Conditional and Relevant Common Information

[†]Amos Lapidoth and Michèle Wigger[‡]
† ETH Zurich, lapidoth@isi.ee.ethz.ch
Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France, michele.wigger@telecom-paristech.fr

***Abstract*—Two variations on Wyner's common information are proposed: *conditional common information* and *relevant common information*. The former characterizes the minimum common rate that is required for lossless source-coding over a one-to-two Gray-Wyner network, when the sum-rate is restricted to be minimal and the terminals all share the side-information. It also characterizes the minimum rate of common randomness that is required for two terminals sharing some side-information to strongly coordinate their outputs according to a target distribution. The latter, relevant common information, is an upper bound on the minimum common rate required for two receivers of a one-to-two Gray-Wyner network to weakly coordinate their reconstruction sequences with the source according to a target distribution. It also characterizes the minimum rate of common randomness that is required for two terminals to produce a target strongly-coordinated sequence at the output of a two-user multiple-access channel.**

I. Introduction

Wyner [1] defined the common information $C(T_1;T_2)$ between two random variables T_1 and T_2 as

$$C(T_1;T_2) \triangleq \min_{W\colon T_1 \to W \to T_2} I(T_1,T_2;W), \tag{1}$$

where $X \to Y \to Z$ indicates that X and Z are conditionally independent given Y, i.e., that X,Y,Z forms a Markov chain. He provided two operational meanings to this quantity: It is the smallest common rate required to losslessly describe a bivariate source $\{(T_{1,i},T_{2,i})\} \sim$ IID $Q_{T_1T_2}$ over a Gray-Wyner network (Fig. 2 with $\{Y_i\}$ null) with the sum-rate at its minimum, and it is also the smallest rate of common randomness required for two terminals to simulate outputs of joint distribution $Q_{T_1T_2}$ (Fig. 1 with $\{Y_i\}$ null).

Here we propose two generalizations of Wyner's common information. The first, the *conditional common information* $C(T_1;T_2|Y)$, accounts for side-information (SI) Y that is available to all terminals. The second, the *relevant common information* $C(T_1;T_2 \to S)$, quantifies the common information that is related to a random variable S.

Definition 1 (Conditional Common Information): Given a triple of random variables Y,T_1,T_2, the *conditional common information* of the pair (T_1,T_2) *given* Y is

$$C(T_1;T_2|Y) \triangleq \min_{W\colon T_1 \to (W,Y) \to T_2} I(T_1,T_2;W|Y). \tag{2}$$

Definition 2 (Relevant Common Information): Given a triple of random variables S,T_1,T_2, the *common information* of the pair (T_1,T_2) *relevant* to S is

$$C(T_1;T_2 \to S) \triangleq \min_{\substack{W\colon T_1 \to W \to T_2 \\ S \to (T_1,T_2) \to W}} I(S;W). \tag{3}$$

Remark 1: For $Y=\emptyset$ and for $S=(T_1,T_2)$, the conditional common information and the relevant common information reduce to Wyner's original common information:

$$Y=\emptyset \quad \Longrightarrow \quad C(T_1;T_2|Y)=C(T_1;T_2), \tag{4}$$

$$S=(T_1,T_2) \quad \Longrightarrow \quad C(T_1;T_2 \to S)=C(T_1;T_2). \tag{5}$$

In Section II, we will provide the following operational meanings to the conditional common information:

- It is the smallest common rate required to describe a bivariate source $\{(T_{1,i},T_{2,i})\}$ over the Gray-Wyner network of Fig. 2, where the side information $\{Y_i\}$ is available to all the terminals, and where the sum of all the rates is at its minimum.
- It is the smallest rate of common randomness required to strongly coordinate [3] the outputs of two terminals according to the target distribution $Q_{T_1T_2}$ when the two terminals are furnished with $\{Y_i\}$.

In Section III we provide operational meanings to the relevant common information:

- It is an upper bound on the smallest common rate required in a one-to-two Gray-Wyner network to weakly coordinate [3] the receivers' reconstructions $\{T_{1,i}\}$ and $\{T_{2,i}\}$ with each other and with the source $\{S_i\}$ according to a target distribution $Q_{T_1T_2S}$, where the sum of all rates needs to be at its minimum.
- It is the smallest rate of common randomness required at two terminals to—through their inputs—strongly coordinate the output of a two-user multiple-access channel (MAC) according to a target distribution Q_S.

Our definition of relevant conditional information in (3) is reminiscent of the definition of *lossy common information* in [4]. However, in [4], the minimization is not only over the auxiliary random variable W but also over all pairs (T_1,T_2) for which T_1 and T_2 reconstruct the source S up to given distortions D_1 and D_2.

II. On Conditional Common Information

We present two operational meanings of the conditional common information (Definition 1).

A. The Strong-Coordination Problem

Consider the scenario of Figure 1, where the side-information $\{Y_i\}$ is IID according to a given distribution Q_Y over a finite set $\mathcal{Y}$. For a given blocklength $n>0$, we define

978-1-5090-2152-9/16/$31.00

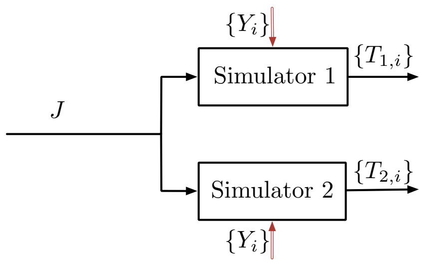

Fig. 1. A simulation problem with side-information.

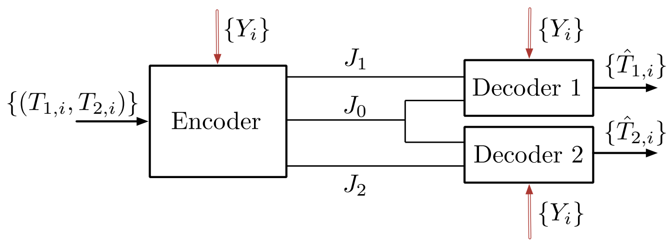

Fig. 2. Lossless Gray-Wyner source coding with side-information Y^n.

$Y^n := (Y_1, \ldots, Y_n)$, and let the common randomness J be uniformly distributed over the index set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$.

We say that a joint distribution $Q_{T_1T_2Y}$ over a finite product set $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{S}$ *can be strongly coordinated with rate* $R > 0$ *and SI* $\{Y_i\}$ if, for each blocklength $n > 0$, there exist functions $\varphi_{\text{SI},1}^{(n)}$ and $\varphi_{\text{SI},2}^{(n)}$ of appropriate domains, for which the sequences

$$T_1^n := \varphi_{\text{SI},1}^{(n)}(J, Y^n, \Theta_1) \tag{6}$$

$$T_2^n := \varphi_{\text{SI},2}^{(n)}(J, Y^n, \Theta_2) \tag{7}$$

satisfy

$$\left\| P_{T_1^n T_2^n Y^n} - Q_{T_1T_2Y}^{\otimes n} \right\|_{\text{TV}} \to 0 \quad \text{as} \quad n \to \infty. \tag{8}$$

Here Θ_1, Θ_2, and J are independent (with Θ_1 and Θ_2 accounting for local randomness); $P_{T_1^n T_2^n Y^n}$ denotes the probability distribution of the tuple (T_1^n, T_2^n, Y^n); and $Q_{T_1T_2Y}^{\otimes n}$ denotes the n-fold product distribution obtained from $Q_{T_1T_2Y}$. Also, $\| \cdot \|_{\text{TV}}$ stands for total variational distance [3].

Theorem 1: The distribution $Q_{T_1T_2Y}$ can be strongly coordinated with rate R and SI $\{Y_i\}$ if, and only if,

$$R \geq I(T_1, T_2; W|Y) \tag{9}$$

for some auxiliary random variable W satisfying

$$T_1 \to (W, Y) \to T_2. \tag{10}$$

Proof: Omitted. ■

As a corollary we obtain the following operational meaning for $C(T_1; T_2|Y)$:

Corollary 1.1: The distribution $Q_{T_1T_2Y}$ can be strongly coordinated with rate R and SI $\{Y_i\}$ if, and only if,

$$R \geq C(T_1; T_2|Y).$$

B. The Lossless Source-Coding Problem

Consider the lossless Gray-Wyner source coding problem of Figure 2, where the sequence of source and side-information triples $\{(T_{1,i}, T_{2,i}, Y_i)\}$ is IID according to a given distribution $Q_{T_1T_2Y}$ over a finite product alphabet $\mathcal{T}_1 \times \mathcal{T}_2 \times \mathcal{Y}$. For a given blocklength $n > 0$, define $T_1^n := (T_{1,1}, \ldots, T_{1,n})$, $T_2^n := (T_{2,1}, \ldots, T_{2,n})$ and $Y^n := (Y_1, \ldots, Y_n)$. The encoder observes all three sequences T_1^n, T_2^n, Y^n and produces the indices J_0, J_1, J_2

$$(J_0, J_1, J_2) = \phi_{\text{SI}}^{(n)}(T_1^n, T_2^n, Y^n), \tag{11}$$

for some encoding function

$$\begin{aligned}\phi_{\text{SI}}^{(n)} &: \mathcal{T}_1^n \times \mathcal{T}_2^n \times \mathcal{Y}^n \\ &\to \{1, \ldots, \lfloor 2^{nR_0} \rfloor\} \times \{1, \ldots, \lfloor 2^{nR_1} \rfloor\} \times \{1, \ldots, \lfloor 2^{nR_2} \rfloor\}.\end{aligned} \tag{12}$$

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

$$\hat{T}_1^n = \psi_{\text{SI},1}^{(n)}(J_0, J_1, Y^n), \tag{13}$$

$$\hat{T}_2^n = \psi_{\text{SI},2}^{(n)}(J_0, J_2, Y^n), \tag{14}$$

for some decoding functions $\psi_{\text{SI},1}^{(n)}$ and $\psi_{\text{SI},2}^{(n)}$ of appropriate domains.

The rate-triple (R_0, R_1, R_2) is *achievable* if, for each blocklength $n > 0$, there exists an encoding function $\phi_{\text{SI}}^{(n)}$ as in (12) and decoding functions $\psi_{\text{SI},1}^{(n)}$ and $\psi_{\text{SI},2}^{(n)}$ of appropriate domains, so that:

$$\lim_{n \to \infty} \Pr\big((T_1^n, T_2^n) \neq (\hat{T}_1^n, \hat{T}_2^n)\big) = 0. \tag{15}$$

Theorem 2: Given a joint distribution $Q_{T_1T_2Y}$, a rate-triple (R_0, R_1, R_2) is achievable if, and only if, there exists an auxiliary random variable W such that

$$R_0 \geq I(W; T_1, T_2|Y) \tag{16a}$$

$$R_1 \geq H(T_1|W, Y) \tag{16b}$$

$$R_2 \geq H(T_2|W, Y). \tag{16c}$$

Proof: Omitted. ■

Let $R_{0,\text{SI}}^\star$ be the minimal R_0 for which for some rates $R_1, R_2 \geq 0$ and auxiliary random variable W the triple (R_0, R_1, R_2) satisfies (16) and

$$R_0 + R_1 + R_2 = H(T_1, T_2|Y). \tag{17}$$

The rate $R_{0,\text{SI}}^\star$ thus indicates the minimum common rate R_0 in the lossless Gray-Wyner source coding problem with side-information so that for some $R_1, R_2 \geq 0$ the triple

(R_0, R_1, R_2) is achievable and (17) holds. Notice that $H(T_1, T_2|Y)$ is the minimum compression rate required for a single receiver knowing $\{Y_i\}$ to losslessly reconstruct both $\{T_{1,i}\}$ and $\{T_{2,i}\}$.

Corollary 2.1: The minimum common rate $R^\star_{0,\text{SI}}$ is

$$R^\star_{0,\text{SI}} = C(T_1; T_2|Y). \tag{18}$$

III. On Relevant Common Information

We present two operational meanings of the relevant common information in Definition 2.

A. The Strong-Coordination Problem

Consider the scenario of Figure 3, where $\Gamma(s|t_1, t_2)$ denotes the channel law of a discrete memoryless multiple-access channel with finite input alphabets $\mathcal{T}_1$ and $\mathcal{T}_2$ and finite output alphabet $\mathcal{S}$. For a given blocklength $n > 0$, we let the common randomness J be uniformly distributed over the index set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$.

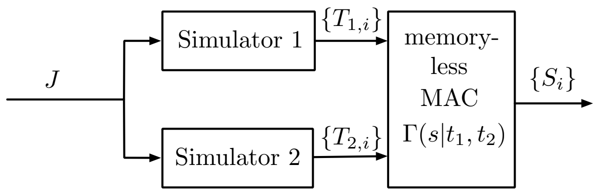

Fig. 3. A remote strong coordination problem.

We say that distribution Q_S over $\mathcal{S}$ can be *remotely strongly-coordinated with rate* R if for each blocklength n there exist simulator functions $\varphi^{(n)}_{\text{Rel},1}$ and $\varphi^{(n)}_{\text{Rel},2}$ of appropriate domains, so that when the sequences

$$T_1^n := \varphi^{(n)}_{\text{Rel},1}(J, \Theta_1) \tag{19}$$

$$T_2^n := \varphi^{(n)}_{\text{Rel},2}(J, \Theta_2) \tag{20}$$

are fed to the MAC $\Gamma(s|t_1, t_2)$, the probability distribution P_{S^n} of the produced output S^n satisfies

$$\left\| P_{S^n} - Q_S^{\otimes n} \right\|_{\text{TV}} \to 0 \quad \text{as} \quad n \to \infty. \tag{21}$$

Here Θ_1, Θ_2, and J are independent (with Θ_1 and Θ_2 accounting for local randomness) and $Q_S^{\otimes n}$ denotes the n-fold product distribution of Q_S.

Theorem 3: The distribution Q_S can be remotely strongly coordinated with rate R if, and only if,

$$R \geq I(S; W) \tag{22}$$

for some auxiliary random variables T_1, T_2, W that satisfy the Markov chains

$$T_1 \to \quad W \quad \to T_2 \tag{23a}$$

$$W \to \quad (T_1, T_2) \quad \to S \tag{23b}$$

and where the conditional probability distribution of S given $T_1 = t_1$ and $T_2 = t_2$ is given by $\Gamma(\cdot|t_1, t_2)$.

Proof: See Section IV. ∎

Corollary 3.1: Let $R^\star_{\text{Rel}}$ be the minimum rate R so that Q_S can be remotely strongly coordinated at the output of the MAC $\Gamma(s|t_1, t_2)$. We find

$$R^\star_{\text{Rel}} = \min_{T_1, T_2} C(T_1; T_2 \to S), \tag{24}$$

where the minimum is taken over all T_1, T_2 that when passed to the MAC $\Gamma(s|t_1, t_2)$ produce an $S \sim Q_S$.

B. The Weak-Coordination Problem

Consider the Gray-Wyner problem in Figure 4, where the source $\{S_i\}$ is IID according to a given distribution Q_S over a finite alphabet $\mathcal{S}$.

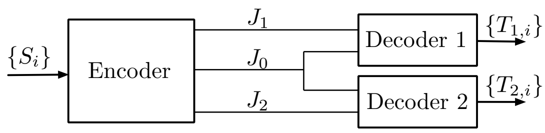

Fig. 4. Gray-Wyner weak-coordination problem.

For a given blocklength n, let $S^n := (S_1, \ldots, S_n)$. The encoder produces three indices

$$(J_0, J_1, J_2) = \phi^{(n)}_{\text{Rel}}(S^n), \tag{25}$$

for some encoding function

$$\phi^{(n)}_{\text{Rel}} \colon \times \mathcal{S}^n \\ \to \{1, \ldots, \lfloor 2^{nR_0} \rfloor\} \times \{1, \ldots, \lfloor 2^{nR_1} \rfloor\} \times \{1, \ldots, \lfloor 2^{nR_2} \rfloor\}. \tag{26}$$

Indices J_0 and J_1 are fed to Decoder 1 and Indices J_0 and J_2 to Decoder 2. The two decoders produce reconstruction sequences

$$T_1^n = \psi^{(n)}_{\text{Rel},1}(J_0, J_1) \tag{27}$$

$$T_2^n = \psi^{(n)}_{\text{Rel},2}(J_0, J_2). \tag{28}$$

We say that the joint distribution $Q_{ST_1T_2}$ *can be weakly-coordinated over a Gray-Wyner network with rates* (R_0, R_1, R_2) if for each blocklength $n > 0$ there exists an encoding function $\phi^{(n)}_{\text{Rel}}$ as in (26) and decoding functions $\psi^{(n)}_{\text{Rel},1}$ and $\psi^{(n)}_{\text{Rel},2}$ of appropriate domains, so that:

$$\|\pi(S^n, T_1^n, T_2^n) - Q_{ST_1T_2}\|_{\text{TV}} \to 0, \qquad \text{as } n \to \infty, \tag{29}$$

where convergence is in probability and where $\pi(S^n, T_1^n, T_2^n)$ denotes the joint type of the tuple (S^n, T_1^n, T_2^n).

Theorem 4: The joint distribution $Q_{ST_1T_2}$ can be weakly coordinated over a Gray-Wyner network with rates (R_0, R_1, R_2) if there exists an auxiliary random variable W such that

$$R_0 \geq I(S; W) \tag{30a}$$

$$R_0 + R_1 \geq I(S; T_1, W) \tag{30b}$$

$$R_0 + R_2 \geq I(S; T_2, W) \tag{30c}$$

$$R_0 + R_1 + R_2 \geq I(S; T_1, T_2, W) + I(T_1; T_2|W). \tag{30d}$$

Proof: Omitted. ∎

Cuff, Permuter, and Cover had considered this problem in the special case without common rate, see [3, Theorem 7]. Both with and without common rate, a matching converse result is missing.

Let $R^\star_{0,\text{Rel}}$ be the minimum common rate $R_0 > 0$ so that for some rates (R_1, R_2) satisfying

$$R_0 + R_1 + R_2 = I(S; T_1, T_2), \tag{31}$$

the distribution $Q_{ST_1T_2}$ can be weakly-coordinated over a Gray-Wyner network with these rates (R_0, R_1, R_2).

Notice that $I(S; T_1, T_2)$ is the smallest rate required to weakly coordinate reconstruction sequences $\{T_{1,i}, T_{2,i}\}$ with the source $\{S_i\}$ according to a joint target distribution $Q_{ST_1T_2}$ when there is only a single decoder that produces both $\{T_{1,i}\}$ and $\{T_{2,i}\}$.

From Theorem 4 we obtain the following.

Corollary 4.1: The minimum common rate $R^\star_{\text{Rel},0}$ is at most equal to the common information of T_1 and T_2 relevant to S in (3):

$$R^\star_{0,\text{Rel}} \leq C(T_1; T_2 \to S). \tag{32}$$

Proof: Fix $Q_{ST_1T_2}$ and consider a rate-tuple (R_0, R_1, R_2) satisfying the constraints in Theorem 4. By the sum-rate constraint (30d) we can have equality in

$$R_0 + R_1 + R_2 = I(S; T_1, T_2), \tag{33}$$

only if for some auxiliary W

$$I(S; W|T_1, T_2) = 0 \qquad \text{and} \qquad I(T_1; T_2|W) = 0.$$

That is, only if for some W the following two Markov chains hold:

$$S \to (T_1, T_2) \to W \tag{34a}$$
$$T_1 \to W \to T_2. \tag{34b}$$

Let W satisfy (34), and set

$$R_0 = I(W; S) \tag{35}$$
$$R_1 = I(T_1; S|W) \tag{36}$$
$$R_2 = I(T_1; S|W). \tag{37}$$

This tuple satisfies all four constraints in Theorem 4 because of the Markov chains (34). By minimizing over all legitimate choices of W, we obtain the desired upper bound on $R^\star_{\text{Rel},0}$.

It can also be shown that no better upper bound on $R^\star_{\text{Rel},0}$ can be obtained from Theorem 4. The relevant common information $C(T_1; T_2 \to S)$ only represents an upper bound on $R^\star_{\text{Rel},0}$, because we are missing a converse proof to Theorem 4. ■

IV. Proof of Theorem 3

We first prove the achievability part, followed by the converse part.

A. Achievability

A main ingredient in the achievability proof is the following lemma from [2].

Lemma 5 (Lemma 19 in [2]): Fix a joint distribution Q_{AB} over the product alphabet $\mathcal{A} \times \mathcal{B}$. Denote its marginal and conditional marginal on B by Q_B and by $Q_{B|A}$. Fix $\delta > 0$ and $R > I(A; B)$, where this mutual information is calculated for $(A, B) \sim Q_{AB}$.

For all sufficiently large n, there is a subset $\{a^n(j)\}_{j=1}^{\lfloor 2^{nR} \rfloor}$ of $\mathcal{A}^n$ such that the average distribution

$$P_{B^n}(b^n) \triangleq \frac{1}{\lfloor 2^{nR} \rfloor} \sum_{j=1}^{\lfloor 2^{nR} \rfloor} Q_{B|A}^{\otimes n}\big(b^n|a^n(j)\big), \quad b^n \in \mathcal{B}^n, \tag{38}$$

(where $Q_{B|A}^{\otimes n}$ denotes the n-fold product of $Q_{B|A}$) is close to $P_B^{\otimes n}(b^n)$ in terms of total variational distance:

$$\big\| P_{B^n} - Q_B^{\otimes n} \big\|_{\text{TV}} \leq \delta. \tag{39}$$

We now prove feasibility of Theorem 3. Fix a rate $R > 0$ and a joint distribution $Q_{WST_1T_2}$ so that $(W, S, T_1, T_2) \sim Q_{WST_1T_2}$ satisfy the Markov chains (23) and

$$R > I(W; S). \tag{40}$$

Consider the construction in Figure 5 where the index J is uniform over the set $\big\{1, \ldots, \lfloor 2^{nR} \rfloor\big\}$ and the n-length sequences $\{w^n(j)\}_{j=1}^{\lfloor 2^{nR} \rfloor}$ are chosen as explained in Lemma 5 above. We feed the random n-length sequence $w^n(J)$ to a discrete memoryless channel $Q_{S|W}$, and denote the output sequence of this channel by S^n. By Lemma 5, the produced S^n satisfies (21) whenever (40) holds.

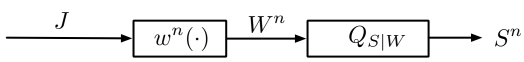

Fig. 5. A simple construction generating the desired random output sequence S^n. The set $\{w^n(\cdot)\}$ needs to be chosen to satisfy the assumptions in Lemma 5 when Q_{AB} is replaced by Q_{WS}.

Since we chose $Q_{WST_1T_2}$ to satisfy Markov chain (23a), the construction in the following Figure 6 is equivalent to the one in Figure 5.

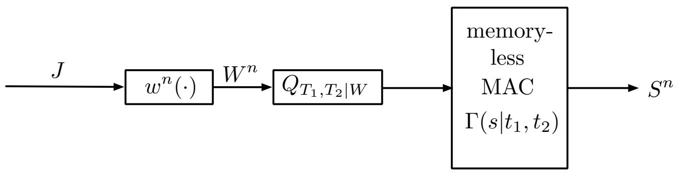

Fig. 6. This construction is equivalent to the one in Figure 5 because of the Markov chain $S \to (T_1, T_2) \to W$.

Since $Q_{WST_1T_2}$ also satisfies Markov chain (23b), the construction in Figure 6 is further equivalent to the construction in Figure 7. The construction in Figure 7 is of the form demanded in the problem setup, and since the generated output sequence satisfies (21), the construction is a solution to our problem. Considering the assumptions we made on R and on the distribution $Q_{WST_1T_2}$, this concludes the proof.

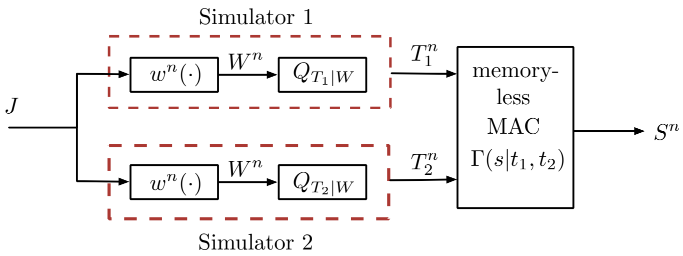

Fig. 7. This construction is equivalent to the one in Figure 6 because of the Markov chain $S \to (T_1, T_2) \to W$. Both simulators use the same codebook $\{w^n(\cdot)\}$. Simulator k feeds the codeword $W^n = w^n(J)$ to a DMC $P_{T_k|W}$.

B. Infeasibility Proof

We will need the following lemmas 20 and 21 from [2]:

Lemma 6 (Lemma 20 in [2]): Let Q_A be a probability law over a finite alphabet $\mathcal{A}$, and let A^n be a random sequence over $\mathcal{A}^n$. If

$$\|P_{A^n} - Q_A^{\otimes n}\|_{\mathrm{TV}} < \epsilon, \tag{41}$$

for some $1/2 > \epsilon > 0$, then

$$\frac{1}{n}\sum_{k=1}^{n} I(A_k; A^{k-1}) \le 2\epsilon\left(\log|\mathcal{A}| + \log\frac{1}{\epsilon}\right). \tag{42}$$

Lemma 7 (Lemma 21 in [2]): Let Q_A be a probability law over a finite alphabet $\mathcal{A}$, and let A^n be a random sequence over $\mathcal{A}^n$. Assume that

$$\|P_{A^n} - Q_A^{\otimes n}\|_{\mathrm{TV}} < \epsilon, \tag{43}$$

for some $1/2 > \epsilon > 0$. Also, let the time-sharing random variable U be uniform over $\{1, \dots, n\}$ and independent of the tuple A^n.

Then,

$$I(A_U; U) \le 2\epsilon\left(\log|\mathcal{A}| + \log\frac{1}{\epsilon}\right). \tag{44}$$

We now prove the infeasibility result in the theorem. Consider a sequence of simulator functions $\{\varphi^{(n)}_{\mathrm{Rel},1}\}_{n=1}^{\infty}$ and $\{\varphi^{(n)}_{\mathrm{Rel},2}\}_{n=1}^{\infty}$ for which the induced MAC outputs $\{S^n\}_{n=1}^{\infty}$ satisfy (21) for a given distribution Q_S.

Fix a large positive integer n, and let $\epsilon_n \in (0, 1/2)$ satisfy

$$\|P_{S^n} - Q_S^{\otimes n}\|_{\mathrm{TV}} < \epsilon_n. \tag{45}$$

Let T_1^n and T_2 be the sequences produced by the chosen $\varphi^{(n)}_{\mathrm{Rel},1}$ and $\varphi^{(n)}_{\mathrm{Rel},2}$, and S^n the corresponding sequence of MAC outputs. Also, let U be uniform over $\{1, \dots, n\}$ independent of J, T_1^n, T_2^n, S^n. Define $S \triangleq S_U$ and $W \triangleq (J, U)$. Then,

$$\begin{aligned}
R = \frac{1}{n}H(J) &\ge \frac{1}{n}I(J; S^n)\\
&\ge \frac{1}{n}H(S^n) - \frac{1}{n}\sum_{k=1}^{n} H(S_k|J)\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[H(S_k|S^{k-1}) - H(S_k|J)\right]\\
&= \frac{1}{n}\sum_{k=1}^{n}\left[H(S_k) - I(S_k; S^{k-1}) - H(S_k|J)\right]\\
&\ge \frac{1}{n}\sum_{k=1}^{n}\Bigg[H(S_k) - 2\epsilon_n\left(\log|\mathcal{S}| + \log\frac{1}{\epsilon_n}\right)\\
&\qquad\qquad - H(S_k|J)\Bigg]\\
&= I(S_U; J|U) - 2\epsilon_n\left(\log|\mathcal{S}| + \log\frac{1}{\epsilon_n}\right)\\
&\ge I(S_U; J, U) - 4\epsilon_n\left(\log|\mathcal{S}| + \log\frac{1}{\epsilon_n}\right),\\
&= I(S; W) - 4\epsilon_n\left(\log|\mathcal{S}| + \log\frac{1}{\epsilon_n}\right).
\end{aligned} \tag{46}$$

where the second inequality follows because conditioning can only reduce entropy; the third inequality by Lemma 6; and the fourth inequality by Lemma 7.

Since the considered sequence of simulators achieves the goal in (21), we can choose the sequence ϵ_n tending to 0 as $n \to \infty$. Therefore,

$$R \ge I(S; W). \tag{47}$$

Notice that by the structure of the problem's setup in Figure 3,

$$T_{1,k} \to J \to T_{2,k} \tag{48}$$

and

$$J \to (T_{1,k}, T_{2,k}) \to S_k. \tag{49}$$

Let $T_1 \triangleq T_{1,U}$ and $T_2 \triangleq T_{2,U}$. Since U is independent of (T_1^n, T_2^n, S^n, J), the above two Markov chains also imply

$$T_1 \to W \to T_2 \tag{50}$$

and

$$W \to (T_1, T_2) \to S. \tag{51}$$

Combined with (47), these two Markov chains conclude the proof of the infeasibility part.

Acknowledgement

The authors thank Ligong Wang for insightful discussions.

References

[1] A. D. Wyner, “The common information of two dependent random variables,” *IEEE Trans. on Inf. Theory*, vol. 21, no. 2, pp. 163–179, March 1975.
[2] P. W. Cuff, “Communication in networks for coordinating behavior,” PhD thesis, Stanford university, July 2009.
[3] P. W. Cuff, H. H. Permuter, and T. Cover, “Coordination capacity,” *IEEE Trans. on Inf. Theory*, vol 56, no. 9, pp. 4181–4206, Sep., 2010.
[4] K. B. Viswanatha, E. Akyol, and K. Rose, “The lossy common information of correlated sources,” *IEEE Trans. on Inf. Theory*, vol 60, no. 6, pp. 3238–3253, June, 2014