{H}(T_1, T_2 \mid Y)$ (see (2.25)), so for such a rate tuple (2.28) implies

$$\mathrm{H}(T_1, T_2 \mid Y, W) \geq \mathrm{H}(T_1 \mid W, Y) + \mathrm{H}(T_2 \mid W, Y).$$

This inequality cannot hold strictly (because the joint entropy never exceeds the sum of the entropies), and it can therefore be replaced with equality. It is thus equivalent to the Markov condition appearing in the minimization defining $\mathrm{C}(T_1; T_2 \mid Y)$ (2.1). The expression being minimized in (2.1) is identical to the right-hand side (RHS) of (2.26a), so (2.27) must hold.

The corollary's second claim follows by choosing W as the auxiliary that achieves $\mathrm{C}(T_1; T_2 \mid Y)$ and setting the rates so that all the inequalities in (2.26) hold with equality. □

2.2 *Simulation Interpretation*

The second interpretation is related to the following strong coordination problem. Consider the network in Fig. 2, where we refer to the sequence $\{Y_i\}$ as side information.

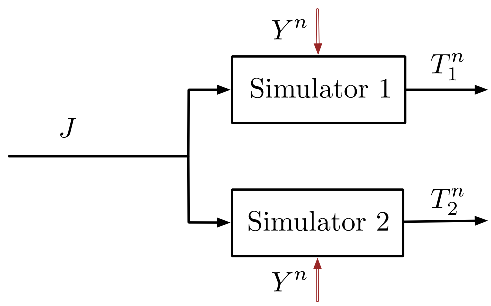

FIG. 2. A simulation problem with side information. We require that $\mathsf{d}_{\mathrm{TV}}\left(P_{T_1^n T_2^n Y^n}; Q_{T_1 T_2 Y}^{\otimes n}\right)$ approach 0.

We say that a joint distribution $Q_{T_1T_2Y} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y})$ can be *strongly coordinated with rate R and SI Y* if, for each blocklength n, there exist a nonempty index set $\mathcal{J}_n$ satisfying

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log |\mathcal{J}_n| \le R \tag{2.29}$$

and *independent random* mappings

$$\Phi_{\mathrm{SI},1}^{(n)} : \mathcal{J}_n \times \mathcal{Y}^n \to \mathcal{T}_1^n \tag{2.30}$$

and

$$\Phi_{\mathrm{SI},2}^{(n)} : \mathcal{J}_n \times \mathcal{Y}^n \to \mathcal{T}_2^n \tag{2.31}$$

such that when $Y^n \sim Q_Y^{\otimes n}$ and $J \sim \mathrm{Unif}(\mathcal{J}_n)$ are independent (and independent of the random mappings $\Phi_{\mathrm{SI},1}^{(n)}$, $\Phi_{\mathrm{SI},2}^{(n)}$) the PMF $P_{T_1^nT_2^nY^n}$ of the sequences T_1^n, T_2^n and Y^n, where the former two are defined by

$$T_1^n = \Phi_{\mathrm{SI},1}^{(n)}(J, Y^n) \tag{2.32}$$

$$T_2^n = \Phi_{\mathrm{SI},2}^{(n)}(J, Y^n), \tag{2.33}$$

is close to the n-fold product distribution $Q_{T_1T_2Y}^{\otimes n}$ in the sense that

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(P_{T_1^nT_2^nY^n}; Q_{T_1T_2Y}^{\otimes n}\right) = 0. \tag{2.34}$$

Note that the Y^n-marginal of both $P_{T_1^n T_2^n Y^n}$ and $Q_{T_1T_2Y}^{\otimes n}$ is $Q_Y^{\otimes n}$, so

$$\begin{aligned}&\mathsf{d}_{\mathrm{TV}}\left(P_{T_1^nT_2^nY^n}; Q_{T_1T_2Y}^{\otimes n}\right)\\&\quad= \sum_{y^n} Q_Y^{\otimes n}(y^n)\, \mathsf{d}_{\mathrm{TV}}\left(P_{T_1^nT_2^n|Y^n=y^n}; Q_{T_1^nT_2^n|Y^n=y^n}^{\otimes n}\right),\end{aligned} \tag{2.35}$$

where $Q_{T_1^nT_2^n|Y^n=y^n}^{\otimes n}$ is the conditional distribution of (T_1^n, T_2^n) given $Y^n = y^n$ under $Q_{T_1T_2Y}^{\otimes n}$,

$$Q_{T_1^nT_2^n|Y^n=y^n}^{\otimes n}(t_1^n, t_2^n) = \prod_{i=1}^{n} Q_{T_1T_2|Y=y_i}(t_{1,i}, t_{2,i}). \tag{2.36}$$

This setup, but without SI, was introduced by Wyner [33], but using the normalized KL-divergence instead of the Total Variation distance in (2.34). Under this KL-divergence constraint, Wyner characterized the set of all PMFs $Q_{T_1T_2}$ that can be strongly coordinated with rate R. From related work [7, 13, 35], it is not difficult to see that Wyner's result continues to hold under the Total Variation distance constraint in (2.34). In fact, in a sense made precise in [7, p. 7076, Eq. (30)], the exponential decay of the normalized KL-divergence is often similar to that of the Total Variation distance.

THEOREM 2.7 The joint PMF $Q_{T_1T_2Y}$ can be strongly coordinated with rate R and SI Y if and only if

$$R \geq \mathrm{C}(T_1; T_2 \mid Y), \tag{2.37}$$

where the RHS is calculated w.r.t. the joint PMF $Q_{T_1T_2Y}$.

Proof. The converse is proved in Appendix A. Here we prove achievability using Wyner's result (under the Total Variation criterion).

Let π_{y^n} denote the empirical distribution of $y^n \in \mathcal{Y}^n$, so $n\,\pi_{y^n}(y)$ is the number of occurrences of $y \in \mathcal{Y}$ in the sequence $y^n \in \mathcal{Y}^n$. Given some $\epsilon > 0$, we say that y^n is typical if $\pi_{y^n}(y)$ is zero whenever $Q_Y(y)$ is zero, and

$$|\pi_{y^n}(y) - Q_Y(y)| < \epsilon, \quad \forall y \in \mathcal{Y}. \tag{2.38}$$

The manner in which the simulations of $(T_{1,i}, T_{2,i})$ are produced depends on whether y^n is typical or not. If not, then Simulator 1 produces its sequence IID $\sim Q_{T_1}$ and Simulator 2 IID $\sim Q_{T_2}$. For such y^n sequences,

$$\mathsf{d}_{\mathrm{TV}}\left(P_{T_1^nT_2^n|Y^n=y^n}; Q_{T_1^nT_2^n|Y^n=y^n}^{\otimes n}\right) \tag{2.39}$$

grows linearly in n, but the probability of their occurrence decays exponentially in n, so their contribution to (2.35) vanishes with n.

We therefore focus on the typical y^n sequences. To address those, we construct a family of Wyner simulators indexed by the SI alphabet $\mathcal{Y}$, with the Wyner simulator indexed by y, 'the y-th Wyner simulator,' designed for the joint distribution $Q_{T_1,T_2|Y=y}$ and required to achieve Total Variation distance smaller than $\epsilon/|\mathcal{Y}|$. The system produces the tuple it reads off from the y-th Wyner simulator whenever the side information Y equals y. This guarantees that the Total Variation distance in (2.39) be smaller than ϵ, because the Total Variation distance between product distributions is upper-bounded by the sum of the Total Variation distances between their respective components (Proposition 1.4).

As the y-th Wyner simulator is used $n\pi_{y^n}(y)$ times, and since the latter is smaller than $n(Q_Y(y)+\epsilon)$, the y-th Wyner simulator can be implemented to produce $n\pi_{y^n}(y)$ tuples with Total Variation distance smaller than $\epsilon/|\mathcal{Y}|$ (for sufficiently large n) with a chance variable J_y that takes on at most $e^{n(Q_Y(y)+\epsilon)(\mathrm{C}(T_1;T_2|Y=y)+\delta)}$ values (where $\delta > 0$ can be arbitrarily small). Using independent such J_y's for the different Wyner simulators, we can perform the overall simulation with a chance variable J that is equiprobably distributed over a set of size

$$\prod_{y\in\mathcal{Y}} e^{n(Q_Y(y)+\epsilon)(\mathrm{C}(T_1;T_2|Y=y)+\delta)} = e^{n(\mathrm{C}(T_1;T_2|Y)+\tilde{\delta}(\epsilon,\delta))},$$

where $\tilde{\delta}(\epsilon,\delta)$ tends to zero as its arguments tend to zero. □

2.3 *Distributed Channel Synthesis Interpretation*

The third interpretation is related to Cuff's *distributed channel synthesis* problem [7]. Consider the network in Fig. 3, where tuples $\{(T_{1,i},Y_i)\}$ of source and SI symbols are drawn IID according to some PMF $Q_{T_1Y} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{Y})$. The goal is for the decoder to produce a sequence $\{T_{2,i}\}$ whose joint PMF $P_{T_1^nT_2^nY^n}$ with $\{(T_{1,i},Y_i)\}$ closely resembles the product distribution $Q_{T_1T_2Y}^{\otimes n}$, where $Q_{T_1T_2Y}$ lies in $\mathcal{P}(\mathcal{T}_1\times\mathcal{T}_2\times\mathcal{Y})$ and is some target PMF having as its (T_1Y)-marginal the PMF Q_{T_1Y} according to which $\{(T_{1,i},Y_i)\}$ are generated.

To achieve this goal, the encoder and decoder share a common randomness K, and the encoder can also convey to the decoder some random index J (that depends on T_1^n and Y^n). The decoder then produces the sequence T_2^n based on K, J and the SI Y^n. For a given blocklength n, the common randomness K is drawn equiprobably from some set $\mathcal{J}_{K,n}$ independently of the source and SI sequences (T_1^n,Y^n), and the index J takes values in some set $\mathcal{J}_n$.

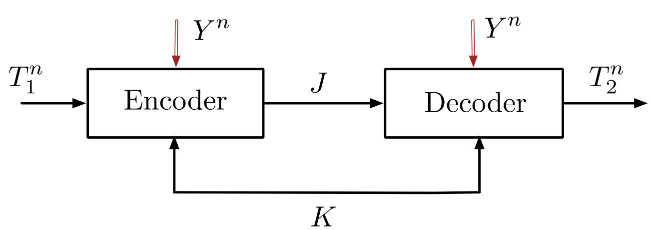

FIG. 3. Distributed channel synthesis with side information. The joint PMF $P_{T_1^nT_2^nY^n}$ of $\{T_{2,i}\}$ with $\{(T_{1,i},Y_i)\}$ should closely resemble $Q_{T_1T_2Y}^{\otimes n}$.

We say that a joint PMF $Q_{T_1T_2Y}$ can be *channel-synthesized with SI Y at communication rate R and common randomness rate* R_K if, for each blocklength n, there exist nonempty sets $\mathcal{J}_n$ and $\mathcal{J}_{K,n}$ satisfying

$$\varlimsup_{n\to\infty} \frac{1}{n}\log|\mathcal{J}_n| \le R \tag{2.40}$$

and

$$\varlimsup_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_{K,n}| \le R_K \tag{2.41}$$

and independent *random* mappings

$$F_{\mathrm{SI}}^{(n)} : \mathcal{T}_1^n \times \mathcal{J}_{K,n} \times \mathcal{Y}^n \to \mathcal{J}_n \tag{2.42}$$

and

$$G_{\mathrm{SI}}^{(n)} : \mathcal{J}_n \times \mathcal{J}_{K,n} \times \mathcal{Y}^n \to \mathcal{T}_2^n \tag{2.43}$$

(that are independent of (T_1^n, Y^n, K)) such that when the tuples $\{(T_{1,i}, Y_i)\}$ are drawn IID $\sim Q_{T_1 Y}$ and the sequence T_2^n is produced as

$$T_2^n = G_{\mathrm{SI}}^{(n)}\big(F_{\mathrm{SI}}^{(n)}\big(T_1^n, K, Y^n\big), K, Y^n\big) \tag{2.44}$$

the resulting joint PMF $P_{T_1^n T_2^n Y^n}$ of (T_1^n, T_2^n, Y^n) satisfies

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(P_{T_1^n T_2^n Y^n}; Q_{T_1 T_2 Y}^{\otimes n}\right) = 0. \tag{2.45}$$

In the absence of SI, the set of PMFs $Q_{T_1 T_2}$ that can be strongly coordinated with rates (R, R_K) was characterized in [7]. The following theorem extends this result to the setup with SI.

THEOREM 2.8 A joint PMF $Q_{T_1 T_2 Y}$ can be channel-synthesized with SI Y at communication rate R and common randomness rate R_K if and only if it is the marginal of some joint PMF $Q_{T_1 T_2 Y W}$ under which

$$T_1 \to (W, Y) \to T_2 \tag{2.46}$$

and

$$R \ge \mathrm{I}(W; T_1 \mid Y) \tag{2.47a}$$

$$R + R_K \ge \mathrm{I}(W; T_1, T_2 \mid Y), \tag{2.47b}$$

where the mutual informations are computed w.r.t. $Q_{T_1 T_2 Y W}$.

Proof. Achievability follows from Cuff's result [7] in much the same way that the achievability part in the proof of Theorem 2.7 followed from Wyner's work. It is therefore omitted. The 'only-if' direction (converse) is proved in Appendix B. □

REMARK 2.3 To exhaust the set of all the rate pairs promised in the theorem, we may restrict W to take values in an alphabet $\mathcal{W}$ of cardinality $|\mathcal{T}_1|\,|\mathcal{T}_2| + 1$, e.g.,

$$\mathcal{W}^* = \big\{1, \ldots, |\mathcal{T}_1|\,|\mathcal{T}_2| + 1\big\}. \tag{2.48}$$

Moreover, said set of rate pairs is closed.

Proof of Remark. We can consider the choice of the auxiliary W separately for each $y \in \mathcal{Y}$. For a fixed $Y = y$, we must choose $Q_{T_1|W,Y=y}$ and $Q_{T_2|W,Y=y}$ subject to the constraints

$$\begin{aligned}\sum_{w\in\mathcal{W}} Q_{W|Y=y}(w)\, Q_{T_1|W=w,Y=y}(t_1)\, Q_{T_2|W=w,Y=y}(t_2) \\ = Q_{T_1T_2|Y=y}(t_1,t_2), \quad (t_1,t_2) \in \mathcal{T}_1 \times \mathcal{T}_2\end{aligned} \tag{2.49}$$

(corresponding to $|\mathcal{T}_1|\,|\mathcal{T}_2| - 1$ constraints, one for all but one pair (t_1, t_2), where one pair can be omitted because the probabilities sum to one). The conditional (on $Y = y$) mutual informations on the RHS of the rate inequalities are determined by $\{Q_{T_1T_2|Y=y}(t_1,t_2)\}$ and

$$\sum_{w\in\mathcal{W}} Q_{W|Y=y}(w)\mathrm{H}(T_1 \mid W = w, Y = y) \tag{2.50}$$

and

$$\sum_{w\in\mathcal{W}} Q_{W|Y=y}(w)\mathrm{H}(T_1, T_2 \mid W = w, Y = y). \tag{2.51}$$

It follows from Carathéodory's theorem (for connected sets) that for each $y \in \mathcal{Y}$ we need at most $|\mathcal{T}_1|\,|\mathcal{T}_2| + 1$ labels for W. Since all three expressions (2.49)–(2.50) do not depend on the labels of W but only on their conditional probabilities, we can choose the same labels under each $y \in \mathcal{Y}$. This establishes the desired cardinality constraint. The second part of the remark follows from the first using a compactness and continuity argument. □

We now focus on the minimum sum-rate $R + R_K$ in the distributed channel synthesis problem.

COROLLARY 2.3 A joint PMF $Q_{T_1T_2Y}$ can be channel-synthesized with SI Y at communication rate R and common randomness rate R_K only if

$$R + R_K \geq \mathrm{C}(T_1; T_2 \mid Y). \tag{2.52}$$

Moreover, there exists a pair (R, R_K) such that (2.52) holds with equality and such that $Q_{T_1T_2Y}$ can be channel-synthesized with SI Y at communication rate R and common randomness rate R_K.

Proof. The necessity of (2.52) follows from the necessity of (2.47b) and from the definition of $\mathrm{C}(T_1; T_2 \mid Y)$ (2.1). The second assertion follows by setting R_K to zero and then using the achievability part of the theorem. □

3. Relevant Common Information

The *relevant common information* $\mathrm{C}(T_1; T_2 \to S)$ quantifies how much of the common information $\mathrm{C}(T_1; T_2)$ is relevant to S. For example, if $T_1 = (X_1, U, T)$ and $T_2 = (X_2, U, T)$ with X_1, X_2, U and (T, S) being independent, then the information that is common to T_1 and T_2 is $\mathrm{H}(U, T)$, but of that only $\mathrm{I}(S; T)$ is relevant to S, so $\mathrm{C}(T_1; T_2) = \mathrm{H}(U, T)$ (cf. Example 2.3) and $\mathrm{C}(T_1; T_2 \to S) = \mathrm{I}(S; T)$.

DEFINITION 3.1 Given a triple of chance variables (S, T_1, T_2) of some joint PMF $P_{ST_1T_2} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2)$, the *common information of the pair* (T_1, T_2) *that is relevant to* S is

$$\mathrm{C}(T_1; T_2 \to S) \triangleq \min_{\substack{W:\, T_1 \to W \to T_2 \\ W \to (T_1, T_2) \to S}} \mathrm{I}(S; W), \tag{3.1}$$

where the minimization is over all finite sets $\mathcal{W}$, all joint PMFs $P_{ST_1T_2W} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{W})$ whose (S, T_1, T_2)-marginal is the given $P_{ST_1T_2}$ and under which both $T_1 \to W \to T_2$ and $W \to (T_1, T_2) \to S$ hold, and where the mutual information $\mathrm{I}(S; W)$ is calculated w.r.t. $P_{ST_1T_2W}$.

REMARK 3.1 The relevant common information has the following basic properties:

1. If $S = (T_1, T_2)$, then the relevant common information reduces to Wyner's common information:

$$\mathrm{C}\big(T_1; T_2 \to (T_1, T_2)\big) = \mathrm{C}(T_1; T_2). \tag{3.2}$$

   (When $S = (T_1, T_2)$, the minimization in (3.1) is identical to the minimization defining Wyner's common information (1.1) except for the extra constraint $W \to (T_1, T_2) \to S$, which—when $S = (T_1, T_2)$—is satisfied irrespective of W.)

2. If T_1 and T_2 are independent, then—irrespective of S— the relevant common information is zero

$$\mathrm{C}(T_1; T_2 \to S) = 0, \quad T_1 \perp\!\!\!\perp T_2 \tag{3.3}$$

   (In this case choosing W to be deterministic satisfies the constraints.)

3. Relevant common information is no larger than Wyner's common information:

$$\mathrm{C}(T_1; T_2 \to S) \leq \mathrm{C}(T_1; T_2). \tag{3.4}$$

   (By the Data Processing inequality, the constraint $W \to (T_1, T_2) \to S$ implies that $\mathrm{I}(S; W) \leq \mathrm{I}(T_1, T_2; W)$. This allows us to upper-bound $\mathrm{C}(T_1; T_2 \to S)$ by a modified expression similar to that for Wyner's common information (1), except for the said constraint. Choosing $P_{W|T_1T_2S}$ equal to $P_{W|T_1T_2}$, where the latter achieves the common information, shows that the extra constraint does not increase the minimum in the modified expression and is, in fact, redundant there.)

4. Relevant common information is no larger than the mutual informations between T_1 or T_2 and S:

$$\mathrm{C}(T_1; T_2 \to S) \leq \min\{\mathrm{I}(T_1; S), \mathrm{I}(T_2; S)\}. \tag{3.5}$$

   (This holds because both $W = T_1$ and $W = T_2$ are admissible choices in the minimization (3.1) defining $\mathrm{C}(T_1; T_2 \to S)$.)

5. In the minimization in (3.1), it suffices to consider auxiliary chance variables W taking values in alphabets of cardinality $|\mathcal{W}| \le |\mathcal{T}_1||\mathcal{T}_2| + 1$. (This holds by Carathéodory's theorem: we have $|\mathcal{T}_1|\,|\mathcal{T}_2| - 1$ constraints on our choice of $P_{T_1|W}$ and $P_{T_2|W}$ analogous to those in (2.49) (without Y); the entropy $\mathrm{H}(S)$ is given; and the Markovity constraint $W \to (T_1, T_2) \to S$ guarantees that $\mathrm{H}(S \mid W)$ can be expressed as an expectation over W of a function of $P_{T_1|W=w}$ and $P_{T_2|w}$.)

6. Relevant common information $\mathrm{C}(T_1; T_2 \to S)$ is continuous in the PMF of the triple (T_1, T_2, S). (The proof of continuity in $P_{T_1T_2}$ for a fixed $P_{S|T_1T_2}$ is very similar to Witsenhausen's proof of the continuity of Wyner's common information [32, Theorem 1(v)]: instead of $H(T_1, T_2 \mid W)$, we maximize $H(S \mid W)$; the term $h_n(\mathbf{p}) + h_m(\mathbf{q})$ in the mapping $(\mathbf{p}, \mathbf{q}) \mapsto \big(\mathbf{p}\mathbf{q}^t, h_n(\mathbf{p}) + h_m(\mathbf{q})\big)$ in [32] is therefore replaced with the entropy $\hbar\big(\mathbf{p}, \mathbf{q}; P_{S|T_1T_2}\big)$ of the distribution on $\mathcal{S}$ that assigns each $s \in \mathcal{S}$ the probability $\sum_{(t_1,t_2)} P_{S|T_1T_2}(s \mid t_1, t_2)\, \mathbf{p}(t_1)\, \mathbf{q}(t_2)$ (with the resulting mapping also being continuous); and the co-domain of the mapping is now $\Delta_{nm} \times [0, \log|\mathcal{S}|]$ instead of $\Delta_{nm} \times [0, \log nm]$. Continuity in $(P_{S|T_1T_2}, P_{T_1T_2})$ is now established by noting that when $P^{(1)}_{S|T_1T_2}$ and $P^{(2)}_{S|T_1T_2}$ are close, $\max_{\mathbf{p},\mathbf{q}}\big|\hbar\big(\mathbf{p}, \mathbf{q}; P^{(1)}_{S|T_1T_2}\big) - \hbar\big(\mathbf{p}, \mathbf{q}; P^{(2)}_{S|T_1T_2}\big)\big|$ is small, e.g., with the help of [4, Theorem 17.3.3)].

7. Relevant common information is related to lossy common information (1.3) in much the same way that weak coordination is related to rate-distortion theory [6]:

$$C_{D_1,D_2}(T_1; T_2) = \min_{\substack{\hat{T}_1,\hat{T}_2:\ \mathbb{E}[d_1(T_1,\hat{T}_1)]\le D_1 \\ \mathbb{E}[d_2(T_2,\hat{T}_2)]\le D_2}} \mathrm{C}(\hat{T}_1; \hat{T}_2 \to (T_1, T_2)). \tag{3.6}$$

Example 3.2 In the setting of Example 2.3, the common information of T_1 and T_2 that is relevant to Y is

$$\mathrm{C}(T_1; T_2 \to Y) = \mathrm{H}(Y). \tag{3.7}$$

Indeed, $\mathrm{C}(T_1; T_2 \to Y) \ge \mathrm{H}(Y)$ because Y must be computable from any auxiliary chance variable W for which $(A_1, Y) \to W \to (A_2, Y)$, and equality holds when W is chosen as (A_1, A_2, Y).

From (2.8), (2.9), and (3.7) we infer that, for the setting of Example 2.3, $\mathrm{C}(T_1; T_2 \mid Y) + \mathrm{C}(T_1; T_2 \to Y)$ equals $\mathrm{C}(T_1; T_2)$. But this does not hold in general. As shown by the following two examples, the LHS can be smaller or larger than the RHS.

Example 3.3 (Example 2.4 Contd.) Since T_1 and T_2 are independent, the common information of T_1 and T_2 relevant to Y is zero (3.3). The conditional common information is $\log(2)$ (2.16), and consequently, in this example,

$$\log 2 = \mathrm{C}(T_1; T_2 \mid Y) + \mathrm{C}(T_1; T_2 \to Y) > \mathrm{C}(T_1; T_2) = 0. \tag{3.8}$$

Example 3.4 Let $Y \sim \mathrm{Ber}(1/2)$, $B_1 \sim \mathrm{Ber}(p)$ and $B_2 \sim \mathrm{Ber}(q)$ be independent Bernoulli random variables, where $p, q \in [0, 1/2]$ and $p \ge q$. Define $T_1 = Y \oplus B_1$ and $T_2 = Y \oplus B_2$. Since (T_1, T_2) is a doubly symmetric binary source whose parameter r equals $p(1-q) + (1-p)q$, Wyner's common information is given by [33, Example on p. 167]

$$\mathrm{C}(T_1; T_2) = \log 2 + H_\mathrm{b}(r) - 2H_\mathrm{b}(r_1), \tag{3.9}$$

where $r_1 = 0.5 - 0.5 \cdot \sqrt{1-2r}$, and $H_\mathrm{b}(\cdot)$ denotes the binary entropy function. Since T_1 and T_2 are conditionally independent given Y, the conditional common information is

$$\mathrm{C}(T_1; T_2 \mid Y) = 0. \tag{3.10}$$

The relevant common information can be upper bounded as (see Item 4 in Remark 3.1)

$$\mathrm{C}(T_1; T_2 \to Y) \le \mathrm{I}(T_1; Y) = \log 2 - H_\mathrm{b}(p). \tag{3.11}$$

Evaluating the bounds in (3.9)–(3.11) for $p = 0.4$ and $q = 0.2$ yields (in nats)

$$\mathrm{C}(T_1; T_2 \mid Y) + \mathrm{C}(T_1; T_2 \to Y) \le 0.020 < 0.115 = \mathrm{C}(T_1; T_2). \tag{3.12}$$

In the following, we present various operational interpretations of relevant common information. The first is presented in Corollary 3.1 ahead and is related to the source-driven weak coordination network depicted in Fig. 4. The second is presented in Corollary 3.2 and is related to combined transmission and weak coordination on a MAC (Fig. 5). The third is related to remote simulation through a MAC (Fig. 6) and is presented in Theorem 3.9.

3.1 *Source-Driven Weak Coordination*

The source-driven weak coordination of a PMF $Q_{ST_1T_2} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2)$ is depicted in Fig. 4. A sequence $\{S_i\}$ is drawn IID according to the marginal distribution Q_S of $Q_{ST_1T_2}$ and is presented to a Gray-Wyner-like encoder.

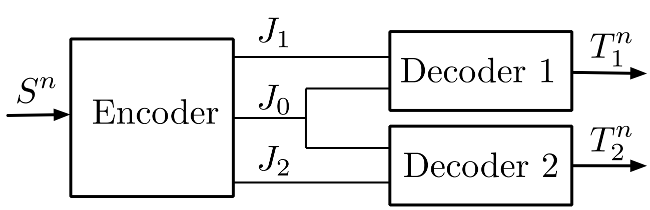

FIG. 4. The source-driven weak-coordination problem. We require that the joint empirical distribution $\pi_{(S^n, T_1^n, T_2^n)}$ converge in probability to $Q_{ST_1T_2}$ (78).

For a given blocklength n, the encoder

$$\phi_{\mathrm{Rel}}{}^{(n)} \colon \mathcal{S}^n \to \mathcal{J}_{0,n} \times \mathcal{J}_{1,n} \times \mathcal{J}_{2,n} \tag{3.13}$$

produces three indices

$$(J_0, J_1, J_2) = \phi_{\mathrm{Rel}}^{(n)}(S^n) \tag{3.14}$$

taking values in the index sets $\mathcal{J}_{0,n}$, $\mathcal{J}_{1,n}$ and $\mathcal{J}_{2,n}$. Indices J_0 and J_1 are presented to Decoder 1 and indices J_0 and J_2 to Decoder 2. The two decoders $\psi^{(n)}_{\mathrm{Rel},1}$ and $\psi^{(n)}_{\mathrm{Rel},2}$ produce the sequences

$$T_1^n = \psi^{(n)}_{\mathrm{Rel},1}(J_0, J_1) \tag{3.15}$$

$$T_2^n = \psi^{(n)}_{\mathrm{Rel},2}(J_0, J_2). \tag{3.16}$$

The joint empirical distribution of (S^n, T_1^n, T_2^n), namely $\pi_{(S^n,T_1^n,T_2^n)}$, takes values in $\mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2)$ and is random because S^n is random. We require that it approach $Q_{ST_1T_2}$ in the sense that

$$\operatorname*{plim}_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(\pi_{(S^n,T_1^n,T_2^n)}; Q_{ST_1T_2}\right) = 0, \tag{3.17}$$

where plim stands for limit in probability.

We say that *the rates (R_0, R_1, R_2) allow for the source-driven weak coordination of $Q_{ST_1T_2}$*, if for every blocklength n, there exist index sets $\mathcal{J}_{0,n}$, $\mathcal{J}_{1,n}$, $\mathcal{J}_{2,n}$ satisfying

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log |\mathcal{J}_{\kappa,n}| \le R_\kappa, \qquad \kappa \in \{0, 1, 2\}; \tag{3.18}$$

an encoding function $\phi^{(n)}_{\mathrm{Rel}}$ as in (3.13); and decoder functions $\psi^{(n)}_{\mathrm{Rel},1}$ and $\psi^{(n)}_{\mathrm{Rel},2}$ such that (3.17) holds.

Similar setups were addressed in [6] and [20]. In [6], however, the encoder only conveys individual indices J_1 and J_2 to the decoders and no common index J_0. In [20] the goal is different: rather than (3.17), the requirement is that the empirical distributions $\pi_{(S^n,T_1^n)}$ and $\pi_{(S^n,T_2^n)}$ approach target PMFs Q_{ST_1} and Q_{ST_2}; no requirement is imposed on the joint empirical distribution $\pi_{(S^n,T_1^n,T_2^n)}$. Like us, [20] only presents a sufficient condition for achievability but no necessary condition. We do, however, provide a complete characterization in the no-excess-rate case (Theorem 3.6 ahead). The work in [6] presents non-matching sufficient and necessary conditions.

The following theorem presents our sufficient conditions for a rate triple (R_0, R_1, R_2) to allow for the source-driven weak coordination of $Q_{ST_1T_2}$.

THEOREM 3.5 The rates (R_0, R_1, R_2) allow for the source-driven weak coordination of $Q_{ST_1T_2}$ whenever there exists a random variable W taking values in a finite set $\mathcal{W}$ and a joint PMF $Q_{WST_1T_2}$ on W, S, T_1, T_2 whose ST_1T_2-marginal is $Q_{ST_1T_2}$ and under which

$$R_0 \ge \mathrm{I}(S; W) \tag{3.19a}$$

$$R_0 + R_1 \ge \mathrm{I}(S; T_1, W) \tag{3.19b}$$

$$R_0 + R_2 \ge \mathrm{I}(S; T_2, W) \tag{3.19c}$$

$$R_0 + R_1 + R_2 \ge \mathrm{I}(S; T_1, T_2, W) + \mathrm{I}(T_1; T_2 \mid W). \tag{3.19d}$$

Proof. Let (S, T_1, T_2, W) be distributed according to the postulated PMF $Q_{WST_1T_2}$. Apply the random coding scheme described in [39, Proof of Theorem 1] with the substitutions

$$X \leftarrow S \qquad X_0 \leftarrow W \qquad X_1 \leftarrow T_1 \qquad X_2 \leftarrow T_2 \tag{3.20}$$

and $\phi_1(a,b) = \phi_2(a,b) = b$. As shown in [39, Eqns. (39)–(48)], the limit (3.17) holds on average over the random choice of the codebooks if the following conditions are satisfied:

$$R_0 \geq I(S; W) \tag{3.21a}$$

$$R_1 \geq I(S; T_1 \mid W) \tag{3.21b}$$

$$R_2 \geq I(S; T_2 \mid W) \tag{3.21c}$$

$$R_1 + R_2 \geq I(S; T_1, T_2 \mid W) + I(T_1; T_2 \mid W). \tag{3.21d}$$

A random-coding argument establishes that Conditions (3.21) gurantee the existence of a sequence of deterministic schemes that attains the weak coordination in (3.17). We next show using a rate-transfer argument [27] that, in fact, Conditions (3.19) suffice.

Key is that the decoders can reproduce the same reconstructions if the encoder splits the private indices J_1 and J_2 into pairs of subindices and—together with the common index J_0—sends one subindex of each pair over the common link. This argument shows that $\left(\tilde{R}_0 + R_1'' + R_2'', R_1', R_2'\right)$ is achievable whenever $\left(\tilde{R}_0, R_1' + R_1'', R_2' + R_2''\right)$ is achievable and hence whenever this latter triple satisfies the sufficient conditions we derived using the random coding argument.

Substituting $R_0' - R_1'' - R_2''$ for $\tilde{R}_0$, we obtain that the nonnegative triple (R_0', R_1', R_2') is achievable whenever there exist $R_1'', R_2'' \geq 0$ such that the triple (R_0, R_1, R_2) given by

$$R_0 = R_0' - R_1'' - R_2'' \tag{3.22a}$$

$$R_1 = R_1' + R_1'' \tag{3.22b}$$

$$R_2 = R_2' + R_2'' \tag{3.22c}$$

is achievable. Using the sufficient condition we obtained via random coding, we conclude that (R_0', R_1', R_2') is achievable whenever there exist $R_1'', R_2'' \geq 0$ such that

$$R_0' - R_1'' - R_2'' \geq I(S; W) \tag{3.23a}$$

$$R_1' + R_1'' \geq I(S; T_1 \mid W) \tag{3.23b}$$

$$R_2' + R_2'' \geq I(S; T_2 \mid W) \tag{3.23c}$$

$$R_1' - R_1'' + R_2' - R_2'' \geq I(S; T_1, T_2 \mid W) + I(T_1; T_2 \mid W). \tag{3.23d}$$

Using Fourier–Motzkin elimination, it can be shown that this condition is equivalent to

$$R'_0 \geq \mathrm{I}(S;W) \tag{3.24a}$$

$$R'_0 + R'_1 \geq \mathrm{I}(S;T_1,W) \tag{3.24b}$$

$$R'_0 + R'_2 \geq \mathrm{I}(S;T_2,W) \tag{3.24c}$$

$$R'_0 + R'_1 + R'_2 \geq \mathrm{I}(S;T_1,T_2,W) + \mathrm{I}(T_1;T_2 \mid W), \tag{3.24d}$$

which, but for the primes, is identical to (3.19). □

The next theorem establishes a converse result under the no-excess-rate condition, i.e., for rate tuples satisfying

$$R_0 + R_1 + R_2 = \mathrm{I}(S;T_1,T_2). \tag{3.25}$$

Notice that $\mathrm{I}(S;T_1,T_2)$ is the smallest rate required to weakly coordinate the reconstruction sequences T_1^n and T_2^n with the source S^n according to a target PMF $Q_{ST_1T_2}$ when a single decoder observes all three indices J_0, J_1, J_2 and produces both T_1^n and T_2^n [6, Thm. 3].

THEOREM 3.6 Consider a PMF $Q_{ST_1T_2} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2)$ and a rate-tuple (R_0, R_1, R_2) satisfying the no-excess-rate condition (3.25) when the RHS of the latter is calculated w.r.t. $Q_{ST_1T_2}$. Said rate tuple allows for the source-driven weak coordination of $Q_{ST_1T_2}$, if and only if there exists some joint PMF on (W, S, T_1, T_2) whose ST_1T_2-marginal is $Q_{ST_1T_2}$ and under which

$$R_0 \geq \mathrm{I}(S;W) \tag{3.26a}$$

$$R_0 + R_1 \geq \mathrm{I}(S;T_1,W) \tag{3.26b}$$

$$R_0 + R_2 \geq \mathrm{I}(S;T_2,W) \tag{3.26c}$$

$$W \to (T_1,T_2) \to S \tag{3.26d}$$

$$T_1 \to W \to T_2. \tag{3.26e}$$

Proof. To prove achievability, we will establish (3.19d) and then invoke Theorem 3.5. To establish (3.19d) we note that the no-excess-rate condition (3.25) implies that its LHS equals $\mathrm{I}(S;T_1,T_2)$, and the Markov conditions (3.26d)–(3.26e) imply that its RHS is also equal to $\mathrm{I}(S;T_1,T_2)$.

The converse is proved in Section 4. □

The following corollary shows that $\mathrm{C}(T_1;T_2 \to S)$ is the smallest common rate that allows achievabity with no excess-rate.

COROLLARY 3.1 Consider a PMF $Q_{ST_1T_2} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2)$. If the rate tuple (R_0, R_1, R_2) satisfies the no-excess-rate condition (3.25) (when the RHS of the latter is calculated w.r.t. $Q_{ST_1T_2}$) and also allows for the source-driven weak coordination of $Q_{ST_1T_2}$, then

$$R_0 \geq \mathrm{C}(T_1; T_2 \to S). \tag{3.27}$$

Moreover, there exists such a rate tuple for which (3.27) holds with equality.

Proof. To establish (3.27), we discard (3.26b)–(3.26c) and optimize over the conditional law of W given (S, T_1, T_2) subject to the Markov conditions (3.26d)–(3.26e).

As to the claim that (3.27) can be achieved with equality, fix some chance variable W and a joint PMF on (W, S, T_1, T_2) that achieves $\mathrm{C}(T_1; T_2 \to S)$, so $\mathrm{I}(S; W)$ equals $\mathrm{C}(T_1; T_2 \to S)$ and the Markov conditions (3.26d)–(3.26e) both hold.

Define R_0 as $\mathrm{I}(S; W)$ and R_1 as $\mathrm{I}(S; T_1, W) - \mathrm{I}(S; W)$, so that $R_0 = \mathrm{C}(T_1; T_2 \to S)$ and both (3.26a) and (3.26b) hold with equality. Define R_2 as $\mathrm{I}(S; T_2, W) - \mathrm{I}(S; W) + \Delta$, so that (3.26c) would hold whenever Δ is positive. Choose Δ so that the no-excess-rate condition (3.25) holds with equality. It remains to establish that, with this choice, Δ is nonnegative or, equivalently, that

$$\mathrm{I}(S; T_1, W) + \mathrm{I}(S; T_2, W) - \mathrm{I}(S; W) \;\leq\; \mathrm{I}(S; T_1, T_2). \tag{3.28}$$

This is, indeed, the case because

$$\begin{aligned}
&\mathrm{I}(S; T_1, W) + \mathrm{I}(S; T_2, W) - \mathrm{I}(S; W) \\
&\quad = \mathrm{I}(S; T_1 \mid W) + \mathrm{I}(S; T_2, W) && (3.29)\\
&\quad = \mathrm{H}(T_1 \mid W) - \mathrm{H}(T_1 \mid W, S) + \mathrm{I}(S; T_2, W) && (3.30)\\
&\quad \leq \mathrm{H}(T_1 \mid W) - \mathrm{H}(T_1 \mid W, S, T_2) + \mathrm{I}(S; T_2, W) && (3.31)\\
&\quad = \mathrm{H}(T_1 \mid W, T_2) - \mathrm{H}(T_1 \mid W, S, T_2) + \mathrm{I}(S; T_2, W) && (3.32)\\
&\quad = \mathrm{I}(S; T_1 \mid W, T_2) + \mathrm{I}(S; T_2, W) && (3.33)\\
&\quad = \mathrm{I}(S; W, T_1, T_2) && (3.34)\\
&\quad = \mathrm{I}(S; T_1, T_2), && (3.35)
\end{aligned}$$

where (3.31) holds because conditioning cannot increase entropy; (3.32) follows from the Markov condition (3.26e); and (3.35) follows from the Markov condition (3.26d). □

3.2 *Combined Transmission and Weak Coordination on a MAC*

The scenario we consider next is the classical two-to-one MAC (with a common message) depicted in Fig. 5, but with the extra twist that we require that the joint empirical distribution of the input and output sequences approximate a given PMF $Q_{ST_1T_2}$. For this to be at all possible, $Q_{ST_1T_2}$ must have the form

$$Q_{ST_1T_2}(s,t_1,t_2) = Q_{T_1T_2}(t_1,t_2)\, p_{\mathrm{c}}(s \mid t_1,t_2), \tag{3.36}$$

where $p_{\mathrm{c}}(s|t_1,t_2)$ is the MAC's law, and $Q_{T_1T_2}$ is some PMF in $\mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2)$. We refer to such a PMF $Q_{ST_1T_2}$ as having conditional law $p_{\mathrm{c}}(s|t_1,t_2)$. Here $\mathcal{T}_1$ and $\mathcal{T}_2$ are the MAC's input alphabets, and $\mathcal{S}$ denotes its output alphabet. The common message is denoted M_0 and the two private messages M_1, M_2. The three are independent and, given a blocklength n, equiprobably distributed over the corresponding message sets $\mathcal{M}_{0,n}$, $\mathcal{M}_{1,n}$, and $\mathcal{M}_{2,n}$. Employing the mapping $\varphi^{(n)}_{\mathrm{Rel},1}$, Encoder 1 maps the pair (M_0, M_1) to the n-tuple of channel inputs

$$T_1^n = \eta^{(n)}_{\mathrm{Rel},1}(M_0, M_1). \tag{3.37}$$

Similarly, Encoder 2 maps (M_0, M_2) to

$$T_2^n = \eta^{(n)}_{\mathrm{Rel},2}(M_0, M_2). \tag{3.38}$$

The decoder observes the MAC's output sequence S^n and, employing the mapping $\zeta^{(n)}_{\mathrm{Rel}}$, produces its guess $(\hat{M}_0, \hat{M}_1, \hat{M}_2) \in \mathcal{M}_{0,n} \times \mathcal{M}_{1,n} \times \mathcal{M}_{2,n}$ of the message triple:

$$(\hat{M}_0, \hat{M}_1, \hat{M}_2) = \zeta^{(n)}_{\mathrm{Rel}}(S^n). \tag{3.39}$$

A MAC $p_{\mathrm{c}}(s|t_1,t_2)$ supports transmission at rates (R_0, R_1, R_2) with weak coordination w.r.t. the PMF $Q_{ST_1T_2}$ of conditional law $p_{\mathrm{c}}(s|t_1,t_2)$ if, for each blocklength n, there exist discrete message sets $\mathcal{M}_{0,n}, \mathcal{M}_{1,n}$ and $\mathcal{M}_{2,n}$; encoding functions $\eta^{(n)}_{\mathrm{Rel},1}$ and $\eta^{(n)}_{\mathrm{Rel},2}$; and a decoding function $\zeta^{(n)}_{\mathrm{Rel}}$ guaranteeing that the following three requirements (3.40)–(3.43) are satisfied:

$$\varliminf_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_{\kappa,n}| \geq R_\kappa, \quad \kappa \in \{0,1,2\}; \tag{3.40}$$

the input and output sequences are weakly coordinated w.r.t. $Q_{ST_1T_2}$

$$\operatorname*{plim}_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(\pi_{(S^n,T_1^n,T_2^n)}; Q_{ST_1T_2} \right) = 0, \tag{3.41}$$

i.e.,

$$\operatorname*{plim}_{n\to\infty} \pi_{(S^n,T_1^n,T_2^n)}(s,t_1,t_2) = Q_{ST_1T_2}(s,t_1,t_2), \quad \forall (s,t_1,t_2) \in \mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2; \tag{3.42}$$

and the decoding error vanishes with the blocklength

$$\lim_{n\to\infty} \Pr\big[(M_0,M_1,M_2) \neq (\hat{M}_0,\hat{M}_1,\hat{M}_2)\big] = 0. \tag{3.43}$$

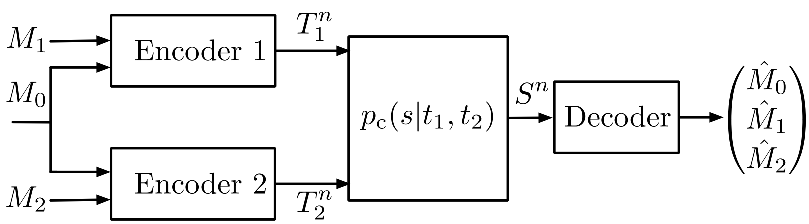

FIG. 5. Combined transmission and weak coordination on a MAC. In addition to reliable communication, we require that the MAC's terminals be weakly coordinated w.r.t. some $Q_{T_1T_2S}$.

THEOREM 3.7 The MAC $p_c(s|t_1,t_2)$ supports transmission at rates (R_0,R_1,R_2) with weak coordination w.r.t. a PMF $Q_{ST_1T_2}$ of conditional law $p_c(s|t_1,t_2)$ if and only if there exists a joint distribution on (W,S,T_1,T_2) of ST_1T_2-marginal $Q_{ST_1T_2}$ satisfying the Markov conditions

$$T_1 \to \ W \ \to T_2 \tag{3.44}$$

$$W \to \ (T_1,T_2) \ \to S \tag{3.45}$$

and the rate constraints

$$R_1 \ \leq \ \mathrm{I}(T_1;S \mid T_2,W) \tag{3.46a}$$

$$R_2 \ \leq \ \mathrm{I}(T_2;S \mid T_1,W) \tag{3.46b}$$

$$R_1+R_2 \ \leq \ \mathrm{I}(T_1,T_2;S \mid W) \tag{3.46c}$$

$$R_0+R_1+R_2 \ \leq \ \mathrm{I}(T_1,T_2;S). \tag{3.46d}$$

Proof. We begin with the proof of achievability. Denote the postulated joint PMF $Q_{ST_1T_2W}$, and let (R_0,R_1,R_2) satisfy (3.46) with strict inequalities (under $Q_{ST_1T_2W}$). Consider the random code construction that was proposed by Slepian and Wolf for the MAC with common and private messages [21]. They showed that if a joint PMF $Q_{T_1T_2W}$ is used in this scheme, then the average probability of error tends to zero

$$\lim_{n\to\infty} \Pr\big[(M_0,M_1,M_2) \neq (\hat{M}_0,\hat{M}_1,\hat{M}_2)\big] = 0, \tag{3.47}$$

where the probability is over the messages (M_0,M_1,M_2), the random code construction and the channel's randomness. Moreover, in this random code construction, the codewords are drawn IID $\sim Q_{T_1T_2W}$ and, consequently, for every triple $(s,t_1,t_2) \in \mathcal{S}\times\mathcal{T}_1\times\mathcal{T}_2$, the distribution of $\pi_{S^nT_1^nT_2^n}(s,t_1,t_2)$ is that of the empirical average of n IID mean-$Q_{ST_1T_2}(s,t_1,t_2)$ random variables. It therefore follows from the Weak Law of Large Numbers that, under random coding,

$$\operatorname*{plim}_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(\pi_{(S^n,T_1^n,T_2^n)};Q_{ST_1T_2}\right) = 0. \tag{3.48}$$

We next need to show the existence of good determinstic codes. Let $\mathtt{C}$ denote a generic code for our network, and $d_{\mathrm{TV}}(\mathtt{C})$ the Total Variation distance induced by it, i.e., the conditional expectation of $\mathsf{d}_{\mathrm{TV}}(\pi_{(S^n,T_1^n,T_2^n)}; Q_{ST_1T_2})$ given that the randomly chosen code is $\mathtt{C}$. By (3.48) there exists an increasing sequence $\{n'_k\}$ such that

$$\Pr\left(\left\{\mathtt{C} : d_{\mathrm{TV}}(\mathtt{C}) < \frac{1}{k}\right\}\right) > \frac{1}{2}, \quad \forall n > n'_k. \tag{3.49}$$

As to the probability of error, (3.47) implies the existence of an increasing sequence $\{n''_k\}$ such that

$$\Pr\left[(M_0, M_1, M_2) \neq (\hat{M}_0, \hat{M}_1, \hat{M}_2)\right] < \frac{1}{2k}, \quad \forall n > n''_k \tag{3.50}$$

and, consequently, by Markov's inequality,

$$\Pr\left(\left\{\mathtt{C} : P_{\mathrm{e}}(\mathtt{C}) < \frac{1}{k}\right\}\right) > \frac{1}{2}, \quad \forall n > n''_k, \tag{3.51}$$

where $P_{\mathrm{e}}(\mathtt{C})$ denotes average probability of error associated with $\mathtt{C}$ (i.e., the conditional probability of $(M_0, M_1, M_2) \neq (\hat{M}_0, \hat{M}_1, \hat{M}_2)$ given that the randomly chosen code is $\mathtt{C}$). It follows from (3.49) and (3.51) that, for every $\max\{n'_k, n''_k\} \le n$, we can find a code $\mathtt{C}$ for which neither $d_{\mathrm{TV}}(\mathtt{C})$ nor $P_{\mathrm{e}}(\mathtt{C})$ exceeds $1/k$. This choice establishes the direct part.

To prove the converse we follow the steps in [21], [14, Sec. 8.4] to obtain that, for every blocklength n,

$$R_1 \le \mathrm{I}(T_{1,U}; S_U \mid T_{2,U}, W) + \epsilon_n \tag{3.52a}$$

$$R_2 \le \mathrm{I}(T_{2,U}; S_U \mid T_{1,U}, W) + \epsilon_n \tag{3.52b}$$

$$R_1 + R_2 \le \mathrm{I}(T_{1,U}, T_{2,U}; S_U \mid W) + \epsilon_n \tag{3.52c}$$

$$R_0 + R_1 + R_2 \le \mathrm{I}(T_{1,U}, T_{2,U}; S_U) + \epsilon_n, \tag{3.52d}$$

where the chance variable U is equiprobably distributed over $[1 : n]$ and independent of $\{(S_i, T_{1,i}, T_{2,i})\}_{i=1}^n$; where W is an auxiliary chance variable satisfying

$$W \to (T_{1,U}, T_{2,U}) \to S \tag{3.53a}$$

$$T_{1,U} \to W \to T_{2,U}; \tag{3.53b}$$

and where ϵ_n tends to zero as n tends to infinity. Carathéodory's theorem shows that there exists a chance variable $\tilde{W}$ taking values in a set of cardinality $|\mathcal{T}_1||\mathcal{T}_2| + 2$ and having some joint PMF with the triple $(S_U, T_{1,U}, T_{2,U})$ such that, when W is replaced by $\tilde{W}$, the rate constraints (3.52) and the Markov conditions (3.53) are satisfied.

The limit in probability in (3.42) is of bounded random variables, so the convergence in probability implies the convergence of the expectations

$$\lim_{n\to\infty} \mathbb{E}\big[\pi_{S^n,T_1^n,T_2^n}(s,t_1,t_2)\big] = Q_{ST_1T_2}(s,t_1,t_2), \quad \forall (s,t_1,t_2) \in \mathcal{S}\times\mathcal{T}_1\times\mathcal{T}_2, \tag{3.54}$$

where the expectation is over the messages (M_0, M_1, M_2) and the randomness in the channel. This expectation equals $P_{S_U T_{1,U} T_{2,U}}(s,t_1,t_2)$, and we thus conclude that

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(P_{S_U T_{1,U} T_{2,U}}; Q_{ST_1T_2}\right) = 0. \tag{3.55}$$

Compactness implies the existence of a subsequence of blocklengths along which the joint PMF of $(S_U, T_{1,U}, T_{2,U}, \tilde{W})$ converges. The converse now follows from continuity by considering limits along this subsequence using (3.55), the rate-constraints in (3.52) and the Markov chains (3.53) where W is replaced by $\tilde{W}$ in both. □

The RHS of (3.46d) is fully determined by $Q_{ST_1T_2}$, and $R_0 + R_1 + R_2 \le \mathrm{I}(T_1, T_2; S)$ is a necessary condition for the channel to support (R_0, R_1, R_2) and $Q_{ST_1T_2}$. Equality can be achieved, for example, by the rate triple $\big(\mathrm{I}(T_1, T_2; S), 0, 0\big)$, where the private messages are absent. But this need not be the only supported tuple with this sum-rate. We say that (R_0, R_1, R_2) is of maximum sum rate (for the law $p_{\mathrm{c}}(s|t_1,t_2)$ and target PMF $Q_{ST_1T_2}$) if

$$R_0 + R_1 + R_2 = \mathrm{I}(T_1, T_2; S), \tag{3.56}$$

where the RHS is computed w.r.t. $Q_{ST_1T_2}$.

How small can the common rate R_0 be in a maximal-sum-rate triple? As the following corollary shows, it can be as low as $\mathrm{C}(T_1; T_2 \to S)$ and no lower.

Corollary 3.2 Consider a PMF $Q_{ST_1T_2}$ whose conditional S-given-(T_1, T_2) distribution is the MAC's channel law $p_{\mathrm{c}}(s|t_1,t_2)$. If the rates (R_0, R_1, R_2) are such that (3.56) holds and that the MAC supports transmission at rates (R_0, R_1, R_2) with weak coordination w.r.t. $Q_{ST_1T_2}$, then

$$R_0 \ge \mathrm{C}(T_1; T_2 \to S). \tag{3.57}$$

Moreover, there exists such a rate tuple for which (3.57) holds with equality.

Proof. To prove that (3.57) is necessary, we note that (3.56) and (3.46c) imply that

$$R_0 \;\ge\; \mathrm{I}(T_1, T_2; S) - \mathrm{I}(T_1, T_2; S \mid W) \tag{3.58}$$

$$= \mathrm{I}(T_1, T_2, W; S) - \mathrm{I}(T_1, T_2; S \mid W) \tag{3.59}$$

$$= \mathrm{I}(W; S), \tag{3.60}$$

where the first equality follows from the Markov condition (3.45) and the second from the chain rule. Minimizing the RHS subject to (3.44)–(3.45) establishes (3.57).

We next turn to the second part of the corollary. Fix a joint distribution achieving $\mathrm{C}(T_1; T_2 \to S)$ and set $R_0 = \mathrm{I}(W; S)$. Now choose R_1 and R_2 so that (3.46c) holds with equality and so that (3.46a) and (3.46b) both hold. This is possible because (3.46a)–(3.46c) and (3.44) are the constraints that appear on a MAC without a common message, and on a MAC the sum-rate constraint is always pinching [4]. □

3.3 *Remote Simulation Through a MAC*

The network depicted in Fig. 6 is required to produce a sequence S^n that appears IID $\sim Q_S$, where $Q_S \in \mathcal{P}(\mathcal{S})$ is an inducible output distribution on the MAC $p_c(s|t_1, t_2)$, i.e., an output distribution for which the set $\mathcal{D}_{T_1 T_2} \subseteq \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2)$ comprising the joint input distributions that induce Q_S, i.e., for which

$$\sum_{t_1, t_2} Q_{T_1 T_2}(t_1, t_2)\, p_c(s|t_1, t_2) = Q_S(s), \quad \forall s \in \mathcal{S}, \tag{3.61}$$

is nonempty. To achieve this goal, a chance variable J that is equiprobably distributed is fed to the two stochastic simulators, which produce the respective channel inputs. We shall see that the least entropy of J (normalized by the blocklength) that makes this possible is

$$\min_{Q_{T_1 T_2 S} \in \mathcal{D}_{T_1 T_2 S}} \mathrm{C}(T_1; T_2 \to S), \tag{3.62}$$

where $\mathcal{D}_{T_1 T_2 S} \subseteq \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{S})$ comprises the joint PMFs whose conditional law is $p_c(s|t_1, t_2)$ and whose S-marginal is the given Q_S, i.e., having the form

$$Q_{T_1 T_2}(t_1, t_2)\, p_c(s|t_1, t_2), \quad Q_{T_1 T_2} \in \mathcal{D}_{T_1 T_2}. \tag{3.63}$$

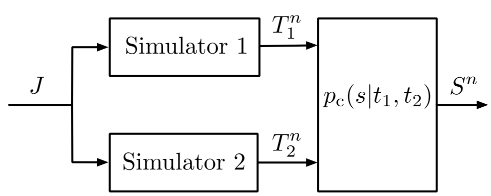

FIG. 6. Remote simulation through a multiple-access channel. The goal is for S^n to appear approximately IID $\sim Q_S$ (131).

The single-user version of this problem—which corresponds to $p_c(s|t_1, t_2)$ being a function of s and t_1 only—was studied by Wyner (under the normalized divergence criterion) [33, Thm. 6.3] and by Han and Verdú [13] and Cuff [7] (under the Total Variation criterion.) They showed that it suffices that the rate of J exceeds the minimum, over all input distributions that induce the given output distribution, of the mutual information between the channel terminals.

A naive approach to our problem would be to choose some $Q_{T_1 T_2}$ from $\mathcal{D}_{T_1 T_2}$ and to use J to induce input sequences of a joint law that closely approximates $Q_{T_1 T_2}^{\otimes n}$. This would require J to have normalized entropy $\mathrm{C}(T_1; T_2)$ or, upon optimizing over the choice of $Q_{T_1 T_2} \in \mathcal{D}_{T_1 T_2}$,

$$\min_{Q_{T_1 T_2} \in \mathcal{D}_{T_1 T_2}} \mathrm{C}(T_1; T_2). \tag{3.64}$$

As the following example shows, this is in general suboptimal: (3.64) can exceed (3.62).

EXAMPLE 3.8 Consider a MAC with binary input alphabets, $\mathcal{T}_1 = \mathcal{T}_2 = \{0, 1\}$, and the four-element output alphabet $\mathcal{S} = \mathcal{T}_1 \cup \{\iota, \delta\}$. If its inputs differ, the MAC produces the output δ (for 'differ'). Otherwise, it behaves like an erasure channel: it produces the output ι (for 'identical') w.p. ρ and the output that is equal to the inputs (which are identical) w.p. $1 - \rho$:

$$\begin{aligned} & p_c(s|t_1, t_2) \\ &= \mathbb{1}\{s = \delta \text{ and } t_1 \neq t_2\} \\ &\quad + (1-\rho) \cdot \mathbb{1}\{s = t_1 = t_2\} + \rho \cdot \mathbb{1}\{s = \iota \text{ and } t_1 = t_2\}. \end{aligned} \tag{3.65}$$

Consider now the target PMF

$$Q_S(s) = \begin{cases} 0 & s = \delta \\ \rho & s = \iota \\ (1-\rho)\frac{1}{2} & s \in \mathcal{T}_1 \end{cases}. \tag{3.66}$$

Since $Q_S(\delta)$ is zero, this output distribution can only be induced by a joint PMF under which T_1 and T_2 never differ. Moreover, to induce this output, T_1 must be distributed equiprobably. Thus, only the PMF

$$\tilde{Q}_{T_1T_2}(t_1, t_2) = \frac{1}{2}\mathbb{1}\{t_1 = t_2\}$$

induces this output distribution, and $\mathcal{D}_{T_1T_2}$ is a singleton. Under this PMF, $T_1 = T_2$ deterministically, so $\mathrm{C}(T_1; T_2)$ is the entropy of T_1, and (3.64) equals $\log(2)$. In contrast, (3.62) equals $\mathrm{C}(T_1; T_2 \to S)$, when the latter is computed under $\tilde{Q}_{T_1T_2}(t_1, t_2)p_c(s|t_1, t_2)$. It thus equals $(1 - \rho)\log(2)$, which is smaller than $\log(2)$ whenever ρ is positive.

The suboptimality of the naive approach is in failing to exploit the randomness introduced by the erasure channel: to simulate its output, it is unnecessary to have T_1^n ($= T_2^n$) be (roughly) uniform over $\{0, 1\}^n$: as we know from the single-user simulation problem, it suffices that it be uniform over a codebook containing approximately $e^{n(1-\rho)\log(2)}$ codewords.

We turn now to a formal statement of the problem. We say that the 'target PMF' $Q_S \in \mathcal{P}(S)$ can be *remotely simulated through the MAC* $p_c(s|t_1, t_2)$ *with rate* R if, for each blocklength n, there exists an index set $\mathcal{J}_n$ satisfying

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log|\mathcal{J}_n| \leq R \tag{3.67}$$

and independent *random mappings* $\Phi^{(n)}_{\mathrm{Rel},1}$ and $\Phi^{(n)}_{\mathrm{Rel},2}$, such that when J is drawn independently of them and equiprobably over $\mathcal{J}_n$, and their outputs

$$T_1^n = \Phi^{(n)}_{\mathrm{Rel},1}(J) \tag{3.68}$$

$$T_2^n = \Phi^{(n)}_{\mathrm{Rel},2}(J) \tag{3.69}$$

are sent over the MAC, the distribution P_{S^n} of the MAC's output sequence S^n closely resembles $Q_S^{\otimes n}$ in the sense that

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right) = 0. \tag{3.70}$$

THEOREM 3.9 Let the target PMF $Q_S \in \mathcal{P}(\mathcal{S})$ be inducible at the output of the MAC $p_{\mathrm{c}}(s|t_1,t_2)$ in the sense that $\mathcal{D}_{T_1T_2}$ above is nonempty. The PMF Q_S can be remotely simulated through the MAC with rate R if and only if

$$R \geq \min_{Q_{T_1T_2S}\in\mathcal{D}_{T_1T_2S}} \mathrm{C}(T_1;T_2 \to S), \tag{3.71}$$

where $\mathcal{D}_{T_1T_2S}$ is defined above.

Proof. The necessity of (3.17) (converse) is proved in Appendix C. Sufficiency (achievability) can be established using the scheme of Fig. 7 as follows. Let $Q_{ST_1T_2W} \in \mathcal{P}(\mathcal{S}\times\mathcal{T}_1\times\mathcal{T}_2\times\mathcal{W})$ be a PMF having a (T_1,T_2)-marginal $Q_{T_1T_2}$ in $\mathcal{D}_{T_1T_2}$ (i.e., for which (3.61) holds) and having the form

$$Q_{ST_1T_2W}(s,t_1,t_2,w) = Q_W(w)\, Q_{T_1|W}(t_1 \mid w)\, Q_{T_2|W}(t_2 \mid w)\, p_{\mathrm{c}}(s|t_1,t_2), \tag{3.72}$$

where W is an auxiliary chance variable that takes values in a set $\mathcal{W}$ and that has the PMF Q_W. This form guarantees that the Markov conditions in (3.1) are satisfied. Consider the scheme depicted in Fig. 7, where J is mapped to the codeword $w(J) \in \mathcal{W}^n$ in a codebook $\{w(j)\}$ indexed by $j \in [1 : e^{nR}]$. Simulator 1, which is random, feeds $w(J)$ to the DMC $Q_{T_1|W}(t_1 \mid w)$ and produces the resulting n-length output sequence. Simulator 2 does the same, but to the DMC $Q_{T_2|W}(t_2 \mid w)$. This setup is reminiscent of the one in Steinberg's resolvability problem over a MAC [22].

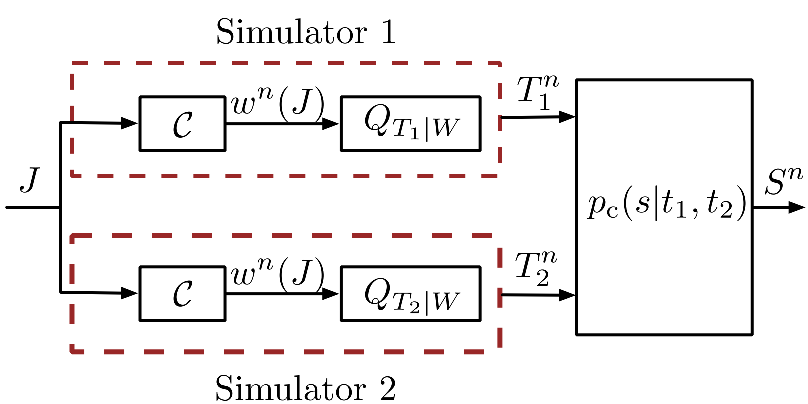

FIG. 7. A scheme for remote simulation through a MAC.

We need to show that if R exceeds $\mathrm{I}(W;S)$, then a codebook as above can be found for which the distibution of the MAC's output sequence P_{S^n} closely resembles $Q_S^{\otimes n}$ in the sense of (3.70). This can be proved using a random coding argument, where the codewords of the codebook $\{w(j)\}$ are drawn IID $\sim Q_W^{\otimes n}$: We claim that if R exceeds $\mathrm{I}(W;S)$, then the expectation (over the codebook) of $\mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right)$

(where P_{S^n} is the PMF of the n-length output sequence induced by the codebook) tends to zero. Once the claim is established, we can infer the existence of a deterministic sequence of codebooks (indexed by the blocklength) for which $\mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right)$ tends to zero. The claim follows directly from [7, Lemma IV.1] with the substitutions

$$V \leftarrow S, \qquad U \leftarrow W, \tag{3.73a}$$

and

$$\Phi_{V|U} \leftarrow Q_{S|W}(s \mid w) = \sum_{t_1,t_2} Q_{T_1|W}(t_1 \mid w) Q_{T_2|W}(t_2 \mid w) p_{\mathrm{c}}(s|t_1,t_2). \tag{3.73b}$$

□

3.4 *Remote Simulation Through a State-Dependent DMC*

In the network of Fig. 8, the relevant common information plays an important, but not decisive, role. A state-dependent discrete memoryless channel (SD-DMC) $\left(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}(t_1)\right)$ is driven by a state sequence $\{T_{1,i}\}$ that is drawn IID $\sim Q_{T_1}$. The goal is to produce a channel output sequence S^n whose law P_{S^n} resembles the product distribution $Q_S^{\otimes n}$ in the sense that

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right) = 0. \tag{3.74}$$

This is accomplished by having the state encoder describe the state sequence to the channel encoder using the codeword J in a rate-R codebook $\mathcal{J}_n$ of cardinality e^{nR}, and by having the shared common randomness K be drawn equiprobably and independently of T_1^n from a rate-R_K set $\mathcal{J}_{K,n}$ of cardinality e^{nR_K}. We seek the rate pairs (R, R_K) that make this possible.

A PMF $Q_S \in \mathcal{P}(\mathcal{S})$ *can be channel-synthesized with state-description rate R and common randomness rate R_K over the SD-DMC* $\left(p_c(s|t_1,t_2), Q_{T_1}(t_1)\right)$ if, for each blocklength n, there exist sets $\mathcal{J}_n$ and $\mathcal{J}_{K,n}$ satisfying

$$\varlimsup_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_n| \le R \tag{3.75}$$

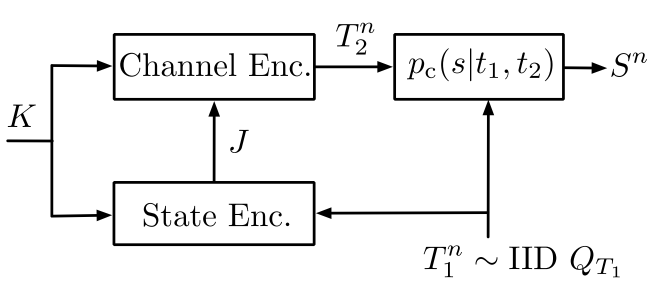

FIG. 8. Remote simulation over a state-dependent channel. The goal is for S^n to appear approximately IID $\sim Q_S$ (135).

and

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log|\mathcal{J}_{K,n}| \leq R_K \tag{3.76}$$

and *independent random* mappings

$$F_{\mathrm{Rel}}^{(n)} \colon \mathcal{T}_1^n \times \mathcal{J}_{K,n} \to \mathcal{J}_n \tag{3.77}$$

and

$$G_{\mathrm{Rel}}^{(n)} \colon \mathcal{J}_n \times \mathcal{J}_{K,n} \times \mathcal{S}^n \to \mathcal{T}_2^n \tag{3.78}$$

(that are independent of (T_1^n, K)) such that when the sequence

$$T_2^n = G_{\mathrm{Rel}}^{(n)}\big(F_{\mathrm{Rel}}^{(n)}(T_1^n, K), K\big) \tag{3.79}$$

is fed to the SD-DMC $\big(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}(t_1)\big)$, the PMF P_{S^n} of the output sequence S^n satisfies (3.74).

We say that the desired output law Q_S is inducible over the SD-DMC $\big(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}(t_1)\big)$ if there exists a joint PMF $Q_{T_1T_2S}$ of the following three properties: its T_1-marginal is the state law, its conditional $Q_{S|T_1T_2}$ is the channel law $p_{\mathrm{c}}(s|t_1,t_2)$, and its S-marginal is the desired output law. The subset of $\mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{S})$ comprising all such joint PMFs is denoted $\mathcal{D}(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}, Q_S)$.

THEOREM 3.8 An output law Q_S that is inducible over the SD-DMC of laws $\big(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}(t_1)\big)$ can be channel-synthesized over the said SD-DMC at rates (R, R_K) if and only if there exists a joint PMF $Q_{T_1T_2SW}$ whose T_1T_2S-marginal is in $\mathcal{D}(p_{\mathrm{c}}(s|t_1,t_2), Q_{T_1}, Q_S)$ and that satisfies the following four conditions:

$$T_1 \to W \to T_2 \tag{3.80a}$$

$$W \to (T_1, T_2) \to S \tag{3.80b}$$

$$R \;\geq\; \mathrm{I}(W; T_1) \tag{3.81a}$$

$$R + R_K \;\geq\; \mathrm{I}(W; S). \tag{3.81b}$$

Before proving the theorem, we make the following remark, whose proof is omitted.

REMARK 3.2 To exhaust the rate pairs promised in the theorem, we may restrict W to take values in an alphabet $\mathcal{W}$ of cardinality $|\mathcal{T}_1|\,|\mathcal{T}_2| + 1$, e.g.,

$$\mathcal{W}^{*'} = \{1, \ldots, |\mathcal{T}_1|\,|\mathcal{T}_2| + 1\}. \tag{3.82}$$

Moreover, said set of rate pairs is closed.

Proof of Theorem. 3.7 The proof of necessity (converse) resembles the one in [7, Section V]. The main differences are that in the steps recovering Inequality (3.81b) the source reconstruction pairs (T_1^n, T_2^n) (called (X^n, Y^n) in [7]) should be replaced by the SD-DMC's output sequence S^n and that the Markov condition (3.80b) requires justification. This and other details are presented in Appendix D.

Also the proof of the sufficiency (achievability) closely follows the proof of the main result in [7]. Let $Q_{ST_1T_2W} \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{W})$ be a PMF whose T_1-marginal is the given state PMF Q_{T_1}, whose S-marginal is the target PMF Q_S, and having the form

$$Q_{ST_1T_2W}(s, t_1, t_1, w) = Q_W(w)\, Q_{T_1|W}(t_1 \mid w)\, Q_{T_2|W}(t_2 \mid w)\, p_c(s|t_1, t_2), \tag{3.83}$$

where W is an auxiliary chance variable that takes values in a set $\mathcal{W}$ and that has the PMF Q_W. This form guarantees that the Markov conditions in (3.1) are satisfied.

Consider the random code construction and the simulators of Fig. 7 that were used to prove sufficiency for the MAC in Theorem 3.9 in Section 3.3, but denote the random index $\tilde{J} = (J, K)$ instead of J and its alphabet $\tilde{\mathcal{J}}_n = \mathcal{J}_n \times \mathcal{J}_{K,n}$ instead of $\mathcal{J}_n$. Let $\tilde{P}_{JKT_1^nT_2^nS^n|\mathcal{C}}$ denote the conditional-on-the-random-codebook-being-$\mathcal{C}$ joint PMF induced by the simulators described in Section 3.3 and the MAC $p_c(s|t_1, t_2)$, when J and K are independent and equiprobably distributed:

$$\tilde{P}_{JKT_1^nT_2^nS^n|\mathcal{C}} = \tilde{P}_{JK}\tilde{P}_{T_1^n|JK\mathcal{C}}\tilde{P}_{T_2^n|JK\mathcal{C}}\tilde{P}_{S^n|T_1^nT_2^n}, \tag{3.84}$$

where $\tilde{P}_{JK}$ is uniform over $\mathcal{J}_{K,n} \times \mathcal{J}_n$, the conditional PMFs $\tilde{P}_{T_1^n|JK\mathcal{C}}$ and $\tilde{P}_{T_2^n|JK\mathcal{C}}$ describe the operations of the two simulators and $\tilde{P}_{S^n|T_1^nT_2^n}$ is the n-fold product of the MAC's transition law (which is also our SD-DMC's transition law) $p_c(s|t_1, t_2)$.

Returning to our SD-DMC, to perform the remote simulation, we propose to randomly draw the codebook as in Section 3.3 for the MAC, and for any given realization of the codebook $\mathcal{C}$ apply the scheme illustrated in Fig. 9 based on the PMF $\tilde{P}_{JKT_1^nT_2^nS^n|\mathcal{C}}$ above.

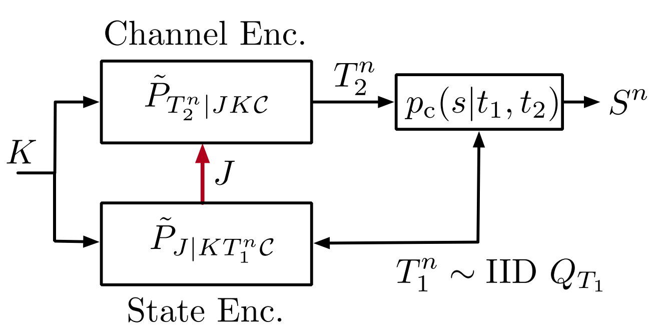

FIG. 9. A coding scheme for remote simulation over a state-dependent discrete memoryless channel.

Specifically, the Channel Encoder performs the same operations as Simulator 2 of Section 3.3, which is characterized by the conditional PMF $\tilde{P}_{T_2^n|JKC}$, and the state encoder uses the reverse encoder corresponding to the conditional PMF $\tilde{P}_{J|KT_1^n\mathcal{C}}$.

We analyze the *expected* Total Variation distance in (3.74) induced by the described state and channel encoders, averaged over the random choice of the codebook. Let $P_{JKT_1^nT_2^nS^n|\mathcal{C}}$ (without tilde) denote the joint PMF induced on $(J, K, T_1^n, T_2^n, S^n)$ by the state and channel encoders of Fig. 9 for a given code $\mathcal{C}$. By the Triangle inequality

$$\begin{aligned}
&\mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(P_{S^n|\mathbb{C}}; Q_S^{\otimes n}\right)\right] \\
&\quad \leq \mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(P_{S^n|\mathbb{C}}; \tilde{P}_{S^n|\mathbb{C}}\right)\right] + \mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(\tilde{P}_{S^n|\mathbb{C}}; Q_S^{\otimes n}\right)\right] && (3.85)\\
&\quad \leq \mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(P_{JKT_1^nT_2^nS^n|\mathbb{C}}; \tilde{P}_{JKT_1^nT_2^nS^n|\mathbb{C}}\right)\right] + \mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(\tilde{P}_{S^n|\mathbb{C}}; Q_S^{\otimes n}\right)\right] && (3.86)\\
&\quad \stackrel{(a)}{=} \mathbb{E}_K\left[\mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(Q_{T_1}^{\otimes n}; \tilde{P}_{T_1^n|K\mathbb{C}}\right)\right]\right] + \mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(\tilde{P}_{S^n|\mathbb{C}}; Q_S^{\otimes n}\right)\right] && (3.87)
\end{aligned}$$

where the second inequality follows from Proposition 1.2, and (a) holds by Proposition 1.3 because for each realization $\mathcal{C}$ of $\mathbb{C}$ the following hold: the PMFs $\tilde{P}_{K|\mathcal{C}}$ and P_K coincide (they are both uniform over the same set); the conditional PMF $\tilde{P}_{J|KT_1^n\mathcal{C}}$ coincides with $P_{J|KT_1^n\mathcal{C}}$; and the conditional PMF $\tilde{P}_{T_2^nS^n|JKT_1^n\mathcal{C}}$ coincides with $P_{T_2^nS^n|JKT_1^n\mathcal{C}}$. We now study the two expectations on the RHS of (3.87) separately, starting with the second. By [7, Lemma IV.1] (with the substitutions in (3.73) and $J \leftarrow (J, K)$), the expectation $\mathbb{E}_{\mathbb{C}}[\mathsf{d}_{\mathrm{TV}}\left(\tilde{P}_{S^n|\mathbb{C}}; Q_S^{\otimes n}\right)]$ tends to 0 as $n \to \infty$ if

$$\frac{1}{n}\log|\mathcal{J}_n| + \frac{1}{n}\log|\mathcal{J}_{K,n}| \geq \mathrm{I}(S;W) + \epsilon. \tag{3.88}$$

As to the first, we fix a realization $K = k$ and employ again Lemma IV.1 of [7], but now only for the random index J and using the substitutions

$$V \leftarrow T_1, \qquad U \leftarrow W, \qquad \Phi_{V|U} \leftarrow Q_{T_1|W}. \tag{3.89}$$

The lemma implies that, for each realization of $K = k$, the expectation $\mathbb{E}_{\mathbb{C}}\left[\mathsf{d}_{\mathrm{TV}}\left(Q_{T_1}^{\otimes n}; \tilde{P}_{T_1^n|K=k,\mathbb{C}}\right)\right]$ tends to 0 as $n \to \infty$ if

$$\frac{1}{n}\log|\mathcal{J}_n| \geq \mathrm{I}(T_1;W) + \epsilon. \tag{3.90}$$

Under the two conditions (3.88) and (3.90) there must thus be a sequence (one for each n) of realizations of the code construction $\mathcal{C}$ such that the Total Variation distance in (3.74) vanishes.

It remains to get rid of the ϵ. This is just a technical matter: Since ϵ can be any positive number, for any (R, R_K) satisfying (142), there exists a sequence $\{\epsilon_n\}_{n=1}^{\infty} \downarrow 0$ such that for each n it is possible to choose sets $\mathcal{J}_n = \{1, \ldots, \lfloor e^{n(R+\epsilon_n)} \rfloor\}$ and $\mathcal{J}_{K,n} = \{1, \ldots, \lfloor e^{n(R_K+\epsilon_n)} \rfloor\}$ and a deterministic codebook $\mathcal{C}$ so that our proposed encoders produce sequences (T_1^n, T_2^n) satisfying (3.74). Now

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\log|\mathcal{J}_n| &= R && (3.91)\\
\lim_{n\to\infty}\frac{1}{n}\log|\mathcal{J}_{K,n}| &= R_K, && (3.92)
\end{aligned}$$

and the achievability proof is complete. □

We now focus on the minimum sum-rate $R + R_K$ that allows an inducible output law Q_S to be channel-synthesized over an SD-DMC $\big(p_c(s|t_1,t_2), Q_{T_1}(t_1)\big)$. This minimum sum-rate is achieved when $R_K = 0$, because a bit of state description is at least as valuable as a bit of common randomness. Indeed, since we allow for random state encoders, the state encoder can always include in its description a random bit that is then common. The minimum sum-rate is thus

$$\min_{Q_{T_1T_2S}\in\mathcal{D}(p_c(s|t_1,t_2),Q_{T_1},Q_S)} \min_{Q_{W|T_1T_2S}} \max\big\{\mathrm{I}(W;T_1), \mathrm{I}(W;S)\big\}, \tag{3.93}$$

where the second minimization is subject to (3.80a) and (3.80b). As the following two examples show, the minimum sum-rate is sometimes, though not always, related to the relevant common information. Whether it is or not depends on which term in the maximum is largest. We begin with an example where the common relevant information is key.

Example 3.11 Consider an SD-DMC that is noiseless in the sense that its output is the tuple comprising its input and state, so $\mathcal{S} = \mathcal{T}_1 \times \mathcal{T}_2$ and

$$S = (T_1, T_2). \tag{3.94}$$

Irrespective of W and of the output PMF Q_S,

$$\begin{aligned} \mathrm{I}(S;W) &= \mathrm{I}(T_1;W) + \mathrm{I}(T_2;W \mid T_1) &(3.95)\\ &\geq \mathrm{I}(T_1;W), &(3.96)\end{aligned}$$

and $\max\{\mathrm{I}(W;T_1), \mathrm{I}(W;S)\}$ thus equals $\mathrm{I}(W;S)$. Consequently, the minimum sum-rate (3.93) for this channel is

$$\min_{Q_{T_1T_2S}\in\mathcal{D}(p_c(s|t_1,t_2),Q_{T_1},Q_S)} \mathrm{C}(T_1;T_2 \to S). \tag{3.97}$$

In our next example the relevant common information does not play a role, because, rather than being $\mathrm{I}(W;S)$, the maximum between $\mathrm{I}(W;T_1)$ and $\mathrm{I}(W;S)$ in (3.93) is $\mathrm{I}(W;T_1)$.

Example 3.12 Consider an SD-DMC whose law is as in (3.65) of Example 3.8 and whose state T_1 is drawn equiprobably from $\{0,1\}$. Let the target output PMF Q_S be as in (3.66) of that example. As in that example, since $Q_S(\delta)$ is zero, this output distribution can only be induced by a joint PMF under which T_1 and T_2 never differ. The sole element of $\mathcal{D}(p_c(s|t_1,t_2), Q_{T_1}, Q_S)$ is thus the PMF

$$Q_{T_1T_2S}(t_1,t_2,s) = \left(\frac{1}{2}\mathbb{1}\{t_1 = t_2 = 0\} + \frac{1}{2}\mathbb{1}\{t_1 = t_2 = 1\}\right) p_c(s|t_1,t_2) \tag{3.98}$$

and the first mimization in (3.93) is superfluous. Moreover, since T_1 and T_2 never differ, the Markov condition (3.80a) implies that T_1 is computable from W, and consequently $\mathrm{I}(W;T_1) = \mathrm{H}(T_1) = \log 2$. The minimum sum-rate in (3.93) thus equals $\max\{\log 2, \mathrm{C}(T_1;T_2 \to S)\}$. Since $\mathrm{C}(T_1;T_2 \to S)$ equals $(1-\rho)\log 2$, the minimum sum-rate is $\log 2$ and unrelated to $\mathrm{C}(T_1;T_2 \to S)$.

4. The Converse Part of the Proof of Theorem 3.2

To prove the converse part of Theorem 3.6, fix a target PMF $Q_{ST_1T_2}$ and an achievable rate triple (R_0, R_1, R_2) satisfying the no-excess-rate condition (3.25). The achievability of the rate triple guarantees the existence, for every blocklength n, of index sets $\mathcal{J}_{0,n}, \mathcal{J}_{1,n}, \mathcal{J}_{2,n}$ satisfying

$$\varlimsup_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_{0,n}| \leq R_0 \tag{4.1a}$$

$$\varlimsup_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_{1,n}| \leq R_1 \tag{4.1b}$$

$$\varlimsup_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_{2,n}| \leq R_2 \tag{4.1c}$$

and the existence of corresponding encoder $\phi^{(n)}_{\mathrm{Rel}}$ and decoders $\psi^{(n)}_{\mathrm{Rel},1}$ and $\psi^{(n)}_{\mathrm{Rel},2}$ for which the sequences T_1^n and T_2^n satisfy the weak coordination constraint (3.17) that the random empirical distribution $\pi_{(S^n,T_1^n,T_2^n)}$ approach the target PMF $Q_{ST_1T_2}$ in probability

$$\operatorname*{plim}_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(\pi_{(S^n,T_1^n,T_2^n)}; Q_{ST_1T_2}\right) = 0, \tag{4.2}$$

or, equivalently,

$$\operatorname*{plim}_{n\to\infty} \pi_{(S^n,T_1^n,T_2^n)}(s,t_1,t_2) = Q_{ST_1T_2}(s,t_1,t_2), \quad \forall (s,t_1,t_2) \in \mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2. \tag{4.3}$$

The convergence in probability of bounded random variables implies their convergence in expectation. Since the expectation of $\pi_{(S^n,T_1^n,T_2^n)}(s,t_1,t_2)$ is the evaluation of the uniform mixture of the PMFs $\{P_{S_iT_{1,i}T_{2,i}}\}_{i=1}^n$ at (s,t_1,t_2), it follows that $n^{-1}\sum_{i=1}^n P_{S_iT_{1,i}T_{2,i}}$ converges componentwise to $Q_{ST_1T_2}$ or, equivalently, $P_{S_UT_{1,U}T_{2,U}}$ converges in Total Variation to $Q_{ST_1T_2}$ whenever the chance variable U is drawn equiprobably from $[1:n]$ and independently of $\{(S_i, T_{1,i}, T_{2,i})\}$:

$$\begin{aligned} &\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(P_{S_UT_{1,U}T_{2,U}}; Q_{ST_1T_2}\right) = 0, \\ &U \sim \mathrm{Uni}([1:n]), \quad U \perp\!\!\!\perp \{(S_i, T_{1,i}, T_{2,i})\}_{i=1}^n. \end{aligned} \tag{4.4}$$

This latter statement will be crucial to the converse. By the continuity of mutual information we also obtain, under the same assumptions on U,

$$\lim_{n\to\infty} \mathrm{I}(T_{1,U}, T_{2,U}; S_U) = I(T_1, T_2; S), \tag{4.5}$$

where the RHS is computed w.r.t. $Q_{T_1T_2S}$.

We shall need the following lemma.

Lemma 4.1 Assume that S^n, T_1^n, T_2^n, J_0 are as above and, in particular, that they are produced under the no-excess-rate condition (3.25) and that the weak coordination constraint (4.2) is satisfied. Let $P_{S^nT_1^nT_2^nJ_0}$ denote their joint PMF. Then for every blocklength n, there exist

- a positive ϵ_n for which $\{\epsilon_n\} \downarrow 0$;
- a chance variable W taking values in an alphabet $\mathcal{W}$ of size

$$|\mathcal{W}| \leq |\mathcal{T}_1|^{n\epsilon_n^{2/5}} |\mathcal{T}_2|^{n\epsilon_n^{2/5}} \tag{4.6}$$

  and having some conditional PMF $P_{W|S^n T_1^n T_2^n J_0}$ given (S^n, T_1^n, T_2^n, J_0); and

- a subset $\mathcal{N} \subseteq [1:n]$ of size

$$|\mathcal{N}| \geq \big(1 - 2\log(2)\, \alpha\, \epsilon_n^{1/5}\big) n, \tag{4.7}$$

where

$$\alpha \triangleq \log(|\mathcal{T}_1||\mathcal{T}_2|) + \epsilon_n^{3/5} \tag{4.8}$$

such that for any $\rho \in (0,1)$ and under the joint PMF

$$P(s^n, t_1^n, t_2^n, j, w) = P_{S^n T_1^n T_2^n J_0}(s^n, t_1^n, t_2^n, j) \cdot P_{W|S^n T_1^n T_2^n J_0}(w \mid s^n, t_1^n, t_2^n, j) \tag{4.9}$$

over $\mathcal{S}^n \times \mathcal{T}_1^n \times \mathcal{T}_2^n \times \mathcal{J}_{0,n} \times \mathcal{W}$ the following three requirements are satisfied

1. $\mathrm{I}(T_{1,i};\ T_{2,i} \mid J_0, W) \leq \epsilon_n^{3/5}, \quad \forall i \in [1:n]$;
2. $\frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i; J_0 \mid T_{1,i}, T_{2,i}, W) \leq \alpha\, \epsilon_n^{2/5}$;
3. $\Pr\big(\{w \in \mathcal{W} : \|P_{S_i|W}(\cdot \mid w) - P_S(\cdot)\|_1 \leq \rho\}\big) \geq 1 - \epsilon_n^{1/10}\rho^{-1}, \quad \forall i \in \mathcal{N}$.

Proof. The proof is based on Dueck's and Ahlswede's Wringing Lemmas [8, 1] and is provided in Appendix E. □

Fix a blocklength n, and let ϵ_n, W, and $\mathcal{N}$ be as in the above lemma. Let U be drawn equiprobably from $[1:n]$ and independently of $\big(S^n, T_1^n, T_2^n, W\big)$. Let $P_{S^n T_1^n T_2^n J_0 U}$, or $\mathbb{P}$ for short, be the extension of the PMF in (4.9) that also includes U:

$$\mathbb{P}(s^n, t_1^n, t_2^n, j, w, i) = P_{S^n T_1^n T_2^n J_0}(s^n, t_1^n, t_2^n, j) \cdot P_{W|S^n T_1^n T_2^n J_0}(w \mid s^n, t_1^n, t_2^n, j) \cdot \frac{1}{n}. \tag{4.10}$$

Define the following subsets of $[1:n] \times \mathcal{W}$:

$$\mathcal{A} = \Big\{(i,w) \colon \mathrm{I}(T_{1,i}; T_{2,i} \mid J_0, W = w) \leq \epsilon_n^{1/5}\Big\} \tag{4.11}$$

$$\mathcal{B} = \Big\{(i,w) \colon \mathrm{I}(S_i; J_0 \mid T_{1,i}, T_{2,i}, W = w) \leq \alpha \epsilon_n^{1/5}\Big\} \tag{4.12}$$

$$\mathcal{C} = \Big\{(i,w) \colon i \in \mathcal{N} \text{ and } \|P_{S_i|W}(\cdot \mid w) - P_S(\cdot)\|_1 \leq \rho\Big\} \tag{4.13}$$

$$\mathcal{D} \;=\; \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}, \tag{4.14}$$

where mutual informations are again w.r.t. the PMF $\mathbb{P}$ in (4.10).

We next show that, for any fixed $\rho \in (0,1)$, $\mathbb{P}(\mathcal{D})$ tends to one when ϵ_n tends to zero and hence

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{D}) = 1, \quad \forall \rho \in (0,1). \tag{4.15}$$

To show this we note that, by Lemma 4.1 and Markov's inequality,

$$\mathbb{P}\big((U,W) \in \mathcal{A}\big) \geq 1 - \epsilon_n^{2/5} \tag{4.16}$$

and

$$\mathbb{P}\big((U,W) \in \mathcal{B}\big) \geq 1 - \epsilon_n^{1/5}. \tag{4.17}$$

Moreover, by (4.7) and Requirement (3) in Lemma 4.1,

$$\begin{aligned}
&\mathbb{P}\big((U,W) \in \mathcal{C}\big) \\
&\quad \geq \mathbb{P}\big((U,W) \in \mathcal{C} \mid U \in \mathcal{N}\big) \cdot \mathbb{P}\big(U \in \mathcal{N}\big) &(4.18)\\
&\quad \geq \mathbb{P}\big((U,W) \in \mathcal{C} \mid U \in \mathcal{N}\big) \cdot (1 - 2\log(2) \cdot \alpha\epsilon_n^{1/5}) &(4.19)\\
&\quad \geq (1 - \epsilon_n^{1/10}\rho^{-1}) \cdot (1 - 2\log(2)\,\alpha\epsilon_n^{1/5}). &(4.20)
\end{aligned}$$

From (4.16), (4.17), (4.20), and the definition of $\mathcal{D}$ (4.14),

$$\begin{aligned}
&\mathbb{P}\big((U,W) \in \mathcal{D}\big) \\
&\quad = 1 - \mathbb{P}\Big(\big((U,W) \in \mathcal{A}^c\big) \cup \big((U,W) \in \mathcal{B}^c\big) \cup \; \big((U,W) \in \mathcal{C}^c\big)\Big) &(4.21)\\
&\quad \geq 1 - \mathbb{P}\big((U,W) \in \mathcal{A}^c\big) - \mathbb{P}\big((U,W) \in \mathcal{C}^c\big) &(4.22)\\
&\quad \geq (1 - \epsilon_n^{1/10}\rho^{-1}) \cdot (1 - 2\log 2\alpha\epsilon_n^{1/5}) - \epsilon_n^{1/5} - \epsilon_n^{2/5}, &(4.23)
\end{aligned}$$

which concludes the proof of (4.15) because, as n tends to infinity, ϵ_n tends to zero.

We turn now to the cardinality constraints. In what follows, all mutual informations are calculated w.r.t. the PMF $\mathbb{P}$ (4.10). Beginning with the common rate,

$$\begin{aligned}
&\frac{1}{n}\log|\mathcal{J}_{0,n}| \\
&\quad \geq \frac{1}{n}\mathrm{H}(J_0 \mid W) &(4.24)\\
&\quad = \frac{1}{n}\mathrm{H}(J_0, W) - \frac{1}{n}\mathrm{H}(W) &(4.25)
\end{aligned}$$

$$\geq \frac{1}{n}\mathrm{I}(S^n; J_0, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.26}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0, W \mid S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.27}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0, W, S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.28}$$

$$\geq \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0 \mid W) - \frac{1}{n}\mathrm{H}(W) \tag{4.29}$$

$$\stackrel{(a)}{=} \sum_{(i,w)\in[1:n]\times\mathcal{W}} \frac{1}{n}\cdot\mathbb{P}(W=w)\cdot\mathrm{I}(S_i; J_0 \mid W=w, U=i) - \frac{1}{n}\mathrm{H}(W) \tag{4.30}$$

$$\stackrel{(b)}{\geq} \sum_{(i,w)\in\mathcal{D}} \mathbb{P}(U=i, W=w)\cdot\mathrm{I}(S_i; J_0 \mid W=w, U=i) - \epsilon_n^{2/5}\log(|\mathcal{T}_1||\mathcal{T}_2|), \tag{4.31}$$

where (a) holds because U is drawn equiprobably from $[1:n]$ and independently of (S^n, J_0, W); and in (b) we restricted the sum and used the cardinality bound (4.6) on $\mathcal{W}$.

Similarly,

$$\frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}|$$

$$\geq \mathrm{H}(J_0, J_1 \mid W) \tag{4.32}$$

$$= \frac{1}{n}\mathrm{H}(J_0, J_1, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.33}$$

$$\geq \frac{1}{n}\mathrm{H}(J_0, T_1^n, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.34}$$

$$\geq \frac{1}{n}\mathrm{I}(S^n; J_0, T_1^n, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.35}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0, T_1^n, W \mid S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.36}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0, T_1^n, W, S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.37}$$

$$\geq \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; J_0, T_{1,i} \mid W) - \frac{1}{n}\mathrm{H}(W) \tag{4.38}$$

$$\geq \sum_{(i,w)\in\mathcal{D}} \mathbb{P}(U=i, W=w) \cdot \mathrm{I}(S_i; J_0, T_{1,i} \mid W=w, U=i)$$
$$- \epsilon_n^{2/5} \cdot \log(|\mathcal{T}_1||\mathcal{T}_2|). \tag{4.39}$$

Likewise, by swapping $\mathcal{J}_{1,n}$ and $\mathcal{J}_{2,n}$,

$$\frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}|$$
$$\geq \sum_{(i,w)\in\mathcal{D}} \mathbb{P}(U=i, W=w) \cdot \mathrm{I}(S_i; J_0, T_{2,i} \mid W=w, U=i)$$
$$- \epsilon_n^{2/5} \cdot \log(|\mathcal{T}_1||\mathcal{T}_2|). \tag{4.40}$$

Finally,

$$\frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}|$$
$$\geq \mathrm{H}(J_0, J_1, J_2 \mid W) \tag{4.41}$$
$$= \frac{1}{n}\mathrm{H}(J_0, J_1, J_2, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.42}$$
$$\geq \frac{1}{n}\mathrm{H}(T_1^n, T_2^n, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.43}$$
$$\geq \frac{1}{n}\mathrm{I}(S^n; T_1^n, T_2^n, W) - \frac{1}{n}\mathrm{H}(W) \tag{4.44}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_1^n, T_2^n, W \mid S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.45}$$
$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_1^n, T_2^n, W, S^{i-1}) - \frac{1}{n}\mathrm{H}(W) \tag{4.46}$$
$$\geq \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_{1,i}, T_{2,i} \mid W) - \frac{1}{n}\mathrm{H}(W) \tag{4.47}$$
$$\geq \sum_{(i,w)\in\mathcal{D}} \mathbb{P}(U=i, W=w) \cdot \mathrm{I}(S_i; T_{1,i}, T_{2,i} \mid W=w, U=i)$$
$$- \epsilon_n^{2/5} \cdot \log(|\mathcal{T}_1||\mathcal{T}_2|). \tag{4.48}$$

Define now the new PMF

$$\begin{aligned}
&\lambda(s^n, t_1^n, t_2^n, i, w) \\
&\quad = \frac{\mathbb{P}(U = i, W = w)}{\mathbb{P}\big((U, W) \in \mathcal{D}\big)} \cdot \mathbb{1}\{(u, w) \in \mathcal{D}\} \\
&\quad \cdot \mathbb{P}(S^n = s^n, T_1^n = t_1^n, T_2^n = t_2^n, J_0 = j \mid W = w),
\end{aligned} \tag{4.49}$$

and note that the mutual informations

$$\mathrm{I}(S_i; J_0 \mid W = w) \tag{4.50}$$

$$\mathrm{I}(S_i; J_0, T_{1,i} \mid W = w) \tag{4.51}$$

$$\mathrm{I}(S_i; J_0, T_{2,i} \mid W = w) \tag{4.52}$$

$$\mathrm{I}(S_i; T_{1,i}, T_{2,i} \mid W = w) \tag{4.53}$$

$$\mathrm{I}(S_i; J_0 \mid T_{1i}, T_{2i}, W = w) \tag{4.54}$$

$$\mathrm{I}(T_{1,i}, T_{2,i} \mid J_0, W = w) \tag{4.55}$$

are the same under the PMFs $\mathbb{P}$ and λ. We can therefore rewrite the inequalities (4.31), (4.39), (4.40), and (4.48) as in Equation (4.56), where the mutual informations are w.r.t. λ. (To make this dependence explicit, we add the subscript λ to the mutual informations.)

$$\begin{aligned}
&\frac{1}{n} \log |\mathcal{J}_{0,n}| \\
&\quad \geq \mathbb{P}\big((U, W) \in \mathcal{D}\big) \sum_{(i,w)\in\mathcal{D}} \lambda(U = i, W = w) \cdot \mathrm{I}_\lambda(S_i; J_0 \mid W = w) \\
&\quad - \epsilon_n^{2/5} \log(|\mathcal{T}_1||\mathcal{T}_2|),
\end{aligned} \tag{4.56a}$$

$$\begin{aligned}
&\frac{1}{n} \log |\mathcal{J}_{0,n}| + \frac{1}{n} \log |\mathcal{J}_{1,n}| \\
&\quad \geq \mathbb{P}\big((U, W) \in \mathcal{D}\big) \cdot \sum_{(i,w)\in\mathcal{D}} \lambda(U = i, W = w) \cdot \mathrm{I}_\lambda(S_i; J_0, T_{1i} \mid W = w) \\
&\quad - \epsilon_n^{2/5} \log(|\mathcal{T}_1||\mathcal{T}_2|),
\end{aligned} \tag{4.56b}$$

$$\frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}|$$
$$\geq \mathbb{P}\big((U,W)\in\mathcal{D}\big)\cdot \sum_{(i,w)\in\mathcal{D}} \lambda(U=i,W=w)\cdot \mathrm{I}_\lambda(S_i;J_0,T_{2i}\mid W=w)$$
$$-\,\epsilon_n^{2/5}\log(|\mathcal{T}_1||\mathcal{T}_2|), \tag{4.56c}$$

$$\frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}|$$
$$\geq \mathbb{P}\big((U,W)\in\mathcal{D}\big)\cdot \sum_{(i,w)\in\mathcal{D}} \lambda(U=i,W=w)\cdot \mathrm{I}_\lambda(S_i;T_{1i},T_{2i}\mid W=w)$$
$$-\,\epsilon_n^{2/5}\log(|\mathcal{T}_1||\mathcal{T}_2|). \tag{4.56d}$$

Notice further that by the definition of the set $\mathcal{D}$, for each pair $(i,w)\in\mathcal{D}$ the following inequalities hold:

$$\mathrm{I}_\lambda(T_{1i};T_{2i}\mid J_0,W=w) \;\leq\; \epsilon_n^{1/5} \tag{4.57a}$$

$$\mathrm{I}_\lambda(S_i;J_0\mid T_{1i},T_{2i},W=w) \;\leq\; \alpha\cdot\epsilon_n^{1/5} \tag{4.57b}$$

$$\|\lambda_{S_i|W=w}-P_S\|_1 \;\leq\; \rho. \tag{4.57c}$$

We next cast (4.4) in terms of λ. To this end, note that by its definition (4.49),

$$\lambda\big(S_U=s,T_{1,U}=t_1,T_{2,U}=t_2\big)$$
$$=\mathbb{P}\big(S_U=s,T_{1,U}=t_1,T_{2,U}=t_2\,\big|\,(U,W)\in\mathcal{D}\big) \tag{4.58}$$

and by the law of total probability

$$\mathbb{P}(S_U=s,T_{1,U}=t_1,T_{2,U}=t_2)$$
$$=\mathbb{P}\big(S_U=s,T_{1,U}=t_1,T_{2,U}=t_2,\text{ and }(U,W)\in\mathcal{D}\big)$$
$$+\,\mathbb{P}\big(S_U=s,T_{1,U}=t_1,T_{2,U}=t_2,\text{ and }(U,W)\notin\mathcal{D}\big). \tag{4.59}$$

Consequently, since $\mathbb{P}\big((U,W)\in\mathcal{D}\big)$ approaches 1 as n tends to ∞ (4.23),

$$\lim_{n\to\infty}\mathsf{d}_{\mathrm{TV}}\left(\lambda_{S_UT_{1,U}T_{2,U}};\mathbb{P}_{S_UT_{1,U}T_{2,U}}\right)=0. \tag{4.60}$$

It follows from this and (4.4), using the Triangle inequality, that

$$\lim_{n\to\infty}\mathsf{d}_{\mathrm{TV}}\left(\lambda_{S_UT_{1,U}T_{2,U}};Q_{ST_1T_2}\right)=0 \tag{4.61}$$

and *a fortiori* (since ρ is positive) that for all sufficiently large values of n

$$\|\lambda_{S_U T_{1,U} T_{2,U}} - Q_{ST_2T_1}\|_1 \leq \rho, \quad n \text{ large.} \tag{4.62}$$

We continue the proof by studying the implications of (4.56), (4.57), and (4.62), which deal with λ rather than $\mathbb{P}$. The next step is to analyze the limiting behavior of these inequalities as $n \to \infty$ (and thus $\epsilon_n \to 0$) and $\rho \to 0$. The main difficulty is in analyzing the limiting behavior of the constraints in (4.56) and the sums in (4.57), because the range of the index i and the alphabets of the chance variables J_0 and W grow with the blocklength n. We circumvent this problem with the following lemma, whose proof requires two consecutive applications of Carathéodory's theorem.

LEMMA 4.2 There exists a subset $\mathcal{E} \subseteq \mathcal{D}$ whose size is at most $|\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2| + 4$ with a corresponding PMF on it $\alpha \in \mathcal{P}(\mathcal{E})$, and for each $(i, w) \in \mathcal{E}$ there exists a subset $\mathcal{J}_{i,w} \subseteq \mathcal{J}_{0,n}$ whose size is at most $|\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2| + 5$ with a corresponding PMF on it $\beta_{i,w} \in \mathcal{P}(\mathcal{J}_{i,w})$, so that the conditions in (4.56), (4.57), and (4.62) remain valid when the PMF λ is replaced by the PMF

$$\begin{aligned} &\nu(s^n, t_1^n, t_2^n, w, i, j) \\ &\quad = \alpha(i, w) \cdot \beta_{i,w}(j) \cdot P_{S^n, T_1^n, T_2^n | W, J_0}(s^n, t_1^n, t_2^n \mid w, j) \end{aligned} \tag{4.63}$$

and the summations is over $(i, w) \in \mathcal{E}$ (instead of over $(i, w) \in \mathcal{D}$).

Proof. See Appendix F. □

Notice that conditions (4.56), (4.57), and (4.62) depend on the elements of the sets $\mathcal{E}$ and $\{\mathcal{J}_{i,w}\}$ only through the conditional probability distribution $P_{S^n, T_1^n, T_2^n | W, J_0}(s^n, t_1^n, t_2^n \mid w, j)$. By relabeling these conditional distributions, we can assume that $\mathcal{E}$ does not depend on n and is equal to $\mathcal{E}^\star$, where

$$\mathcal{E}^\star = \{1, \ldots, |\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2| + 4\}. \tag{4.64}$$

Similarly, we can assume that $\mathcal{J}_{i,w}$ depends on neither n, i, or w and is equal to $\mathcal{J}^\star$, where

$$\mathcal{J}^\star = \{1, \ldots, |\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2| + 5\}. \tag{4.65}$$

Since the alphabets are now all fixed and finite, the class of joint PMFs on them is compact, and we can pick a subsequence of blocklengths along which they converge. Let $\nu^\star \in \mathcal{P}(\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{E}^\star \times \mathcal{J}^\star)$ denote the limiting PMF, and let $(S^\star, T_1^\star, T_2^\star, \Xi^\star, J^\star) \sim \nu^\star$.

We now consider the limits of the relevant quantities in (4.56), (4.57), and (4.62) (with λ replaced by ν and with the summations in (4.56) being over $(i, w) \in \mathcal{E}^\star$) along this subsequence (with ϵ_n consequently tending to zero) and then let ρ approach zero. Since all involved chance variables are over fixed and finite alphabets, standard continuity arguments allow us to conclude that a rate-triple (R_0, R_1, R_2) is achievable with no excess-rate only if the following two-auxiliary condition holds: there exists a joint distribution satisfying

$$S^\star \perp\!\!\!\perp \Xi^\star; \tag{4.66}$$

$$T_2^\star \to \ (J^\star, \Xi^\star) \ \to T_1^\star \tag{4.67}$$

$$J^\star \to \ (\Xi^\star, T_1^\star, T_2^\star) \ \to S^\star, \tag{4.68}$$

under which the following inequalities hold

$$R_0 \ \geq \ \mathrm{I}(S^\star; J^\star \mid \Xi^\star) \tag{4.69a}$$

$$R_1 + R_0 \ \geq \ \mathrm{I}(S^\star; J^\star, T_1^\star \mid \Xi^\star) \tag{4.69b}$$

$$R_2 + R_0 \ \geq \ \mathrm{I}(S^\star; J^\star, T_2^\star \mid \Xi^\star) \tag{4.69c}$$

$$R_2 + R_1 + R_0 \ \geq \ \mathrm{I}(S^\star; T_1^\star, T_2^\star \mid \Xi^\star) \tag{4.69d}$$

$$R_2 + R_1 + R_0 \ = \ \mathrm{I}(S^\star; T_1^\star, T_2^\star), \tag{4.69e}$$

where the last equality accounts for the no-excess-rate condition and follows from (3.25) and (4.61).

We next show that this two-auxiliary condition implies the following one-auxiliary condition: there exists a joint distribution satisfying

$$T_1^\star \to (J^\star, \Xi^\star) \to T_2^\star \tag{4.70}$$

and

$$(J^\star, \Xi^\star) \to (T_1^\star, T_2^\star) \to S^\star \tag{4.71}$$

under which

$$R_0 \ \geq \ \mathrm{I}(S^\star; J^\star, \Xi^\star) \tag{4.72a}$$

$$R_1 + R_0 \ \geq \ \mathrm{I}(S^\star; T_1^\star, J^\star, \Xi^\star) \tag{4.72b}$$

$$R_2 + R_0 \ \geq \ \mathrm{I}(S^\star; T_2^\star, J^\star, \Xi^\star). \tag{4.72c}$$

From this the converse will follow by defining $W^\star$ as

$$W^\star = (J^\star, \Xi^\star). \tag{4.73}$$

Condition (4.70) is just a restatement of (4.67); (4.72a) follows from (4.69a) and the independence condition (4.66); (4.72b) follows from (4.69b) and (4.66); and (4.72c) follows from (4.69c) and (4.66). It remains to establish (4.71).

To this end, we first observe that (4.69d) and the independence condition (4.66) imply that

$$R_2 + R_1 + R_0 \geq \mathrm{I}(S^\star; T_1^\star, T_2^\star, \Xi^\star). \tag{4.74}$$

This, (4.69e), and the chain rule imply that

$$\mathrm{I}(S^\star; \Xi^\star \mid T_1^\star, T_2^\star) = 0. \tag{4.75}$$

Consequently,

$$\mathrm{I}(S^\star; J^\star, \Xi^\star \mid T_1^\star, T_2^\star) = \mathrm{I}(S^\star; \Xi^\star \mid T_1^\star, T_2^\star) + \mathrm{I}(S^\star; J^\star \mid T_1^\star, T_2^\star, \Xi^\star) \tag{4.76}$$

$$= \mathrm{I}(S^\star; J^\star \mid T_1^\star, T_2^\star, \Xi^\star) \tag{4.77}$$

$$= 0, \tag{4.78}$$

where the first equality follows from the chain rule, the second from (4.75) and the last from (4.68). This establishes (4.71) and concludes the proof of the converse.

Data Availability Statement

No new data were generated or analyzed in support of this work.

Acknowledgment

The authors acknowledge helpful discussions with M. Bloch and thank the anonymous reviewers for helpful pointers to the literature.

Funding

This work was supported by the European Research Council [715111].

References

1. Ahlswede, R. (1985) The rate-distortion region for multiple descriptions without excess rate. *IEEE Trans. Inform. Theory*, **31**, 721–726.
2. Braun, G., Jain, R., Lee, T. & Pokutta, S. (2017) Information-theoretic approximations of the non-negative rank. *Comput. Complexity*, **26**, 147–197.
3. Braun, G. & Pokutta, S. (2016) Common information and unique disjointness. *Algorithmica*, **76**, 597–629.
4. Cover, T. M. & Thomas, J. A. (2006) *Elements of Information Theory*, 2nd edn. John Wiley & Sons.
5. Cuff, P. W. (2009) *Communication in networks for coordinating behavior*. PhD thesis, Stanford University.
6. Cuff, P. W., Permuter, H. H. & Cover, T. M. (2010) Coordination capacity. *IEEE Trans. Inform. Theory*, **56**, 4181–4206.
7. Cuff, P. W. (2013) Distributed channel synthesis. *IEEE Trans. Inform. Theory*, **59**, 7071–7096.
8. Dueck, G. (1981) The strong converse of the coding theorem for the multiple-access channel. *J. Comb. Inf. Syst. Sci.*, **6**, 187–196.
9. El Gamal, A. & Kim, Y.-H. (2006) *Network Information Theory*. Cambridge University Press.

10. Gács, P. & Körner, J. (1973) Common information is far less than mutual information. *Probl. Contr. Inform. Theory*, **2**, 149–162.

11. Graczyk, R. & Lapidoth, A. (2020) *Gray-Wyner and Slepian-Wolf guessing*. Los Angeles: Proc. IEEE Int. Symp. on Inform. Theory. USA, pp. 2207–2211.

12. Gray, R. M. & Wyner, A. D. (1974) Source coding for a simple network. *Bell Syst. Tech. J.*, **53**, 1681–1721.

13. Han, T. S. & Verdú, S. (1993) Approximation theory of output statistics. *IEEE Trans. Inform. Theory*, **39**, 752–772.

14. Kramer, G. (2007) Topics in multi-user information theory. *Found. Trends Commun. Inform. Theory*, **4**, 265–444.

15. Kumar, G. R., Li, C. T. & El Gamal, A. (2014) *Exact common information*. Hawaii: Proc. IEEE Int. Symp. on Inform. Theory. USA, pp. 161–165.

16. Lapidoth, A. & Wigger, M. (2016) *Conditional and relevant common information*. Israel, Eilat: Proc. ICSEE.

17. Liu, W. & Xu, G. (2010) *The common information of N dependent random variables*. Monticello: Proc. Annual Allerton Conference. USA, pp. 836–843.

18. Op t Veld, G. J. & Gastpar, M. (2015) *Caching Gaussians: minimizing total correlation on the Gray-Wyner network*. Proc. CISS. USA: Princeton, pp. 478–483.

19. Op t Veld, G. J. & Gastpar, M. (2015) *Caching two Gaussians. Proc. 36th WIC Symp. on Inform*. Belgium, Brussels: Theory in the Benelux, pp. 4–11.

20. Shayevitz, O. & Wigger, M. (2013) On the capacity of the discrete memoryless broadcast channel with feedback. *IEEE Trans. Inform. Theory*, **59**, 1329–1345.

21. Slepian, D. & Wolf, J. K. (1973) A coding theorem for multiple-access channels with correlated sources. *Bell Syst. Tech. J.*, **52**, 1037–1076.

22. Steinberg, Y. (1998) Resolvability theory for the multiple-access channel. *IEEE Trans. Inform. Theory*, **44**, 472–487.

23. Sula, E. & Gastpar, M. (2019) *Relaxed Wyner's common information*. Visby: Proc. IEEE ITW. Sweden, pp. 489–493.

24. Sula, E. & Gastpar, M. (2020) On Wyner's common information in the Gaussian case. Online: arXiv:1912.07083.

25. Timo, R., Grant, A., Chan, T. & Kramer, G. (2008) *Source coding for a simple network with receiver side information*. Toronto: Proc. IEEE Int. Symp. on Inform. Theory. Canada, pp. 2307–2311.

26. Timo, R. & Saeedi Bidokthi, S., Wigger, M. & Geiger, B. A. (2018) Rate-distortion approach to caching. *IEEE Trans. Inform. Theory*, **64**, 1957–1976.

27. Tuncel, E. (2009) *The rate transfer argument in two-stage scenarios: when does it matter?* Seoul: Proc. IEEE Int. Symp. on Inform. Theory. South Korea, pp. 41–45.

28. Tyagi, H. (2013) Common information and secret key capacity. *IEEE Trans. Inform. Theory*, **59**, 5627–5640.

29. Viswanathak, B., Akyol, E. & Rose, K. (2014) The lossy common information of correlated sources. *IEEE Trans. Inform. Theory*, **60**, 3238–3253.

30. Wang, C.-Y., Lim, S. H. & Gastpar, M. (2016) Information-theoretic caching: sequential coding for computing. *IEEE Trans. Inform. Theory*, **62**, 6393–6406.

31. Witsenhausen, H. S. (1975) On sequences of pairs of dependent random variables. *SIAM J. Appl. Math.*, **28**, 110–113.

32. Witsenhausen, H. S. (1976) Values and bounds for the common information of two discrete random variables. *SIAM J. Appl. Math.*, **31**, 313–333.

33. Wyner, A. D. (1975) The common information of two dependent random variables. *IEEE Trans. Inform. Theory*, **21**, 163–179.

34. Xu, G., Liu, W. & Chen, B. (2016) A lossy source coding interpretation of Wyner's common information. *IEEE Trans. Inform. Theory*, **62**, 754–768.

35. Yassaee, M. H., Aref, M. R. & Gohari, A. (2014) Achievability proof via output statistics of random binning. *IEEE Trans. Inform. Theory*, **60**, 6760–6786.
36. Yu, L., Li, H. & Chen, C. W. (2016) Generalized common informations: measuring commonness by the conditional maximal correlation. *Online: arXiv*, **1610**, 09289.
37. Yu, L. & Tan, V. (2018) Wyner's common information under Rényi divergence measures. *IEEE Trans. Inform. Theory*, **64**, 3616–3632.
38. Yu, L. & Tan, V. (2020) On exact and ∞-Rényi common informations. *IEEE Trans. Inform. Theory*, **66**, 3366–3406.
39. Zhang, Z. & Berger, T. (1987) New results in binary multiple descriptions. *IEEE Trans. Inform. Theory*, **33**, 502–521.

A. The Converse Part of Theorem 2.7

Before proceeding to the converse part of the proof of Theorem 2.7, we recall a lemma from [7].

Lemma A.1 (Lemma VI-3 in [7]). Let $\mathcal{A}$ be a finite set, and let $A^n \sim P_{A^n} \in \mathcal{P}(\mathcal{A}^n)$ be approximately IID in the sense that there exists some $Q \in \mathcal{P}(\mathcal{A})$ for which

$$\mathsf{d}_{\mathrm{TV}}\left(P_{A^n}; Q^{\otimes n}\right) \leq \epsilon \tag{A.1}$$

for some $\epsilon < 1/4$. Let the time-sharing chance variable U be uniform over $[1:n]$ and independent of A^n. Then,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(A_i; A^{i-1}) \leq 4\epsilon \log \frac{|\mathcal{A}|}{\epsilon}, \tag{A.2}$$

and

$$\mathrm{I}(A_U; U) \leq 4\epsilon \log \frac{|\mathcal{A}|}{\epsilon}. \tag{A.3}$$

We now establish the desired converse to Theorem 2.7.

The converse part of the proof of Theorem 2.7.
Consider a sequence of simulators $\{\Phi_{\mathrm{SI},1}^{(n)}\}_{n=1}^{\infty}$ and $\{\Phi_{\mathrm{SI},2}^{(n)}\}_{n=1}^{\infty}$ for which the induced outputs $\{T_1^n\}_{n=1}^{\infty}$ and $\{T_2^n\}_{n=1}^{\infty}$ and the SI sequence $\{Y^n\}_{n=1}^{\infty}$ satisfy (2.34), i.e., for which there exists a positive sequence $\{\epsilon_n\}_{n=1}^{\infty} \downarrow 0$ such that

$$\mathsf{d}_{\mathrm{TV}}\left(P_{T_1^n T_2^n Y^n}; Q_{T_1 T_2 Y}^{\otimes n}\right) < \epsilon_n. \tag{A.4}$$

Fix a blocklength n sufficiently large such that

$$\epsilon_n \leq \frac{1}{4}, \tag{A.5}$$

and note that for the chosen blocklength:

$$\frac{1}{n}\log|\mathcal{J}_n| = \frac{1}{n}\mathrm{H}(J) \tag{A.6}$$

$$\geq \frac{1}{n}\mathrm{I}(J; T_1^n, T_2^n \mid Y^n) \tag{A.7}$$

$$\geq \frac{1}{n} H(T_1^n, T_2^n \mid Y^n) - \frac{1}{n} \sum_{k=1}^{n} H(T_{1,k}, T_{2,k} \mid J, Y^n) \tag{A.8}$$

$$= \frac{1}{n} \big[H(T_1^n, T_2^n, Y^n) - H(Y^n) \big] - \frac{1}{n} \sum_{k=1}^{n} H(T_{1,k}, T_{2,k} \mid J, Y^n) \tag{A.9}$$

$$= \frac{1}{n} \Big[H(T_1^n, T_2^n, Y^n) - \sum_{k=1}^{n} H(Y_k) \Big] - \frac{1}{n} \sum_{k=1}^{n} H(T_{1,k}, T_{2,k} \mid J, Y^n) \tag{A.10}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \Big[H(T_{1,k}, T_{2,k}, Y_k) - I(T_{1,k}, T_{2,k}, Y_k; T_1^{k-1}, T_2^{k-1}, Y^{k-1}) - H(Y_k) - H(T_{1,k}, T_{2,k} \mid J, Y^n) \Big] \tag{A.11}$$

$$\stackrel{(a)}{\geq} \frac{1}{n} \sum_{k=1}^{n} \Big[H(T_{1,k}, T_{2,k}, Y_k) - 4\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n} - H(Y_k) - H(T_{1,k}, T_{2,k} \mid J, Y^n) \Big] \tag{A.12}$$

$$\stackrel{(b)}{=} H(T_{1,U}, T_{2,U}, Y_U \mid U) - H(Y_U \mid U) - H(T_{1,U}, T_{2,U} \mid J, Y_U, U, Y^{U-1}, Y_{U+1}^n) - 4\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_{2,U}||\mathcal{Y}|}{\epsilon_n} \tag{A.13}$$

$$\stackrel{(c)}{=} H(T_{1,U}, T_{2,U}, Y_U \mid U) - H(Y_U) - H(T_{1,U}, T_{2,U} \mid W_n, Y_U) - 4\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_{2,U}||\mathcal{Y}|}{\epsilon_n} \tag{A.14}$$

$$= H(T_{1,U}, T_{2,U}, Y_U) - I(T_{1,U}, T_{2,U}, Y_U; U) - H(Y_U) - H(T_{1,U}, T_{2,U} \mid W_n, Y_U) - 4\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_{2,U}||\mathcal{Y}|}{\epsilon_n} \tag{A.15}$$

$$\stackrel{(d)}{\geq} H(T_{1,U}, T_{2,U}, Y_U) - H(Y_U) - H(T_{1,U}, T_{2,U} \mid W_n, Y_U) - 8\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n} \tag{A.16}$$

$$= I(T_{1,U}, T_{2,U}; W_n \mid Y_U) - 8\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n}, \tag{A.17}$$

where (a) follows from Lemma A.1 because $\{(T_{1,k}, T_{2,k}, Y_k)\}$ are nearly IID (A.4), and $\epsilon_n < 1/4$; (b) holds when we draw U equiprobably from $[1 : n]$ and independently of the other chance variables (J, T_1^n, T_2^n, Y^n); (c) follows from the independence between U and Y_U and by defining $W_n \triangleq (J, U, Y^{U-1}, Y_{U+1}^n)$; and (d) follows from the second part of Lemma A.1.

To relate the RHS of (A.17) to $\mathrm{C}(T_1; T_2 \mid Y)$ (under $Q_{T_1 T_2 Y}$), we note that, with the above definitions of U and W_n, the independence between the encoding functions implies that

$$T_{1,U} \to (W_n, Y_U) \to T_{2,U} \tag{A.18}$$

forms a Markov chain. Consequently, if $\tilde{Q}_n$ denotes the joint PMF of $(T_{1,U}, T_{2,U}, Y_U)$, then

$$I(T_{1,U}, T_{2,U}; W_n \mid Y_U) \geq \mathrm{C}_{\tilde{Q}_n}(T_1; T_2 \mid Y), \tag{A.19}$$

where the conditional common information on the right is calculated under $\tilde{Q}_n$. This and (A.17) imply that

$$\frac{1}{n} \log |\mathcal{J}_n| \geq \mathrm{C}_{\tilde{Q}_n}(T_1; T_2 \mid Y) - 8\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n}. \tag{A.20}$$

The converse now follows by letting n tend to infinity because (2.34) and Proposition 1.5 imply that

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}} \left(\tilde{Q}_n; Q_{T_1 T_2 Y} \right) = 0 \tag{A.21}$$

and hence, by the continuity of the conditional common information,

$$\lim_{n\to\infty} \mathrm{C}_{\tilde{Q}_n}(T_1; T_2 \mid Y) = \mathrm{C}(T_1; T_2 \mid Y). \tag{A.22}$$

□

B. The Converse Part of Theorem 2.8

Consider a sequence of encoders and decoders $\{F_{\mathrm{SI}}^{(n)}\}_{n=1}^{\infty}$ and $\{G_{\mathrm{SI}}^{(n)}\}_{n=1}^{\infty}$ for which the sequences $\{T_1^n\}_{n=1}^{\infty}$, $\{T_2^n\}_{n=1}^{\infty}$, and $\{Y^n\}_{n=1}^{\infty}$ satisfy (2.45), i.e., for which there exists a positive sequence $\{\epsilon_n\}_{n=1}^{\infty}$ decaying to zero such that

$$\mathsf{d}_{\mathrm{TV}} \left(P_{T_1^n T_2^n Y^n}; Q_{T_1 T_2 Y}^{\otimes n} \right) < \epsilon_n. \tag{B.1}$$

A close inspection of the converse part of the proof of Theorem 2.7 (Appendix A) reveals that if one replaces $\frac{1}{n} \log |\mathcal{J}_n|$ by the sum $\frac{1}{n} \log |\mathcal{J}_n| + \frac{1}{n} \log |\mathcal{J}_{K,n}|$ and the index J by the pair (J, K), then all the steps remain valid except that in (A.6) the equality needs to be replaced by the inequality $\geq$. We thus conclude that for the setup under consideration here, for any blocklength n:

$$\frac{1}{n} \log |\mathcal{J}_n| + \frac{1}{n} \log |\mathcal{J}_{K,n}| \geq \mathrm{I}(W_n; T_{1,U}, T_{2,U} \mid Y_U) - 8\epsilon_n \log \frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n}, \tag{B.2}$$

where U is independent of $(J, K, T_1^n, T_2^n, Y^n)$ and equiprobable over $[1 : n]$, and $W_n \triangleq (J, K, U, Y^{U-1}, Y_{U+1}^n)$. Notice that the Markov chain $T_{1,U} \to (W_n, Y_U) \to T_{2,U}$ continues to hold, because T_2^n is a random (independent of T_1^n) function of (J, K, Y^n), so $T_{2,U}$ is a random mapping of (W_n, Y_U).

We next derive an additional inequality. Since J takes values in $\mathcal{J}_n$, for every blocklength n,

$$\begin{aligned}
\frac{1}{n} \log |\mathcal{J}_n| &\geq \frac{1}{n} \mathrm{I}(J; T_1^n \mid Y^n, K) && \text{(B.3)}\\
&= \frac{1}{n} \mathrm{I}(J, K; T_1^n \mid Y^n) && \text{(B.4)}\\
&= \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}(J, K; T_{1,i} \mid Y^n, T_1^{i-1}) && \text{(B.5)}
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(J,K,Y^{i-1},Y_{i+1}^{n};T_{1,i} \mid Y_i) \tag{B.6}$$

$$= \mathrm{I}(J,K,Y^{U-1},Y_{U+1}^{n};T_{1,U} \mid U,Y_U) \tag{B.7}$$

$$= \mathrm{I}(J,K,Y^{U-1},Y_{U+1}^{n},U;T_{1,U} \mid Y_U) \tag{B.8}$$

$$= \mathrm{I}(W_n;T_{1,U} \mid Y_U), \tag{B.9}$$

where (B.4) holds because the common randomness K is independent of (T_1^n, Y^n), so $\mathrm{I}(K;T_1^n \mid Y^n)$ is zero; (B.5) follows from the chain rule; and (B.6) holds because $\mathrm{H}(T_{1,i} \mid Y^n, T_1^{i-1})$ equals $\mathrm{H}(T_{1,i} \mid Y_i)$ (because $\{(T_{1,i}, Y_i)\}$ are IID).

Let $\tilde{Q}_n$ denote the joint PMF of $(T_{1,U}, T_{2,U}, Y_U)$. By (2.45) and Proposition 1.5

$$\lim_{n\to\infty} \mathsf{d}_{\mathrm{TV}}\left(\tilde{Q}_n; Q_{T_1T_2Y}\right) = 0. \tag{B.10}$$

It follows from Remark 2.3 and from (B.2) and (B.9) that there exists a chance variable W_n^* taking values in the finite set $\mathcal{W}^*$ of (2.48) and a joint distribution $\tilde{Q}_{T_{1,U}T_{2,U}Y_UW_n^*} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y} \times \mathcal{W}^*)$ under which

$$T_{1,U} \to (W_n^*, Y_U) \to T_{2,U} \tag{B.11a}$$

$$\frac{1}{n}\log|\mathcal{J}_n| + \frac{1}{n}\log|\mathcal{J}_{K,n}| \geq \mathrm{I}(W_n^*;T_{1,U},T_{2,U} \mid Y_U) - 8\epsilon_n \log\frac{|\mathcal{T}_1||\mathcal{T}_2||\mathcal{Y}|}{\epsilon_n} \tag{B.11b}$$

$$\frac{1}{n}\log|\mathcal{J}_n| \geq \mathrm{I}(W_n^*;T_{1,U} \mid Y_U). \tag{B.11c}$$

We next consider a subsequence $\{n_\nu\}$ along which $\tilde{Q}_{T_{1,U}T_{2,U}Y_UW_n^*}$ converges in Total Variation. Its marginal converges to $Q_{T_1T_2Y}$ by (B.10), and continuity implies that the limiting distribution satisfies the required Markov condition. Taking the limit superior of (B.11b) and (B.11c) along the subsequence establishes the converse.

C. The Converse Part of the Proof of Theorem 3.9

Proof of the necessity of (3.71). Consider a sequence of simulators $\{\Phi_{\mathrm{Rel},1}^{(n)}\}_{n=1}^{\infty}$ and $\{\Phi_{\mathrm{Rel},2}^{(n)}\}_{n=1}^{\infty}$ for which the induced MAC outputs $\{S^n\}_{n=1}^{\infty}$ satisfy (3.70), i.e., for which there exists a positive sequence $\{\epsilon_n\}_{n=1}^{\infty} \downarrow 0$ such that

$$\mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right) < \epsilon_n. \tag{C.1}$$

Fix a blocklength n sufficiently large so

$$\epsilon_n \leq 1/4. \tag{C.2}$$

Let T_1^n and T_2^n be the sequences produced by the encoders $\Phi_{\mathrm{Rel},1}^{(n)}$ and $\Phi_{\mathrm{Rel},2}^{(n)}$ when fed J. Let U be drawn equiprobably from $[1:n]$ and independently of the tuple (J, T_1^n, T_2^n, S^n), and define

$$W_n \triangleq (J, U). \tag{C.3}$$

Then,

$$\begin{aligned}
\frac{1}{n}\log|\mathcal{J}_n| &= \frac{1}{n}\mathrm{H}(J) \geq \frac{1}{n}\mathrm{I}(J; S^n) && \text{(C.4)}\\
&\geq \frac{1}{n}\mathrm{H}(S^n) - \frac{1}{n}\sum_{k=1}^{n}\mathrm{H}(S_k \mid J) && \text{(C.5)}\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[\mathrm{H}(S_k \mid S^{k-1}) - \mathrm{H}(S_k \mid J)\right] && \text{(C.6)}\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[\mathrm{I}(S_k; J) - \mathrm{I}(S_k; S^{k-1})\right] && \text{(C.7)}\\
&\overset{(a)}{\geq} \frac{1}{n}\sum_{k=1}^{n}\left[\mathrm{I}(S_k; J) - 4\epsilon_n\left(\log\frac{|\mathcal{S}|}{\epsilon_n}\right)\right] && \text{(C.8)}\\
&= \mathrm{I}(S_U; J \mid U) - 4\epsilon_n\left(\log\frac{|\mathcal{S}|}{\epsilon_n}\right) && \text{(C.9)}\\
&\overset{(b)}{\geq} \mathrm{I}(S_U; J, U) - 8\epsilon_n\left(\log\frac{|\mathcal{S}|}{\epsilon_n}\right) && \text{(C.10)}\\
&= \mathrm{I}(S_U; W_n) - 8\epsilon_n\left(\log\frac{|\mathcal{S}|}{\epsilon_n}\right), && \text{(C.11)}
\end{aligned}$$

where (a) follows by invoking the first part of Lemma A.1 and (b) the second.

Since the (possibly random) encoders $\Phi_{\mathrm{Rel},1}^{(n)}$ and $\Phi_{\mathrm{Rel},2}^{(n)}$ are independent,

$$T_{1,i} \to J \to T_{2,i}. \tag{C.12a}$$

And, since S_i is the output of a memoryless MAC of inputs $(T_{1,i}, T_{2,i})$,

$$J \to (T_{1,i}, T_{2,i}) \to S_i. \tag{C.12b}$$

These two Markov conditions together with the definition of W_n (C.3) and the independence between U and (T_1^n, T_2^n, S^n, J) imply

$$T_{1,U} \to W_n \to T_{2,U} \tag{C.13a}$$

and

$$W_n \to (T_{1,U}, T_{2,U}) \to S_U. \tag{C.13b}$$

Denoting by $\tilde{Q}_n$ the joint PMF of $(T_{1,U}, T_{2,U}, S_U)$, it now follows from (C.11) and (C13) that

$$\frac{1}{n}\log|\mathcal{J}_n| \;\geq\; \mathrm{C}_{\tilde{Q}_n}(T_1; T_2 \to S) - 8\epsilon_n\left(\log\frac{|\mathcal{S}|}{\epsilon_n}\right), \tag{C.14}$$

where the relevant common information is calculated under $\tilde{Q}_n$.

To derive the large-n limiting behavior of (C.14), we first note that, by compactness, from every subsequence of blocklengths, we can pick a subsequnce $\{n_k\}$ under which the (T_1, T_2)-marginal of $\tilde{Q}_n$ converges to some $Q^*_{T_1T_2} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2)$. It then follows from Proposition 1.3 that $\tilde{Q}_{n_k}$ converges to the PMF $Q^*_{T_1T_2}(t_1, t_2)\, p_{\mathrm{c}}(s|t_1, t_2)$, which we denote $Q^*_{T_1T_2S}$.

As we next argue, the S-marginal of the latter must be the target PMF Q_S, and $Q^*_{T_1T_2}$ must consequently be in $\mathcal{D}_{T_1T_2}$ and $Q^*_{T_1T_2S}$ in $\mathcal{D}_{T_1T_2S}$ of (3.71). To establish this it suffices to show that the S-marginal of $\tilde{Q}_n$ converges to the target PMF Q_S (Proposition 1.2), which is indeed the case by (C.1) and Proposition 1.5.

Having established that Q^* has the right form, we conclude that

$$\mathrm{C}_{Q^*}(T_1; T_2 \to S) \geq \min_{Q_{T_1T_2S} \in \mathcal{D}_{T_1T_2S}} \mathrm{C}(T_1; T_2 \to S). \tag{C.15}$$

Using this and a continuity argument establishes that we can deduce from (C.14) that

$$\varliminf_{k\to\infty} \frac{1}{n_k} \log |\mathcal{J}_{n_k}| \;\geq\; \min_{Q_{T_1T_2S} \in \mathcal{D}_{T_1T_2S}} \mathrm{C}(T_1; T_2 \to S). \tag{C.16}$$

Since this holds for every subsequence of blocklengths,

$$\varliminf_{n\to\infty} \frac{1}{n} \log |\mathcal{J}_n| \;\geq\; \min_{Q_{T_1T_2S} \in \mathcal{D}_{T_1T_2S}} \mathrm{C}(T_1; T_2 \to S), \tag{C.17}$$

and the necessity of (3.71) is established. □

D. The Converse Part of the Proof of Theorem 3.10

The converse part of the proof of Theorem 3.10. Consider sequences $\{\mathcal{J}_n\}$ and $\{\mathcal{J}_{K,n}\}$ of sets satisfying (3.75) and (3.76) and a sequence of state encoders $\{F^{(n)}_{\mathrm{Rel}}\}_{n=1}^{\infty}$ and channel encoders $\{G^{(n)}_{\mathrm{rel}}\}_{n=1}^{\infty}$ such that—with the channel state being $T_1^n \sim Q_{T_1}^{\otimes n}$, its description being $F^{(n)}_{\mathrm{Rel}}\big(T_1^n, K\big), K\big)$, and the channel input being $T_2^n = G^{(n)}_{\mathrm{Rel}}\big(F^{(n)}_{\mathrm{Rel}}\big(T_1^n, K\big), K\big)$—the channel output sequence $\{S^n\}_{n=1}^{\infty}$ satisfies (3.74). There then exists a positive sequence $\{\epsilon_n\}_{n=1}^{\infty} \downarrow 0$ such that, for each blocklength n,

$$\mathsf{d}_{\mathrm{TV}}\left(P_{S^n}; Q_S^{\otimes n}\right) < \epsilon_n. \tag{D.1}$$

We proceed as in Appendix C, but with the index J there replaced by the pair (J, K) here. Thus—rather than as in (C.3)—we now define

$$W_n \triangleq (J, K, U), \tag{D.2}$$

with U drawn equiprobably from $[1 : n]$ and independently of $(J, K, T_1^n, T_2^n, S^n)$. We repeat the steps leading from (C.4) to (C.11), but with the LHS of (C.4) replaced by $\frac{1}{n}\log|\mathcal{J}_n| + \frac{1}{n}\log|\mathcal{J}_{K,n}|$, with the index J replaced by the pair (J, K) and with the equality sign in (C.4) replaced by a $\geq$ sign. In this way we conclude that, in our current setup, for any blocklength n,

$$\frac{1}{n}\log|\mathcal{J}_n| + \frac{1}{n}\log|\mathcal{J}_{K,n}| \;\geq\; \mathrm{I}(W_n; S_U) - 8\epsilon_n \log \frac{|\mathcal{S}|}{\epsilon_n}. \tag{D.3}$$

In analogy to (C13), but with W_n defined in (D.2),

$$T_{1,U} \to W_n \to T_{2,U} \tag{D.4a}$$

$$W_n \to (T_{1,U}, T_{2,U}) \to S_U. \tag{D.4b}$$

We need an additional rate inequality, which we derive using the independence between J and K, the independence between K and T_1^n, the chain rule for mutual information, and the fact that $\{T_{1,i}\}$ are IID:

$$\begin{aligned}
\frac{1}{n}\log|\mathcal{J}_n| &= \frac{1}{n}\mathrm{H}(J \mid K) && \text{(D.5)}\\
&\geq \frac{1}{n}\mathrm{I}(J; T_1^n \mid K) && \text{(D.6)}\\
&= \frac{1}{n}\mathrm{I}(J, K; T_1^n) && \text{(D.7)}\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(J, K; T_{1,i} \mid T_1^{i-1}) && \text{(D.8)}\\
&\geq \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(J, K; T_{1,i}) && \text{(D.9)}\\
&= \mathrm{I}(J, K; T_{1,U} \mid U) && \text{(D.10)}\\
&= \mathrm{I}(J, K, U; T_{1,U}) && \text{(D.11)}\\
&= \mathrm{I}(W_n; T_{1,U}). && \text{(D.12)}
\end{aligned}$$

By the rate inequalities (D.3) and (D.12), the Markov conditions (D4), and the cardinality remark (Remark 3.2), we can extend the joint PMF of $(T_{1,U}, T_{2,U}, S_U)$ to a joint distribution $\tilde{Q}_{T_{1,U}T_{2,U}S_U W_n^*} \in \mathcal{P}(\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{S} \times \mathcal{W}^{*'})$ of $(T_{1,U}, T_{2,U}, S_U, W_n^*)$, where W_n^* takes values in the blocklength-independent finite set $\mathcal{W}^{*'}$ of (3.82), and where

$$T_{1,U} \to W_n^* \to T_{2,U} \tag{D.13a}$$

$$W_n^* \to (T_{1,U}, T_{2,U}) \to S_U \tag{D.13b}$$

$$\frac{1}{n}\log|\mathcal{J}_n| + \frac{1}{n}\log|\mathcal{J}_{K,n}| \geq \mathrm{I}(W_n^*; S_U) - 8\epsilon_n \log\frac{|\mathcal{T}_1||\mathcal{T}_2|}{\epsilon_n} \tag{D.13c}$$

and

$$\frac{1}{n}\log|\mathcal{J}_n| \geq \mathrm{I}(W_n^*; T_{1,U}). \tag{D.13d}$$

Consider a subsequence $\{n_k\}$ along which the sequences $\{n^{-1}\log|\mathcal{J}_n|\}$ and $\{n^{-1}\log|\mathcal{J}_{K,n}|\}$ both converge (to limits that by (3.75) and (3.76) lower-bound R and R_K) and along which $\tilde{Q}_{T_{1,U}T_{2,U}S_U W_n^*}$ converges in total variation to some $Q^*_{T_1T_2SW^*}$. Taking limits in (D.13) along this subsequence and using

(3.75), (3.76), and a continuity argument, we establish the validity of (3.80) and the necessity of (3.81) when those are calculated w.r.t. $Q^*_{T_1T_2SW^*}$.

It thus remains to show that the T_1T_2S-marginal of $Q^*_{T_1T_2SW^*}$ is in $\mathcal{D}(p_c(s|t_1,t_2), Q_{T_1}, Q_S)$.

Since $\{\tilde{Q}_{T_{1,U}T_{2,U}S_UW^*_{n_k}}\}_{k=1}^{\infty}$ converges to $\{\tilde{Q}_{T_{1,U}T_{2,U}S_UW^*_{n_k}}\}_{k=1}^{\infty}$, the same is true for the corresponding marginals (Corollary 1.1). The T_1-marginal of $Q^*_{T_1T_2SW^*}$ must thus be Q_{T_1}, because the sequence of $T_{1,U}$-marginals of $\{\tilde{Q}_{T_{1,U}T_{2,U}S_UW^*_{n_k}}\}_{k=1}^{\infty}$ is constant and equal to Q_{T_1}. Likewise the S-marginal of $Q^*_{T_1T_2SW^*}$ must be Q_S, because, by (3.74) and Proposition 1.5, the S_U-marginals of $\{\tilde{Q}_{T_{1,U}T_{2,U}S_UW^*_{n_k}}\}_{k=1}^{\infty}$ converge to Q_S.

Finally, $Q^*_{T_1T_2S}(t_1,t_2,s)$ factorizes as $Q^*_{T_1T_2}(t_1,t_2)\,p_c(s|t_1,t_2)$ by Corollary 1.2, because the PMF of $T_{1,U}T_{2,U}S_U$ factorizes in this way. □

E. Proof of Lemma 4.1

To prove Lemma 4.1, we begin by observing that the no-excess-rate condition (3.25) and the rate inequalities (3.18) imply that

$$\mathrm{I}(S;T_1,T_2) = R_0 + R_1 + R_2 \tag{E.1}$$

$$\geq \overline{\lim_{n\to\infty}} \frac{1}{n}\log|\mathcal{J}_{0,n}| + \overline{\lim_{n\to\infty}} \frac{1}{n}\log|\mathcal{J}_{1,n}| + \overline{\lim_{n\to\infty}} \frac{1}{n}\log|\mathcal{J}_{2,n}|. \tag{E.2}$$

Consequently, there exists a positive sequence $\{\epsilon_n^{(1)}\}$ converging to zero such that for all blocklengths n,

$$\mathrm{I}(S;T_1,T_2) \geq \frac{1}{n}|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}| - \epsilon_n^{(1)}. \tag{E.3}$$

Draw U equiprobably from $[1:n]$ and independently of $\{(S_i, T_{1,i}, T_{2,i})\}$. By (4.5), $\mathrm{I}(S_U;T_{1,U},T_{2,U})$ approaches $\mathrm{I}(S;T_1,T_2)$, so there exists a positive sequence $\{\epsilon_n^{(2)}\}$ converging to zero such that for all blocklengths n,

$$\mathrm{I}(S_U;T_{1,U},T_{2,U}) \geq \mathrm{I}(S;T_1,T_2) - \epsilon_n^{(2)}. \tag{E.4}$$

We now define

$$\epsilon_n = \epsilon_n^{(1)} + \epsilon_n^{(2)} \tag{E.5}$$

and begin with (E.3):

$$\mathrm{I}(S;T_1,T_2)$$
$$\geq \frac{1}{n}|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}| - \epsilon_n^{(1)} \tag{E.6}$$

$$\geq \frac{1}{n}\big(\mathrm{H}(J_0) + \mathrm{H}(J_1 \mid J_0) + \mathrm{H}(J_2 \mid J_0)\big) - \epsilon_n^{(1)} \tag{E.7}$$

$$\geq \frac{1}{n}\big(\mathrm{H}(J_0) + \mathrm{H}(T_1^n \mid J_0) + \mathrm{H}(T_2^n \mid J_0)\big) - \epsilon_n^{(1)} \tag{E.8}$$

$$\geq \frac{1}{n}\big(\mathrm{H}(J_0) + \mathrm{H}(T_1^n, T_2^n \mid J_0)\big) - \epsilon_n^{(1)} \tag{E.9}$$

$$= \frac{1}{n}\mathrm{H}(T_1^n, T_2^n, J_0) - \epsilon_n^{(1)} \tag{E.10}$$

$$\geq \frac{1}{n}\mathrm{H}(T_1^n, T_2^n) - \epsilon_n^{(1)} \tag{E.11}$$

$$\geq \frac{1}{n}\mathrm{I}(S^n; T_1^n, T_2^n) - \epsilon_n^{(1)} \tag{E.12}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_1^n, T_2^n \mid S^{i-1}) - \epsilon_n^{(1)} \tag{E.13}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_1^n, T_2^n, S^{i-1}) - \epsilon_n^{(1)} \tag{E.14}$$

$$\geq \frac{1}{n}\sum_{i=1}^{n}\mathrm{I}(S_i; T_{1,i}, T_{2,i}) - \epsilon_n^{(1)} \tag{E.15}$$

$$= \mathrm{I}(S_U; T_{1,U}, T_{2,U} \mid U) - \epsilon_n^{(1)} \tag{E.16}$$

$$= \mathrm{I}(S_U; T_{1,U}, T_{2,U}, U) - \epsilon_n^{(1)} \tag{E.17}$$

$$\geq \mathrm{I}(S_U; T_{1,U}, T_{2,U}) - \epsilon_n^{(1)} \tag{E.18}$$

$$\geq \mathrm{I}(S; T_1, T_2) - \epsilon_n^{(1)} - \epsilon_n^{(2)} \tag{E.19}$$

$$= \mathrm{I}(S; T_1, T_2) - \epsilon_n, \tag{E.20}$$

where the last inequality follows from (E.4) and the last equality from (E.5).

Since the RHS of (E.20) is within ϵ_n of its LHS, the RHS of (E.9) must be within ϵ_n of the RHS of (E.8). Consequently,

$$\mathrm{I}(T_1^n; T_2^n \mid J_0) \leq n\epsilon_n, \tag{E.21}$$

and, by Dueck's Wringing Lemma [8], there exists an index set $\tilde{\mathcal{N}}_W \subseteq [1:n]$ satisfying

$$\left|\tilde{\mathcal{N}}_W\right| \leq n\,\epsilon_n^{2/5} \tag{E.22}$$

and

$$\mathrm{I}(T_{1,i}; T_{2,i} \mid J_0, W) \leq \epsilon_n^{3/5}, \quad \forall i \in [1:n], \tag{E.23}$$

where W is defined as

$$W = \{(T_{1,i}, T_{2,i})\}_{i \in \tilde{\mathcal{N}}_W} \tag{E.24}$$

and therefore takes values in a set $\mathcal{W}$ whose cardinality is upper-bounded as in (4.6).

The chosen chance variable W thus fulfills Requirement (1) in the lemma. We now show that it also fulfills Requirement (2). To this end, observe that by (E.6):

$$
\begin{aligned}
&\mathrm{I}(S;T_1,T_2) \\
&\geq \frac{1}{n}\log|\mathcal{J}_{0,n}| + \frac{1}{n}\log|\mathcal{J}_{1,n}| + \frac{1}{n}\log|\mathcal{J}_{2,n}| - \epsilon_n^{(1)} && \text{(E.25)}\\
&\geq \frac{1}{n}\mathrm{H}(J_0,J_1,J_2) - \epsilon_n^{(1)} && \text{(E.26)}\\
&\geq \frac{1}{n}\mathrm{H}(T_1^n,T_2^n,J_0) + \frac{1}{n}\mathrm{H}(W) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.27)}\\
&\geq \frac{1}{n}\mathrm{H}(T_1^n,T_2^n,J_0,W) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.28)}\\
&\geq \frac{1}{n}\mathrm{I}(S^n;T_1^n,T_2^n,J_0,W) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.29)}\\
&= \frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;T_1^n,T_2^n,J_0,W \mid S^{i-1}) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.30)}\\
&= \frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;T_1^n,T_2^n,J_0,W,S^{i-1}) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.31)}\\
&\geq \frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;T_{1,i},T_{2,i},J_0,W) - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.32)}\\
&\geq \frac{1}{n}\sum_{i=1}^n \Big(\mathrm{I}(S_i;T_{1,i},T_{2,i}) + \mathrm{I}(S_i;J_0 \mid T_{1,i},T_{2,i},W)\Big) \\
&\quad - \frac{1}{n}\mathrm{H}(W) - \epsilon_n^{(1)} && \text{(E.33)}\\
&\geq \mathrm{I}(S;T_1,T_2) + \frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;J_0 \mid T_{1,i},T_{2,i},W) \\
&\quad - \frac{1}{n}\mathrm{H}(W) - \epsilon_n && \text{(E.34)}\\
&\geq \frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;J_0 \mid T_{1,i},T_{2,i},W) + \mathrm{I}(S;T_1,T_2) - \alpha\epsilon_n^{2/5}, && \text{(E.35)}
\end{aligned}
$$

where (E.34) can be argued by following the steps leading from (E.15) to (E.19), and where the last inequality holds because W takes on at most $(|\mathcal{T}_1||\mathcal{T}_2|)^{n\epsilon_n^{2/5}}$ distinct values and by the definition of α in (4.8). Inequality (E.35) establishes that Requirement (2) is also satisfied:

$$\frac{1}{n}\sum_{i=1}^n \mathrm{I}(S_i;J_0 \mid T_{1,i},T_{2,i},W) \leq \alpha\epsilon_n^{2/5}. \tag{E.36}$$

We next turn to the existence of the set $\mathcal{N}$ and to the fulfillment of Requirement (3). These follow by applying Ahlswede's Wringing Lemma [1, Lemma 2) with the substitution of S_t for X_t there; of W for Y there, and with the choice of γ there as

$$\gamma = (2\log 2)^{-1}\alpha^{-1}\epsilon_n^{-1/5} - 1. \tag{E.37}$$

Requirement (3) then follows from [1, Eq. (3.8c)) upon upper-bounding the entropy of W by that of a uniform (over the same support) and then noting that $\alpha^{-1}\log(|\mathcal{T}_1|\cdot|\mathcal{T}_2|) < 1$ by (4.8). The cardinality bound (4.7) follows from [1, Eq. (3.8a)). We conclude Appendix E with a remark on our application of Dueck's Wringing Lemma: Key is the choice of the exponent $3/5$ on the RHS of Requirement (1) (E.23); the exponent $2/5$ on the RHS of Requirement (2) ((E.36), implied by (E.22)); and the exponent $1/10$ on the RHS of Requirement (3) (also implied by (E.22)). To justify these exponents, observe that the main step in the proof of the converse part of is showing that

$$\lim_{n\to\infty} \Pr[(U,W)\in\mathcal{A}] = 1; \tag{E.38a}$$

$$\lim_{n\to\infty} \Pr[(U,W)\in\mathcal{B}] = 1; \tag{E.38b}$$

$$\lim_{n\to\infty} \Pr[(U,W)\in\mathcal{C}] = 1, \tag{E.38c}$$

where the chance variables U and W, and the events $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are defined in (4.10)–(4.13). Due to our judiciously chosen exponents, (E.38a) and (E.38b) follow from Requirements (1) and (2) via Markov's inequality (see (4.16) and (4.17)); and (E.38c) follows directly from (E.22) and Requirement 3) (see (4.21)–(4.23)).

F. Proof of Lemma 4.2

Proof of Lemma 4.2: Fix a triple $(s', t_1', t_2') \in \mathcal{S}\times\mathcal{T}_1\times\mathcal{T}_2$. A first application of Carathéodory's theorem establishes the existence of a subset $\mathcal{E}\subseteq\mathcal{D}$ of size not exceeding $|\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2|+4$ and a PMF $\alpha\in\mathcal{P}(\mathcal{E})$ such that

$$\sum_{(i,w)\in\mathcal{E}} \alpha(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0\mid W=w) = \sum_{(i,w)\in\mathcal{D}} \lambda_{UW}(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0\mid W=w) \tag{F.1}$$

$$\sum_{(i,w)\in\mathcal{E}} \alpha(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0,T_{1,i}\mid W=w) = \sum_{(i,w)\in\mathcal{D}} \lambda_{UW}(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0,T_{1,i}\mid W=w) \tag{F.2}$$

$$\sum_{(i,w)\in\mathcal{E}} \alpha(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0,T_{2,i}\mid W=w) = \sum_{(i,w)\in\mathcal{D}} \lambda_{UW}(i,w)\cdot \mathrm{I}_\lambda(S_i;J_0,T_{2,i}\mid W=w) \tag{F.3}$$

$$\sum_{(i,w)\in\mathcal{E}} \alpha(i,w)\cdot \mathrm{I}_\lambda(S_i;T_{1,i},T_{2,i}\mid W=w)$$
$$= \sum_{(i,w)\in\mathcal{D}} \lambda_{UW}(i,w)\cdot \mathrm{I}_\lambda(S_i;T_{1,i},T_{2,i}\mid W=w) \tag{F.4}$$

and such that for every $(s,x,y)\in\mathcal{S}\times\mathcal{T}_1\times\mathcal{T}_2$ other than (s',t_1',t_2')

$$\sum_{(i,w)\in\mathcal{E}} \alpha(i,w)\cdot \lambda_{S_iT_{1,i}T_{2,i}|W}(s,t_1,t_2\mid w)$$
$$= \sum_{(i,w)\in\mathcal{D}} \lambda_{UW}(i,w)\cdot \lambda_{S_iT_{1,i}T_{2,i}|W}(s,t_1,t_2\mid w). \tag{F.5}$$

Because probabilities sum to 1, these latter $|\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2|-1$ equalities ensure that (F.5) holds also for the triple (s',x',y').

We now apply Carathéodory's theorem a second time for J_0. Consider an arbitrary pair $(i,w)\in\mathcal{E}$. By Carathéodory's theorem, there exists a subset $\mathcal{J}_{i,w}\subset\mathcal{J}_{0,n}$ of size not exceeding $|\mathcal{S}||\mathcal{T}_1||\mathcal{T}_2|+5$ and a PMF $\beta_{i,w}\in\mathcal{P}(\mathcal{J}_{i,w})$ satisfying the conditions that

$$\sum_{j\in\mathcal{J}_{i,w}} \beta_{i,w}(j)\cdot \mathrm{H}_\lambda(S_i\mid J_0=j,W=w)$$
$$= \sum_{j\in\mathcal{J}_{0,n}} \lambda_{J_0|UW}(j\mid u,w)\cdot \mathrm{H}_\lambda(S_i\mid J_0=j,W=w) \tag{F.6}$$

$$\sum_{j\in\mathcal{J}_{i,w}} \beta_{i,w}(j)\cdot \mathrm{H}_\lambda(S_i\mid T_{1,i},J_0=j,W=w)$$
$$= \sum_{j\in\mathcal{J}_{0,n}} \lambda_{J_0|UW}(j\mid u,w)\cdot \mathrm{H}_\lambda(S_i\mid T_{1,i},J_0=j,W=w) \tag{F.7}$$

$$\sum_{j\in\mathcal{J}_{i,w}} \beta_{i,w}(j)\cdot \mathrm{H}_\lambda(S_i\mid T_{2,i},J_0=j,W=w)$$
$$= \sum_{j\in\mathcal{J}_{0,n}} \lambda_{J_0|UW}(j\mid u,w)\cdot \mathrm{H}({}_\lambda S_i\mid T_{2,i},J_0=j,W=w) \tag{F.8}$$

$$\sum_{j\in\mathcal{J}_{i,w}} \beta_{i,w}(j)\cdot \mathrm{H}_\lambda(S_i\mid T_{1,i},T_{2,i},J_0=j,W=w)$$
$$= \sum_{j\in\mathcal{J}_{0,n}} \lambda_{J_0|UW}(j\mid u,w)\cdot \mathrm{H}_\lambda(S_i\mid T_{1,i},T_{2,i},J_0=j,W=w) \tag{F.9}$$

$$\sum_{j\in\mathcal{J}_{i,w}} \beta_{i,w}(j)\cdot \mathrm{I}_\lambda(T_{1,i};T_{2,i}\mid J_0=j,W=w)$$
$$= \sum_{j\in\mathcal{J}_{0,n}} \lambda_{J_0|UW}(j\mid u,w)\cdot \mathrm{I}_\lambda(T_{1,i};T_{2,i}\mid J_0=j,W=w) \tag{F.10}$$

and that for every triple (s, t_1, t_2) in $\mathcal{S} \times \mathcal{T}_1 \times \mathcal{T}_2$ other than (s', t_1', t_2')

$$\sum_{j \in \mathcal{J}_{i,w}} \beta_{i,w}(j) \cdot \lambda_{S_i T_{1,i} T_{2,i} | J_0, W}(s, t_1, t_2 \mid j, w)$$
$$= \sum_{j \in \mathcal{J}_{i,w}} \lambda_{J_0|UW}(j \mid i, w) \cdot \lambda_{S_i T_{1,i} T_{2,i} | J_0 W}(s, t_1, t_2 \mid j, w). \tag{F.11}$$

(Again, because probabilities sum to 1, Equality (F.11) must also hold when (s, t_1, t_2) equals (s', t_1', t_2').) These conditions guarantee that the conditional entropies $\mathrm{H}(S_i \mid W = w)$, $\mathrm{H}(S_i \mid T_{1,i}, T_{2,i}, W = w)$, and the conditional joint PMF on $(T_{1,i}, T_{2,i}, S)$ given $W = w$ are the same under the PMF

$$\beta_{S_i, T_{1,i} T_{2,i} J_0 | W}(s, t_1, t_2, j \mid w) = \beta_{i,w}(j) \cdot \lambda_{S_i T_{1,i} T_{2,i} | J_0, W}(s, t_1, t_2 \mid j, w) \tag{F.12}$$

and

$$\lambda_{S_i T_{1,i} T_{2,i} J_0 | W}(s, x, y, j \mid w). \tag{F.13}$$

These guarantee that the terms in (4.57) do not change when we replace λ with the above β. □