Conditional and Relevant Common Information

†Amos Lapidoth and Michèlle 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 reconstructions 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

$$C(T_1; T_2) \triangleq \min_{W: T_1 \rightarrow W \rightarrow T_2} I(T_1, T_2; W),$$

(1)

where $X \rightarrow Y \rightarrow Z$ 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 $\{(T_1, i, T_2, i)\}$ IID $Q_{T_1; T_2}$ 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 simultaneously produce outputs of joint distribution $Q_{T_1; T_2}$ (Fig. 3 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 $C(T_1; T_2 \rightarrow S)$, 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

$$C(T_1; T_2 | Y) \triangleq \min_{W: T_1 \rightarrow (W,Y) \rightarrow T_2} I(T_1, T_2; W | Y).$$

(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

$$C(T_1; T_2 \rightarrow S) \triangleq \min_{W: T_1 \rightarrow W \rightarrow T_2 \rightarrow S} I(S; W).$$

(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:

$$Y = \emptyset \implies C(T_1; T_2 | Y) = C(T_1; T_2),$$

(4)

$$S = (T_1, T_2) \implies C(T_1; T_2 \rightarrow S) = C(T_1; T_2).$$

(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 the outputs of two terminals according to a target distribution $Q_{T_1; T_2}$ 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 the receivers’ reconstructions $\{T_1, i\}$ and $\{T_2, i\}$ with each other and with the source $\{S_i\}$ according to a target distribution $Q_{T_1; T_2; S}$, 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 [2]. However, in (4), the minimization is not only over the auxilary 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
Let \( Y^n := (Y_1, \ldots, Y_n) \), and let the common randomness \( J \) be uniformly distributed over the index set \( \{1, \ldots, [2^n R]\} \).

We say that a joint distribution \( Q_{T_1, T_2, Y} \) over a finite product set \( T_1 \times T_2 \times S \) can be strongly coordinated with rate \( R > 0 \) and SI \( \{Y_i\} \) if, for each blocklength \( n > 0 \), there exist functions \( \psi^{(n)}_{SI1} \) and \( \psi^{(n)}_{SI2} \) of appropriate domains, for which the sequences

\[
\begin{align*}
T^n_1 &:= \psi^{(n)}_{SI1}(J, Y^n, \Theta_1) \\
T^n_2 &:= \psi^{(n)}_{SI2}(J, Y^n, \Theta_2)
\end{align*}
\]

satisfy

\[
\| P_{T^n_1 T^n_2 Y^n} - Q_{T^n_1 T^n_2 Y}^{\otimes n} \|_{TV} \to 0 \quad \text{as} \quad n \to \infty. \tag{8}
\]

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 \( Q_{T_1, T_2, Y} \) can be strongly coordinated with rate \( R \) and SI \( \{Y_i\} \) if, and only if,

\[
R \geq C(T_1; T_2|Y). \tag{11}
\]

**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 \( Q_{T_1, T_2, Y} \) over a finite product alphabet \( T_1 \times T_2 \times Y \). For a given blocklength \( n > 0 \), define \( T^n_1 := (T_{1,1}, \ldots, T_{1,n}) \), \( T^n_2 := (T_{2,1}, \ldots, T_{2,n}) \) and \( Y^n := (Y_1, \ldots, Y_n) \). The encoder observes all three sequences \( T^n_1, T^n_2, Y^n \) and produces the indices \( J_0, J_1, J_2 \)

\[
(J_0, J_1, J_2) = \phi^{(n)}_{SI}(T^n_1, T^n_2, Y^n), \tag{12}
\]

for some encoding function

\[
\phi^{(n)}_{SI} : T^n_1 \times T^n_2 \times Y^n \to \{1, \ldots, [2^n R_0]\} \times \{1, \ldots, [2^n R_1]\} \times \{1, \ldots, [2^n R_2]\}. \tag{13}
\]

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

\[
T^n_1 = \psi^{(n)}_{SI1}(J_0, J_1, Y^n), \tag{14}
\]

\[
T^n_2 = \psi^{(n)}_{SI2}(J_0, J_2, Y^n), \tag{15}
\]

for some decoding functions \( \psi^{(n)}_{SI1} \) and \( \psi^{(n)}_{SI2} \) 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 \( \phi^{(n)}_{SI} \) as in (12) and decoding functions \( \psi^{(n)}_{SI1} \) and \( \psi^{(n)}_{SI2} \) of appropriate domains, so that:

\[
\lim_{n \to \infty} \Pr((T^n_1, T^n_2) \neq (\hat{T}^n_1, \hat{T}^n_2)) = 0. \tag{16}
\]

**Theorem 2:** Given a joint distribution \( Q_{T_1, T_2, Y} \), a rate-triple \((R_0, R_1, R_2)\) is achievable if, and only if, there exists an auxiliary random variable \( W \) such that

\[
\begin{align*}
R_0 &\geq I(W; T_1, T_2|Y) \tag{16a} \\
R_1 &\geq H(T_1|W, Y) \tag{16b} \\
R_2 &\geq H(T_2|W, Y). \tag{16c}
\end{align*}
\]

**Proof:** Omitted.

Let \( R^*_{0, SI} \) 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) if

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

The rate \( R^*_{0, SI} \) 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...
(R₀, R₁, R₂) is achievable and \([17]\) holds. Notice that \(H(T₁, T₂|Y)\) is the minimum compression rate required for a single receiver knowing \(\{Y_i\}\) to losslessly reconstruct both \(\{T₁,i\}\) and \(\{T₂,i\}\).

**Corollary 2.1:** The minimum common rate \(R_{0,SI}^*\) is
\[
R_{0,SI}^* = C(T₁; T₂|Y).
\]  

### 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 \(\Gamma(s|t₁, t₂)\) denotes the channel law of a discrete memoryless multiple-access channel with finite input alphabets \(T₁\) and \(T₂\) and finite output alphabet \(S\). For a given blocklength \(n > 0\), we let the common randomness \(J\) be uniformly distributed over the index set \(\{1, \ldots, [2ⁿR]\}\).

We say that distribution \(Q_S\) over \(S\) can be remotely strongly-coordinated with rate \(R\) if for each blocklength \(n\) there exist simulator functions \(ϕ_{Rel,1}^{(n)}\) and \(ϕ_{Rel,2}^{(n)}\) of appropriate domains, so that when the sequences
\[
T₁^n := ϕ_{Rel,1}^{(n)}(J, Θ₁)
\]
\[
T₂^n := ϕ_{Rel,2}^{(n)}(J, Θ₂)
\]
are fed to the MAC \(\Gamma(s|t₁, t₂)\), the probability distribution \(Pₜ\) of the produced output \(S^n\) satisfies
\[
∥Pₜ - Q_S^n∥_{TV} → 0 \quad \text{as} \quad n → ∞.
\]  

Here \(Θ₁, Θ₂,\) and \(J\) are independent (with \(Θ₁\) and \(Θ₂\) accounting for local randomness) and \(Q_S^n\) 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,
\[
R ≥ I(S; W)
\]  

for some auxiliary random variables \(T₁, T₂, W\) that satisfy the Markov chains
\[
T₁ → W → T₂ \quad (23a)
\]
\[
W → (T₁, T₂) → S \quad (23b)
\]
and where the conditional probability distribution of \(S\) given \(T₁ = t₁\) and \(T₂ = t₂\) is given by \(Γ(·|t₁, t₂)\).

**Proof:** See Section IV

**Corollary 3.1:** Let \(R_{rel}\) be the minimum rate \(R\) so that \(Qₜ\) can be remotely strongly coordinated at the output of the MAC \(Γ(s|t₁, t₂)\). We find
\[
R_{rel}^* = \min_{T₁, T₂} C(T₁; T₂ → S),
\]  

where the minimum is taken over all \(T₁, T₂\) that when passed to the MAC \(Γ(s|t₁, t₂)\) 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 \(S\).

![Gray-Wyner weak-coordination problem](image)

For a given blocklength \(n\), let \(S^n := (S₁, \ldots, S_n)\). The encoder produces three indices
\[
(J₀, J₁, J₂) = ϕ_{Rel}^{(n)}(S^n),
\]  

for some encoding function
\[
ϕ_{Rel} : × S^n → \{1, \ldots, [2ⁿR₀]\} × \{1, \ldots, [2ⁿR₁]\} × \{1, \ldots, [2ⁿR₂]\}.
\]  

Indices \(J₀\) and \(J₁\) are fed to Decoder 1 and Indices \(J₀\) and \(J₂\) to Decoder 2. The two decoders produce reconstruction sequences
\[
T₁^n = ϕ_{Rel,1}^{(n)}(J₀, J₁)
\]
\[
T₂^n = ϕ_{Rel,2}^{(n)}(J₀, J₂).
\]

We say that the joint distribution \(Q_{ST₁T₂}\) can be weakly-coordinated over a Gray-Wyner network with rates \((R₀, R₁, R₂)\) if for each blocklength \(n > 0\) there exists an encoding function \(ϕ_{Rel}^{(n)}\) as in (26) and decoding functions \(ψ_{Rel,1}^{(n)}\) and \(ψ_{Rel,2}^{(n)}\) of appropriate domains, so that:
\[
∥π(S^n, T₁^n, T₂^n) - Q_{ST₁T₂}∥_{TV} → 0 \quad \text{as} \quad n → ∞,
\]  

where convergence is in probability and where \(π(S^n, T₁^n, T₂^n)\) denotes the joint type of the tuple \((S^n, T₁^n, T₂^n)\).

**Theorem 4:** The joint distribution \(Q_{ST₁T₂}\) can be weakly-coordinated over a Gray-Wyner network with rates \((R₀, R₁, R₂)\) if there exists an auxiliary random variable \(W\) such that
\[
R₀ ≥ I(S; W) \quad (30a)
\]
\[
R₀ + R₁ ≥ I(S; T₁, W) \quad (30b)
\]
\[
R₀ + R₂ ≥ I(S; T₂, W) \quad (30c)
\]
\[
R₀ + R₁ + R₂ ≥ I(S; T₁, T₂, W) + I(T₁; T₂|W) \quad (30d)
\]

**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 \( R_{0,\text{Rel}} \) be the minimum common rate \( R_0 > 0 \) so that for some rates \((R_1, R_2)\) satisfying
\[
R_0 + R_1 + R_2 = I(S; T_1, T_2),
\]
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 \( R^*_{\text{Rel,0}} \) is at most equal to the common information of \( T_1 \) and \( T_2 \) relevant to \( S \) in [3]:
\[
R^*_{\text{Rel,0}} \leq C(T_1; T_2 \rightarrow S).
\]

**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
\[
R_0 + R_1 + R_2 = I(S; T_1, T_2),
\]
only if for some auxiliary \( W \)
\[
I(S; W|T_1, T_2) = 0 \quad \text{and} \quad I(T_1; T_2|W) = 0.
\]
That is, only if for some \( W \) the following two Markov chains hold:
\[
S \rightarrow (T_1, T_2) \rightarrow W \tag{34a}
\]
\[
T_1 \rightarrow W \rightarrow T_2. \tag{34b}
\]
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 \( R^*_{\text{Rel,0}} \).

It can also be shown that no better upper bound on \( R^*_{\text{Rel,0}} \) can be obtained from Theorem 4. The relevant common information \( C(T_1; T_2 \rightarrow S) \) only represents an upper bound on \( R^*_{\text{Rel,0}} \), 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 [3].

**Lemma 5 (Lemma 19 in [2]):** Fix a joint distribution \( Q_{AB} \) over the product alphabet \( A \times B \). 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 \( \{a^n(j)\}_{j=1}^{[2^nR]} \) of \( A^n \) such that the average distribution
\[
P^B_{\text{avg}}(b^n) \triangleq \frac{1}{[2^nR]^n} \sum_{j=1}^{[2^nR]} Q^n_{B|A}(b^n|a^n(j)), \quad b^n \in B^n, \tag{38}
\]
where \( Q^n_{B|A} \) denotes the \( n \)-fold product of \( Q_{B|A} \) is close to \( P^B_{\text{avg}}(b^n) \) in terms of total variational distance:
\[
\|P^B_{\text{avg}} - Q^n_{B|A}\|_{TV} \leq \delta. \tag{39}
\]

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 \( \{1, \ldots, [2^nR]\} \) and the \( n \)-length sequences \( \{w^n(j)\}_{j=1}^{[2^nR]} \) 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.

\[
J \xrightarrow{} w^n(\cdot) \xrightarrow{} W^n \xrightarrow{} Q_{S|W} \xrightarrow{} S^n
\]

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.

\[
J \xrightarrow{} u^n(\cdot) \xrightarrow{} W^n \xrightarrow{} Q^n_{T_1,T_2|W} \xrightarrow{} \text{memory-less MAC} \xrightarrow{} \Gamma(s(t_1,t_2)) \xrightarrow{} S^n
\]

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

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 (24), 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.

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

\[
\|P_{A^n} - Q_A^{\otimes n}\|_{TV} < \epsilon,
\]

for some \( 1/2 > \epsilon > 0 \), then

\[
\frac{1}{n} \sum_{k=1}^{n} I(A_k; A^{k-1}) \leq 2\epsilon \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon} \right).
\]

**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

\[
\|P_{A^n} - Q_A^{\otimes n}\|_{TV} < \epsilon,
\]

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

Then,

\[
I(A_U; U) \leq 2\epsilon \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon} \right).
\]

We now prove the infeasibility result in the theorem. Consider a sequence of simulator functions \( \{\varphi_{\text{Rel,1}}^{(n)}\}_{n=1}^{\infty} \) and \( \{\varphi_{\text{Rel,2}}^{(n)}\}_{n=1}^{\infty} \), for which the induced MAC outputs \( \{S^n\}_{n=1}^{\infty} \) satisfy (24) for a given distribution \( Q_S \).

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

\[
\|P_{S^n} - Q_S^{\otimes n}\|_{TV} < \epsilon_n.
\]

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

\[
R = \frac{1}{n} H(J) \geq \frac{1}{n} I(J; S^n) \geq \frac{1}{n} H(S^n) - \frac{1}{n} \sum_{k=1}^{n} H(S_k | J)
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} [H(S_k | S^{k-1}) - H(S_k | J)]
\]

\[
= \frac{1}{n} \sum_{k=1}^{n} [H(S_k) - I(S_k; S^{k-1}) - H(S_k | J)]
\]

\[
\geq \frac{1}{n} \sum_{k=1}^{n} [H(S_k) - 2\epsilon_n \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon_n} \right) - H(S_k | J)]
\]

\[
= I(S_U; J | U) - 2\epsilon_n \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon_n} \right)
\]

\[
\geq I(S_U; J | U) - 4\epsilon_n \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon_n} \right),
\]

\[
= I(S; W) - 4\epsilon_n \left( \log |\mathcal{A}| + \log \frac{1}{\epsilon_n} \right).
\]

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 (24), we can choose the sequence \( \epsilon_n \) tending to 0 as \( n \to \infty \). Therefore,

\[
R \geq I(S; W).
\]

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

\[
T_{1,k} \rightarrow J \rightarrow T_{2,k}
\]

and

\[
J \rightarrow (T_{1,k}, T_{2,k}) \rightarrow S_k.
\]

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

\[
T_1 \rightarrow W \rightarrow T_2
\]

and

\[
W \rightarrow (T_1, T_2) \rightarrow S.
\]

Combined with (27), these two Markov chains conclude the proof of the infeasibility part.

**Acknowledgement**

The authors thank Ligong Wang for insightful discussions.

**References**


