



HAL
open science

Coding Theorems for Empirical Coordination

Mael Le Treust

► **To cite this version:**

Mael Le Treust. Coding Theorems for Empirical Coordination: Technical Report. 2015. hal-01865569

HAL Id: hal-01865569

<https://hal.science/hal-01865569>

Preprint submitted on 30 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coding Theorems for Empirical Coordination – Technical Report

Maël Le Treust

ETIS, UMR 8051 / ENSEA, Université Cergy-Pontoise, CNRS,

6, avenue du Ponceau,

95014 CERGY-PONTOISE CEDEX,

FRANCE

Email: mael.le-treust@ensea.fr

CONTENTS

I Non-Causal Encoding and Decoding	3
I-A Achievability Proof	5
I-B Converse Proof	6
II Perfect Channel	8
II-A Achievability Proof	9
II-B Converse Proof	11
III Lossless Decoding	14
III-A Achievability Proof	16
III-B Converse Proof	18
IV Separation between Source and Channel	22
IV-A Achievability Proof	24
IV-B Converse Proof	26
V Causal Decoding	30
V-A Achievability Proof	32
V-B Converse Proof	36

VI Causal Encoding	38
VI-A Achievability Proof	40
VI-B Converse Proof	46
VII Strictly Causal Decoding	49
VII-A Achievability Proof	51
VII-B Converse Proof	54
VIII Strictly Causal Encoding	56
VIII-A Achievability Proof	58
VIII-B Converse Proof	60
Appendix	65
References	76

The probability distribution $\mathcal{Q} \in \mathbb{Q}$ decomposes as follows:

$$\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(w_1, w_2|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_1, w_2).$$

The supports of the auxiliary random variables (W_1, W_2) are bounded by $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq (|\mathcal{B}| + 1) \cdot (|\mathcal{B}| + 2)$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark I.2 The mutual informations in equation (2) are continuous over the set of probability distributions \mathbb{Q} . Moreover, \mathbb{Q} is compact since the supports of the auxiliary random variables (W_1, W_2) are finite and \mathbb{Q} has equality constraints. As mentioned in [8] pp. 7083 and in [2] pp. 9, we can consider the maximum instead of the supremum in equation (2).

Remark I.3 The achievability result of Theorem I.1 without state informations S and Z was already stated in [1], with a unique auxiliary random variable $W = (W_1, W_2)$.

Remark I.4 As mentioned in Theorem I.1 1), probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ should satisfy the marginal distributions over the source $\mathcal{P}_{\text{usz}}(u, s, z)$, the channel $\mathcal{T}(y|x, s)$ and the Markov chains representing the network topology. If a probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ does not decomposes with $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$, then the error probability can not converge to zero and this probability distribution is not achievable. This remark is valid for all coding theorems presented in this document.

A. Achievability Proof

Denote by $\mathcal{Q}(u, s, z, x, w_1, w_2, y, v) \in \mathbb{Q}$ the joint probability distribution that achieves the maximum in equation (2). There exists $\delta > 0$ and rate $R \geq 0$ such that :

$$R \geq I(W_1, W_2; U, S) + \delta, \quad (4)$$

$$R \leq I(W_1, W_2; Y, Z) - \delta. \quad (5)$$

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ pairs of sequences $(W_1^n(m), W_2^n(m))$ with index $m \in \mathcal{M}$ drawn from the i.i.d. marginal probability distribution $\mathcal{Q}_{w_1 w_2}^{\otimes n}$.
- *Encoding function.* The encoder observes the sequences of source symbols $U^n \in \mathcal{U}^n$ and state symbols $S^n \in \mathcal{S}^n$. It finds the index $m \in \mathcal{M}$ such that the sequences $(U^n, S^n, W_1^n(m), W_2^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence X^n drawn from the conditional probability distribution $\mathcal{Q}_{x|usw_1 w_2}^{\otimes n}$ depending on sequences $(U^n, S^n, W_1^n(m), W_2^n(m))$.
- *Decoding function.* The decoder observes the pair of sequences (Y^n, Z^n) and finds the index $m \in \mathcal{M}$ such that the sequences $(Y^n, Z^n, W_1^n(m), W_2^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Decoder returns the sequence V^n drawn from the conditional probability distribution $\mathcal{Q}_{v|yzw_1 w_2}^{\otimes n}$ depending on sequences $(Y^n, Z^n, W_1^n(m), W_2^n(m))$.

From the properties of typical sequences, packing and covering Lemmas stated in [9] pp. 27, 46 and 208, equations (4), (5), imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (6)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, (U^n, S^n, W_1^n(m), W_2^n(m)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (7)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } (Y^n, Z^n, W_1^n(m'), W_2^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (8)$$

For all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U^n, S^n, Z^n, W_1^n(m), W_2^n(m), X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{usz}(u, s, z) \otimes \mathcal{Q}(x, w_1, w_2|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_1, w_2)$ with probability more than $1 - 3\varepsilon$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem I.1.

B. Converse Proof

We introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (9)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, v)$, *i.e.* for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ is small. We have equations:

$$\begin{aligned} 0 &= \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) \\ &\quad - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}; U_i, S_i, Y_i, Z_i | U_{i+1}^n, S_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n) \end{aligned} \quad (10)$$

$$\begin{aligned} &\leq \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) \\ &\quad - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}; U_i, S_i | U_{i+1}^n, S_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n) \end{aligned} \quad (11)$$

$$= \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | W_{2,i}) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | W_{1,i}). \quad (12)$$

Equation (10) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (11) comes from the properties of the mutual information.

Equation (12) comes from the introduction of the auxiliary random variables $W_{1,i} = (U_{i+1}^n, S_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n)$ and $W_{2,i} = (Y^{i-1}, Z^{i-1})$. The pair of random variables $(W_{1,i}, W_{2,i})$ satisfy the three Markov Chains that correspond to the set of probability distributions \mathbb{Q} :

$$Y_i \text{---} (X_i, S_i) \text{---} (U_i, Z_i, W_{1,i}, W_{2,i}), \quad (13)$$

$$Z_i \text{---} (U_i, S_i) \text{---} (X_i, Y_i, W_{1,i}, W_{2,i}), \quad (14)$$

$$V_i \text{---} (Y_i, Z_i, W_{1,i}, W_{2,i}) \text{---} (U_i, S_i, X_i). \quad (15)$$

- The first Markov chain comes from the memoryless property of the channel and the fact that Y_i does not belong to $(W_{1,i}, W_{2,i})$.
- The second Markov chain comes from the i.i.d. property of the source and the fact that Z_i does not belong to $(W_{1,i}, W_{2,i})$.
- The third Markov chain comes from the non-causal decoding: V_i is a function of the pair (Y^n, Z^n) that is included in $(Y_i, Z_i, W_{1,i}, W_{2,i})$.

$$0 \leq \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | W_{2,i}) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | W_{1,i})$$

$$\leq n \cdot \left(I(W_{1,T}; Y_T, Z_T | W_{2,T}, T) - I(W_{2,T}; U_T, S_T | W_{1,T}, T) \right) \quad (16)$$

$$\leq n \cdot \left(I(W_{1,T}, T; Y_T, Z_T | W_{2,T}) - I(W_{2,T}; U_T, S_T | W_{1,T}, T) \right) \quad (17)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | W_2) - I(W_2; U_T, S_T | W_1) \right) \quad (18)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_2; U_T, S_T | W_1, E = 0) + \varepsilon \right) \quad (19)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y, Z | W_2) - I(W_2; U, S | W_1) + 2\varepsilon \right). \quad (20)$$

Equation (16) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, Z_T$.

Equation (17) comes from the properties of the mutual information.

Equation (18) comes from identifying W_1 with $(W_{1,T}, T)$ and W_2 with $W_{2,T}$ and taking the maximum over the probability distributions \mathcal{Q} that belong to \mathbb{Q} . This is possible since the random variables $(W_{1,T}, T)$ and $W_{2,T}$ satisfy the Markov chains of the set of probability distributions \mathbb{Q} , as stated in Lemma 7 in the Appendix.

Equation (19) comes from the empirical coordination requirement as stated in Lemma 8. Sequences are not jointly typical with small error probability $\mathcal{P}(E = 1)$.

Equation (20) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, Y_T, V_T) = (u, s, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes the converse proof of Theorem I.1.

II. PERFECT CHANNEL

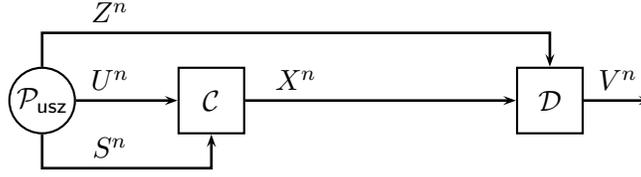


Fig. 2. The perfect channel is defined by $\mathcal{T}_{y|xs} = \mathbb{1}(y|x)$ and the decoding is lossy.

Theorem II.1 (Perfect Channel)

1) The joint probability distribution $\mathcal{Q}(u, s, z, x, v)$ is achievable if and only if it decomposes as follows:

$$\begin{cases} \mathcal{Q}(u, s, z) = \mathcal{P}_{\text{usz}}(u, s, z), \\ Z \text{---} (U, S) \text{---} X, \end{cases} \quad (21)$$

and $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(v|u, s, z, x)$ is achievable.

2) The probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(v|u, s, z, x)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_p} \left(I(W_2; Z|X) + H(X) - I(X, W_2; U, S) \right) > 0, \quad (22)$$

3) The probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(v|u, s, z, x)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_p} \left(I(W_2; Z|X) + H(X) - I(X, W_2; U, S) \right) < 0, \quad (23)$$

where \mathbb{Q}_p is the set of distributions $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{W}_2 \times \mathcal{X} \times \mathcal{V})$ with auxiliary random variable W_2 that satisfies:

$$\begin{cases} \sum_{w_2 \in \mathcal{W}_2} \mathcal{Q}(u, s, z, x, w_2, v) \\ = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(v|u, s, z, x), \\ Z \text{---} (U, S) \text{---} (X, W_2), \\ V \text{---} (X, Z, W_2) \text{---} (U, S). \end{cases}$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}_p$ decomposes as follows:

$$\mathcal{Q}(u, s, z, x, w_2, v) = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(w_2|u, s, x) \otimes \mathcal{Q}(v|x, z, w_2).$$

The support of the auxiliary random variable W_2 is bounded by $|\mathcal{W}_2| \leq |\mathcal{B}| + 1$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark II.2 This result generalizes the coding theorem of Wyner Ziv stated in [2].

A. Achievability Proof

This proof can be obtained from achievability result of Sec. I-A, by replacing random variables W_1 and Y by X . Note that $I(W_2; Z|X) + H(X) - I(X, W_2; U, S) = I(X, W_2; X, Z) - I(X, W_2; U, S)$.

Denote by $\mathcal{Q}(u, s, z, x, w_2, v) \in \mathbb{Q}_p$ the joint probability distribution that achieves the maximum in equation (22). There exists $\delta > 0$ and rate $R \geq 0$ such that :

$$R \geq I(X, W_2; U, S) + \delta, \quad (24)$$

$$R \leq I(X, W_2; X, Z) - \delta. \quad (25)$$

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ pairs of sequences $(X^n(m), W_2^n(m))$ with index $m \in \mathcal{M}$ drawn from the i.i.d. marginal probability distribution $\mathcal{Q}_{xw_2}^{\otimes n}$.
- *Encoding function.* The encoder observes the sequences of source symbols $U^n \in \mathcal{U}^n$ and state symbols $S^n \in \mathcal{S}^n$. It finds the index $m \in \mathcal{M}$ such that the sequences $(U^n, S^n, X^n(m), W_2^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the corresponding sequence $X^n(m)$ through the channel.
- *Decoding function.* The decoder observes the pair of sequences (X^n, Z^n) and finds the index $m \in \mathcal{M}$ such that the sequences $(X^n, Z^n, X^n(m), W_2^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical (for probability distribution $\mathcal{Q}_{xzw_2} \otimes \mathbb{1}_{x|x}$). Decoder returns the sequence V^n drawn from the conditional probability distribution $\mathcal{Q}_{v|xzw_2}^{\otimes n}$ depending on sequences $(X^n, Z^n, W_2^n(m))$.

From the properties of typical sequences, packing and covering Lemmas stated in [9] pp. 27, 46 and 208, equations (24), (25), imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (26)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, (U^n, S^n, X^n(m), W_2^n(m)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (27)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } (X^n, Z^n, X^n(m'), W_2^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (28)$$

For all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U^n, S^n, Z^n, X^n(m), W_2^n(m), V^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{usz}(u, s, z) \otimes \mathcal{Q}(x, w_2|u, s) \otimes \mathcal{Q}(v|x, z, w_2)$ with probability more than $1 - 3\varepsilon$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem II.1.

B. Converse Proof

We introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (29)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, v)$, *i.e.* for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ is small. We have equations:

$$0 = H(X^n, Z^n | E = 0) - I(X^n, Z^n; U^n, S^n | E = 0) - H(X^n, Z^n | U^n, S^n, E = 0) \quad (30)$$

$$\begin{aligned} &\leq \sum_{i=1}^n H(X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^n; U_i, S_i | U_{i+1}^n, S_{i+1}^n, E = 0) \\ &\quad - H(X^n, Z^n | U^n, S^n, E = 0) \end{aligned} \quad (31)$$

$$\begin{aligned} &= \sum_{i=1}^n I(X^n, Z^n, U_{i+1}^n, S_{i+1}^n; X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^n, U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \\ &\quad - H(X^n, Z^n | U^n, S^n, E = 0) + n \cdot \varepsilon \end{aligned} \quad (32)$$

$$\begin{aligned} &= \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \\ &\quad + \sum_{i=1}^n I(Z_i; X_i, Z_i | X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n, E = 0) - \sum_{i=1}^n I(Z_i; U_i, S_i | X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n, E = 0) \\ &\quad - H(X^n, Z^n | U^n, S^n, E = 0) + n \cdot \varepsilon \end{aligned} \quad (33)$$

$$\begin{aligned} &= \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \\ &\quad + \sum_{i=1}^n H(Z_i | X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n, U_i, S_i, E = 0) - H(X^n, Z^n | U^n, S^n, E = 0) + n \cdot \varepsilon \end{aligned} \quad (34)$$

$$\begin{aligned} &\leq \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \\ &\quad + \sum_{i=1}^n H(Z_i | U_i, S_i, E = 0) - H(Z^n | U^n, S^n, E = 0) + n \cdot 2\varepsilon \end{aligned} \quad (35)$$

$$\begin{aligned} &\leq \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; X_i, Z_i | E = 0) - \sum_{i=1}^n I(X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) + n \cdot 3\varepsilon \end{aligned} \quad (36)$$

$$= \sum_{i=1}^n I(X_i, W_{2,i}; X_i, Z_i | E = 0) - I(X_i, W_{2,i}; U_i, S_i | E = 0) + n \cdot 3\varepsilon. \quad (37)$$

Equations (30) and (31) come from the properties of the mutual information.

Equation (32) comes from the i.i.d. property of the information source and Lemma 10 in the Appendix

that implies $\sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \leq n \cdot \varepsilon$.

Equations (33) and (34) come from the properties of the mutual information.

Equation (35) comes from the i.i.d. property of the information source (U, S, Z) and Lemma 10 in the Appendix that implies $\sum_{i=1}^n H(Z_i | X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n, U_i, S_i, E = 0) - \sum_{i=1}^n H(Z_i | U_i, S_i, E = 0) \leq n \cdot \varepsilon$.

Equation (36) comes from the i.i.d. property of the information source (U, S, Z) and Lemma 10 in the Appendix that implies $\sum_{i=1}^n H(Z_i | U_i, S_i, E = 0) - H(Z^n | U^n, S^n, E = 0) \leq n \cdot \varepsilon$.

Equation (37) comes from the introduction of the auxiliary random variable $W_{2,i} = (X^{-i}, Z^{-i}, U_{i+1}^n, S_{i+1}^n)$. The random variable $W_{2,i}$ satisfy the Markov Chains that correspond to the set of probability distributions \mathbb{Q}_p corresponding to the perfect channel:

$$Z_i \text{---} (U_i, S_i) \text{---} (X_i, W_{2,i}), \quad (38)$$

$$V_i \text{---} (X_i, Z_i, W_{2,i}) \text{---} (U_i, S_i). \quad (39)$$

- The first Markov chain comes from i.i.d. property of the source and the fact that Z_i does not belong to $W_{2,i}$.
- The second Markov chain comes from the non-causal decoding: V_i is a function of (X^n, Z^n) that is included in $(X_i, Z_i, W_{2,i}) = (X_i, Z_i, X^{-i}, Z^{-i}, U_{i+1}^n, S_{i+1}^n)$.

$$\begin{aligned} 0 &\leq \sum_{i=1}^n I(X_i, W_{2,i}; X_i, Z_i | E = 0) - I(X_i, W_{2,i}; U_i, S_i | E = 0) + 3n\varepsilon \\ &= n \cdot \left(I(X_T, W_{2,T}; X_T, Z_T | T, E = 0) - I(X_T, W_{2,T}; U_T, S_T | T, E = 0) + 3\varepsilon \right) \end{aligned} \quad (40)$$

$$\leq n \cdot \left(I(X_T, W_{2,T}, T; X_T, Z_T | E = 0) - I(X_T, W_{2,T}, T; U_T, S_T | E = 0) + 3\varepsilon \right) \quad (41)$$

$$\leq n \cdot \max_{\mathbb{Q} \in \mathbb{Q}_p} \left(I(X_T, W_2; X_T, Z_T | E = 0) - I(X_T, W_2; U_T, S_T | E = 0) + 3\varepsilon \right) \quad (42)$$

$$\leq n \cdot \max_{\mathbb{Q} \in \mathbb{Q}_p} \left(I(X, W_2; X, Z) - I(X, W_2; U, S) + 4\varepsilon \right). \quad (43)$$

Equation (40) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{2,T}, X_T, V_T$.

Equation (41) comes from the i.i.d. property of the information source (U, S) and Lemma 11 in the Appendix that implies $I(T; U_T, S_T | E = 0) = 0$.

Equation (42) comes from identifying W_2 with $(W_{2,T}, T)$ and taking the maximum over the probability distributions that belong to \mathbb{Q}_p . This is possible since the random variables $(W_{2,T}, T)$ satisfy the Markov chains of the set of probability distributions \mathbb{Q}_p , as stated in Lemma 7 in the Appendix.

Equation (43) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, V_T) = (u, s, z, x, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes the converse proof of Theorem II.1.

III. LOSSLESS DECODING

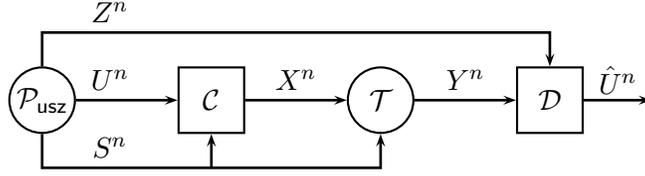


Fig. 3. Noisy channel $\mathcal{T}(y|x, s)$ and lossless decoding $\mathbb{1}(\hat{u}|u)$.

Theorem III.1 (Lossless Decoding)

1) The joint probability distribution $\mathcal{Q}(u, s, z, x, y, \hat{u})$ is achievable if and only if it decomposes as follows:

$$\left\{ \begin{array}{l} \mathcal{Q}(u, s, z) = \mathcal{P}_{\text{usz}}(u, s, z), \\ \mathcal{Q}(y|x, s) = \mathcal{T}(y|x, s), \\ \mathcal{Q}(\hat{u}|u) = \mathbb{1}(\hat{u}|u), \\ Y \ominus (X, S) \ominus (U, Z), \\ Z \ominus (U, S) \ominus (X, Y), \end{array} \right. \quad (44)$$

and $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathbb{1}(\hat{u}|u)$ is achievable.

2) The probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathbb{1}(\hat{u}|u)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_1} \left(I(U, W_1; Y, Z) - I(W_1; S|U) - H(U) \right) > 0, \quad (45)$$

3) The probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathbb{1}(\hat{u}|u)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_1} \left(I(U, W_1; Y, Z) - I(W_1; S|U) - H(U) \right) < 0, \quad (46)$$

where \mathbb{Q}_1 is the set of distributions $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{W}_1 \times \mathcal{X} \times \mathcal{Y} \times \mathcal{U})$ with auxiliary random variable W_1 that satisfies:

$$\left\{ \begin{array}{l} \sum_{w_1 \in \mathcal{W}_1} \mathcal{Q}(u, s, z, w_1, x, y, \hat{u}) \\ = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathbb{1}(\hat{u}|u), \\ Y \ominus (X, S) \ominus (U, Z, W_1), \\ Z \ominus (U, S) \ominus (X, Y, W_1). \end{array} \right.$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}_1$ decomposes as follows:

$$\mathcal{Q}(u, s, z, w_1, x, y, \hat{u}) = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(w_1|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(\hat{u}|u).$$

The support of the auxiliary random variables W_1 is bounded by $|\mathcal{W}_1| \leq |\mathcal{B}| + 1$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark III.2 This result was already stated in [4] and [5] with a more restrictive lossless decoding constraint: $\mathcal{P}(\hat{U}^n \neq U^n) \leq \varepsilon$. It generalizes the coding theorem of Gel'fand Pinsker stated in [3].

A. Achievability Proof

This proof can be obtained from the achievability result of Sec. I-A, by replacing random variables W_2 and V by U . Note that $I(U, W_1; Y, Z) - I(W_1; S|U) - H(U) = I(W_1, U; Y, Z) - I(W_1, U; U, S)$.

Denote by $\mathcal{Q}(u, s, z, w_1, x, y, \hat{u}) \in \mathbb{Q}_1$ the joint probability distribution that achieves the maximum in equation (45). There exists $\delta > 0$ and rate $R \geq 0$ such that :

$$R \geq I(W_1, U; U, S) + \delta, \quad (47)$$

$$R \leq I(W_1, U; Y, Z) - \delta. \quad (48)$$

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ pairs of sequences $(W_1^n(m), U^n(m))$ with index $m \in \mathcal{M}$ drawn from the i.i.d. marginal probability distribution $\mathcal{Q}_{w_1 u}^{\otimes n}$.
- *Encoding function.* The encoder observes the sequences of source symbols $U^n \in \mathcal{U}^n$ and state symbols $S^n \in \mathcal{S}^n$. It finds the index $m \in \mathcal{M}$ such that the sequences $(U^n, S^n, W_1^n(m), U^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical (for probability distribution $\mathcal{Q}_{usw_1} \otimes \mathbb{1}_{u|u}$). Encoder sends a sequence X^n drawn from the conditional probability distribution $\mathcal{Q}_{x|usw_1}^{\otimes n}$ depending on sequences $(S^n, W_1^n(m), U^n(m))$.
- *Decoding function.* The decoder observes the pair of sequences (Y^n, Z^n) and finds the index $m \in \mathcal{M}$ such that the sequences $(Y^n, Z^n, W_1^n(m), U^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Decoder returns the sequence $\hat{U}^n = U^n(m)$.

From the properties of typical sequences, packing and covering Lemmas stated in [9] pp. 27, 46 and 208, equations (47), (48), imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (49)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, (U^n, S^n, W_1^n(m), U^n(m)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (50)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } (Y^n, Z^n, W_1^n(m'), U^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (51)$$

For all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U^n, S^n, Z^n, W_1^n(m), X^n, Y^n, \hat{U}^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{usz}(u, s, z) \otimes \mathcal{Q}(x, w_1|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathbb{1}(\hat{u}|u)$ with probability more than $1 - 3\varepsilon$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem III.1.

An alternative achievability proof based on superposition coding can be found in [4] and [5].

B. Converse Proof

We introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, \hat{U}^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, \hat{U}^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (52)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, \hat{u})$, *i.e.* for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ is small. We have equations:

$$n \cdot H(U) = H(U^n) \quad (53)$$

$$= H(E) + H(U^n|E) - H(E|U^n) \quad (54)$$

$$\leq 1 + \mathcal{P}(E = 0) \cdot H(U^n|E = 0) + \mathcal{P}(E = 1) \cdot H(U^n|E = 1) \quad (55)$$

$$\leq 1 + H(U^n|E = 0) + \mathcal{P}(E = 1) \cdot n \cdot \log_2 |\mathcal{U}| \quad (56)$$

$$\leq H(U^n|E = 0) + n \cdot \varepsilon. \quad (57)$$

Equation (53) comes from the i.i.d. property of the source.

Equations (54), (55) and (56) comes from the properties of the mutual information.

Equation (57) comes from the hypothesis of small error probability $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ and large length $n \in \mathbb{N}$ of the codewords, hence $\frac{1}{n} + \mathcal{P}(E = 1) \cdot \log_2 |\mathcal{U}| \leq \varepsilon$.

$$H(U^n|E=0) = I(U^n; Y^n, Z^n|E=0) + H(U^n|Y^n, Z^n, E=0) \quad (58)$$

$$= I(U^n; Y^n, Z^n|E=0) + n \cdot \varepsilon \quad (59)$$

$$= \sum_{i=1}^n I(U^n; Y_i, Z_i|Y^{i-1}, Z^{i-1}, E=0) + n \cdot \varepsilon \quad (60)$$

$$\leq \sum_{i=1}^n I(U^n, Y^{i-1}, Z^{i-1}; Y_i, Z_i|E=0) + n \cdot \varepsilon \quad (61)$$

$$= \sum_{i=1}^n I(U^n, Y^{i-1}, Z^{i-1}, S_{i+1}^n; Y_i, Z_i|E=0) \\ - \sum_{i=1}^n I(S_{i+1}^n; Y_i, Z_i|U^n, Y^{i-1}, Z^{i-1}, E=0) + n \cdot \varepsilon \quad (62)$$

$$= \sum_{i=1}^n I(U^n, Y^{i-1}, Z^{i-1}, S_{i+1}^n; Y_i, Z_i|E=0) \\ - \sum_{i=1}^n I(S_i; Y^{i-1}, Z^{i-1}|U^n, S_{i+1}^n, E=0) + n \cdot \varepsilon \quad (63)$$

$$= \sum_{i=1}^n I(U_i, U^{-i}, Y^{i-1}, Z^{i-1}, S_{i+1}^n; Y_i, Z_i|E=0) \\ - \sum_{i=1}^n I(S_i; U^{-i}, Y^{i-1}, Z^{i-1}, S_{i+1}^n|U_i, E=0) + n \cdot 2\varepsilon \quad (64)$$

$$= \sum_{i=1}^n I(U_i, W_{1,i}; Y_i, Z_i|E=0) - \sum_{i=1}^n I(W_{1,i}; S_i|U_i, E=0) + n \cdot 2\varepsilon. \quad (65)$$

Equation (58) comes from the properties of the mutual information.

Equation (59) comes from Fano's inequality that implies $H(U^n|Y^n, Z^n, E=0) \leq n \cdot \varepsilon$, as stated in Lemma 1.

Equation (60), (61) and (62) come from the properties of the mutual information.

Equation (63) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (64) comes from the independence of random variables (U_i, S_i) with (U^{-i}, S_{i+1}^n) that implies that $\sum_{i=1}^n I(S_i; U^{-i}, S_{i+1}^n|U_i, E=0) \leq n \cdot \varepsilon$, as stated in Lemma 10 in the Appendix.

Equation (65) comes from the introduction of the auxiliary random variable $W_{1,i} = (U^{-i}, Y^{i-1}, S_{i+1}^n)$. The Markov chain property $Y_i \text{---} (X_i, S_i) \text{---} (U_i, W_{1,i})$ is satisfied for all $i \in \{1, \dots, n\}$ since the channel is memoryless and Y_i do not belong to $W_{1,i}$. The random variable $W_{1,i}$ belongs to the set of probability distributions \mathbb{Q}_1 for lossless decoding.

$$\begin{aligned}
H(U^n|E=0) &\leq \sum_{i=1}^n I(U_i, W_{1,i}; Y_i, Z_i|E=0) - \sum_{i=1}^n I(W_{1,i}; S_i|U_i, E=0) + n \cdot 2\varepsilon \\
&= n \cdot \left(I(U_T, W_{1,T}; Y_T, Z_T|T, E=0) - I(W_{1,T}; S_T|U_T, T, E=0) + 2\varepsilon \right) \tag{66}
\end{aligned}$$

$$\begin{aligned}
&= n \cdot \left(I(U_T, W_{1,T}, T; Y_T, Z_T|E=0) - I(W_{1,T}, T; S_T|U_T, E=0) \right. \\
&\quad \left. - I(T; Y_T, Z_T|E=0) + I(T; S_T|U_T, E=0) + 2\varepsilon \right) \tag{67}
\end{aligned}$$

$$\leq n \cdot \left(I(U_T, W_{1,T}, T; Y_T, Z_T|E=0) - I(W_{1,T}, T; S_T|U_T, E=0) + 2\varepsilon \right) \tag{68}$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathcal{Q}} \left(I(U_T, W_1; Y_T, Z_T|E=0) - I(W_1; S_T|U_T, E=0) + 2\varepsilon \right) \tag{69}$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathcal{Q}} \left(I(U, W_1; Y, Z) - I(W_1; S|U) + 3\varepsilon \right). \tag{70}$$

Equation (66) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, X_T, Y_T, \hat{U}_T$.

Equation (67) comes from the properties of the mutual information.

Equation (68) comes from the independence of T with (U_T, S_T) that implies $I(T; S_T|U_T, E=0) = 0$, as stated in Lemma 11 in the Appendix. The memoryless property of the channel guarantees that the pair of random variable $(W_{1,T}, T)$ satisfies the Markov chain $Y_T \text{---} (X_T, S_T) \text{---} (U_T, W_{1,T}, T)$. Hence the pair $(W_{1,T}, T)$ belongs to the set of probability distributions \mathcal{Q}_1 , as stated in Lemma 7 in the Appendix. Equation (69) comes from taking the maximum over the probability distributions \mathcal{Q} that belong to \mathcal{Q}_1 . Equation (70) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, Y_T, \hat{U}_T) = (u, s, z, x, y, \hat{u})|E=0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, \hat{u})$. The continuity of the entropy function stated pp. 33 in [10] concludes.

Combining equations (57) and (70) gives equation (71). It is satisfied for all $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, \hat{u})$.

$$H(U) \leq \max_{\mathcal{Q} \in \mathcal{Q}} \left(I(U, W_1; Y, Z) - I(W_1; S|U) + 4\varepsilon \right). \tag{71}$$

This concludes the converse proof of Theorem III.1.

Lemma 1 *By Fano's inequality, we have the following equation:*

$$H(U^n|Y^n, Z^n, E = 0) \leq n \cdot \varepsilon. \quad (72)$$

Proof III.3 (Proof of Lemma 1) The decoding $g : \mathcal{Y}^n \times \mathcal{Z}^n \mapsto \mathcal{U}^n$ is deterministic, hence $H(\hat{U}^n|Y^n, Z^n, E = 0) = 0$. We have the following equations:

$$H(U^n|Y^n, Z^n, E = 0) \leq H(U^n, \hat{U}^n|Y^n, Z^n, E = 0) \quad (73)$$

$$\leq H(\hat{U}^n|Y^n, Z^n, E = 0) + H(U^n|\hat{U}^n, Y^n, Z^n, E = 0) \quad (74)$$

$$\leq H(U^n|\hat{U}^n, E = 0) \quad (75)$$

$$\leq n \cdot \left(H(U|\hat{U}) + \varepsilon \right) \quad (76)$$

$$= n \cdot \varepsilon. \quad (77)$$

Equations (73) and (74) come from the properties of the entropy.

Equation (75) comes from the deterministic decoding: \hat{U}^n is a deterministic function of (Y^n, Z^n) .

Equation (76) comes from the cardinality bound $\log_2 \left| \{u^n \text{ s.t. } u^n \in A_\varepsilon^{*n}(\hat{u}^n)\} \right| \leq n \cdot \left(H(U|\hat{U}) + \varepsilon \right)$, on the set of sequences u^n that are jointly typical with \hat{u}^n .

Equation (77) comes from the target joint distribution $\mathcal{Q}(u, s, z, x, y, \hat{u})$ that satisfy $\mathcal{Q}(\hat{u}|u) = \mathbb{1}(\hat{u}|u)$ hence $H(U|\hat{U}) = 0$.

IV. SEPARATION BETWEEN SOURCE AND CHANNEL

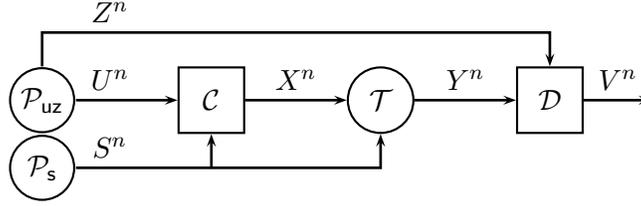


Fig. 4. Random variables of the source (U, Z, V) are independent of the random variables of the channel (S, X, Y).

Theorem IV.1 (Separation between Source and Channel)

1) The product of probability distribution $\mathcal{Q}(u, z, v) \otimes \mathcal{Q}(s, x, y)$ is achievable if and only if it decomposes as follows:

$$\begin{cases} \mathcal{Q}(u, z) = \mathcal{P}_{uz}(u, z), \\ \mathcal{Q}(s) = \mathcal{P}_s(s), \\ \mathcal{Q}(y|x, s) = \mathcal{T}(y|x, s). \end{cases} \quad (78)$$

and $\mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(v|u, z) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{T}(y|x, s)$ is achievable.

2) Joint probability distribution $\mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(v|u, z) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{T}(y|x, s)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_s} \left(I(W_1; Y) + I(W_2; Z) - I(W_1; S) - I(W_2; U) \right) > 0, \quad (79)$$

3) Joint probability distribution $\mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(v|u, z) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{T}(y|x, s)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_s} \left(I(W_1; Y) + I(W_2; Z) - I(W_1; S) - I(W_2; U) \right) < 0, \quad (80)$$

where \mathbb{Q}_s is the set $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{Z} \times \mathcal{W}_2 \times \mathcal{V}) \times \Delta(\mathcal{S} \times \mathcal{W}_1 \times \mathcal{X} \times \mathcal{Y})$ of product of probability distributions with auxiliary random variables (W_1, W_2) that satisfies:

$$\begin{cases} \sum_{(w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2} \mathcal{Q}(u, z, w_2, v) \otimes \mathcal{Q}(s, x, w_1, y) \\ \quad = \mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(v|u, z) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{T}(y|x, s), \\ Y \perp\!\!\!\perp (X, S) \perp\!\!\!\perp W_1, \\ Z \perp\!\!\!\perp U \perp\!\!\!\perp W_2, \\ V \perp\!\!\!\perp (Z, W_2) \perp\!\!\!\perp U. \end{cases}$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}_s$ decomposes as follows:

$$\mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(w_2|u) \otimes \mathcal{Q}(v|z, w_2) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{Q}(w_1|s, x) \otimes \mathcal{T}(y|x, s).$$

The supports of the auxiliary random variables (W_1, W_2) are bounded by $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq (|\mathcal{B}| + 1) \cdot (|\mathcal{B}| + 2)$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark IV.2 This separation result was already stated in [6] for a given distortion level and a given channel cost.

A. Achievability Proof

The achievability proof is based on the combination of the achievability proof of Gel'fand Pinsker [3] and the achievability proof of Wyner Ziv [2].

Denote by $\mathcal{Q}(u, z, w_2, v) \otimes \mathcal{Q}(s, x, w_1, y) \in \mathbb{Q}_s$ the joint probability distribution that achieves the maximum in equation (79). There exists $\delta > 0$ and rates $R \geq 0$, $R_{L_1} \geq 0$ and $R_{L_2} \geq 0$ such that :

$$R + R_{L_2} \geq I(W_2; U) + \delta, \quad (81)$$

$$R_{L_1} \geq I(W_1; S) + \delta, \quad (82)$$

$$R + R_{L_1} \leq I(W_1; Y) - \delta, \quad (83)$$

$$R_{L_2} \leq I(W_2; Z) - \delta. \quad (84)$$

- *Random codebook.* Source Codebook: We generate $|\mathcal{M} \times \mathcal{M}_{L_2}| = 2^{n(R+R_{L_2})}$ sequences $W_2^n(m, l_2)$ drawn from the i.i.d. probability $\mathcal{Q}_{w_2}^{\otimes n}$ with indexes by $(m, l_2) \in \mathcal{M} \times \mathcal{M}_{L_2}$. Channel Codebook: We generate $|\mathcal{M} \times \mathcal{M}_{L_1}| = 2^{n(R+R_{L_1})}$ sequences $W_1^n(m, l_1)$ drawn from the i.i.d. probability $\mathcal{Q}_{w_1}^{\otimes n}$ with indexes $(m, l_1) \in \mathcal{M} \times \mathcal{M}_{L_1}$.
- *Encoding function.* Encoder observes the source sequence U^n and find the indexes $(m, l_2) \in \mathcal{M} \times \mathcal{M}_{L_2}$ such that $(U^n, W_2^n(m, l_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. For each index $m \in \mathcal{M}$, it finds an index $l_1 \in \mathcal{M}_{L_1}$ such that $(S^n, W_1^n(m, l_1)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence X^n drawn from the probability $\mathcal{Q}_{x|sw_1}^{\otimes n}$ depending on $(S^n, W_1^n(m, l_1))$.
- *Decoding function.* The decoder observes the output sequence Y^n and the sequence of state information Z^n . It finds the pair of indexes $(m, l_1) \in \mathcal{M} \times \mathcal{M}_{L_1}$ such that $(Y^n, W_1^n(m, l_1)) \in A_\varepsilon^{*n}(\mathcal{Q})$. It finds the index $l_2 \in \mathcal{M}_{L_2}$ such that $(Z^n, W_2^n(m, l_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$. It returns the sequence V^n drawn from the probability $\mathcal{Q}_{v|zw_2}^{\otimes n}$ depending on $(Z^n, W_2^n(m, l_2))$.

From the properties of typical sequences, packing and covering Lemmas stated in [9] pp. 27, 46 and 208, equations (81), (82), (83), (84) imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of

error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left(U^n \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (85)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(S^n \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (86)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall (m, l_2) \in \mathcal{M} \times \mathcal{M}_{L_2}, \quad (U^n, W_2^n(m, l_2)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (87)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall l_1 \in \mathcal{M}_{L_1}, \quad (S^n, W_1^n(m, l_1)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (88)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists (m', l'_1) \neq (m, l_1), \text{ s.t. } (Y^n, W_1^n(m', l'_1)) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (89)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists l'_2 \neq l_2, \text{ s.t. } (Z^n, W_2^n(m, l'_2)) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (90)$$

For all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U^n, S^n, Z^n, W_1^n(m, l_1), W_2^n(m, l_2), X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{uz}(u, z) \otimes \mathcal{Q}(w_2|u) \otimes \mathcal{Q}(v|z, w_2) \otimes \mathcal{P}_s(s) \otimes \mathcal{Q}(x|s) \otimes \mathcal{Q}(w_1|s, x) \otimes \mathcal{T}(y|x, s)$, with probability more than $1 - 6\varepsilon$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem IV.1.

B. Converse Proof

We consider the product of probability distribution $\mathcal{Q}(u, z, v) \otimes \mathcal{Q}(s, x, y)$ and we introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (91)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, z, v) \otimes \mathcal{Q}(s, x, y)$, *i.e.* for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ is small.

Upper Bound. For every code, the random variables satisfy the following equations:

$$\begin{aligned} & I(U^n; Y^n | E = 0) \\ &= \sum_{i=1}^n I(U^n; Y_i | Y^{i-1}, E = 0) \end{aligned} \quad (92)$$

$$\leq \sum_{i=1}^n I(U^n, Y^{i-1}; Y_i | E = 0) \quad (93)$$

$$= \sum_{i=1}^n I(U^n, Y^{i-1}, S_{i+1}^n; Y_i | E = 0) - \sum_{i=1}^n I(S_{i+1}^n; Y_i | U^n, Y^{i-1}, E = 0) \quad (94)$$

$$= \sum_{i=1}^n I(U^n, Y^{i-1}, S_{i+1}^n; Y_i | E = 0) - \sum_{i=1}^n I(S_i; Y^{i-1} | U^n, S_{i+1}^n, E = 0) \quad (95)$$

$$= \sum_{i=1}^n I(U^n, Y^{i-1}, S_{i+1}^n; Y_i | E = 0) - \sum_{i=1}^n I(S_i; U^n, Y^{i-1}, S_{i+1}^n | E = 0) + n \cdot \varepsilon \quad (96)$$

$$= \sum_{i=1}^n I(W_{1,i}; Y_i | E = 0) - \sum_{i=1}^n I(W_{1,i}; S_i | E = 0) + n \cdot \varepsilon. \quad (97)$$

Equations (92), (93) and (94) come from the properties of the mutual information.

Equation (95) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (96) comes from the i.i.d. property of the information source U and the independence with the channel states S , hence by Lemma 10 in the Appendix: $\sum_{i=1}^n I(S_i; U^n, S_{i+1}^n | E = 0) \leq n \cdot \varepsilon$.

Equation (97) comes from the introduction of the auxiliary random variable $W_{1,i} = (U^n, Y^{i-1}, S_{i+1}^n)$.

The Markov chain property $Y_i \text{---} (X_i, S_i) \text{---} W_{1,i}$ is satisfied since the channel is memoryless and Y_i is not included in $W_{1,i}$.

Lower Bound. For every code, the random variables satisfy the following equations:

$$\begin{aligned} & I(U^n; Y^n | E = 0) \\ = & I(U^n, Z^n; Y^n | E = 0) \end{aligned} \quad (98)$$

$$\geq I(U^n; Y^n | Z^n, E = 0) \quad (99)$$

$$= \sum_{i=1}^n I(U_i; Y^n | Z^n, U^{i-1}, E = 0) \quad (100)$$

$$\geq \sum_{i=1}^n I(U_i; Y^n, Z^{-i}, U^{i-1} | Z_i, E = 0) - n \cdot \varepsilon \quad (101)$$

$$= \sum_{i=1}^n I(U_i, Z_i; Y^n, Z^{-i}, U^{i-1} | E = 0) - \sum_{i=1}^n I(Z_i; Y^n, Z^{-i}, U^{i-1} | E = 0) - n \cdot \varepsilon \quad (102)$$

$$= \sum_{i=1}^n I(U_i; Y^n, Z^{-i}, U^{i-1} | E = 0) - \sum_{i=1}^n I(Z_i; Y^n, Z^{-i}, U^{i-1} | E = 0) - n \cdot \varepsilon \quad (103)$$

$$= \sum_{i=1}^n I(U_i; W_{2,i} | E = 0) - \sum_{i=1}^n I(Z_i; W_{2,i} | E = 0) - n \cdot \varepsilon. \quad (104)$$

Equation (98) comes from the Markov chain $Z^n \circlearrowleft U^n \circlearrowleft Y^n$.

Equations (99) and (100) come from the properties of the mutual information.

Equation (101) comes from the i.i.d. property of the information sources (U, Z) and Lemma 10 in the Appendix that implies $\sum_{i=1}^n I(U_i; Z^{-i}, U^{i-1} | Z_i, E = 0) \leq n \cdot \varepsilon$.

Equation (102) comes from the properties of the mutual information.

Equation (103) comes from the Markov chain property $Z_i \circlearrowleft U_i \circlearrowleft (Y^n, Z^{-i}, U^{i-1})$ that is valid for all $i \in \{1, \dots, n\}$. This comes from the i.i.d. property of the source (U, Z) .

Equation (104) comes from the introduction of the auxiliary random variable $W_{2,i} = (Y^n, Z^{-i}, U^{i-1})$.

The random variable $W_{2,i}$ satisfies the Markov Chains that correspond to the set of product probability distributions \mathbb{Q}_s for the separation between source and channel:

$$Z_i \circlearrowleft U_i \circlearrowleft W_{2,i}, \quad (105)$$

$$V_i \circlearrowleft (Z_i, W_{2,i}) \circlearrowleft U_i. \quad (106)$$

- The first Markov chain comes from the i.i.d. property of the source and the fact that Z_i does not belong to $W_{2,i}$.
- The second Markov chain comes from the non-causal decoding: V_i is a function of (Y^n, Z^n) that is included in $(Z_i, W_{2,i}) = (Z_i, Y^n, Z^{-i}, U^{i-1})$.

Combining upper and lower bounds. Equations (97) and (104) give equation (107):

$$\begin{aligned}
0 &\leq \sum_{i=1}^n I(W_{1,i}; Y_i | E = 0) - \sum_{i=1}^n I(W_{1,i}; S_i | E = 0) \\
&+ \sum_{i=1}^n I(W_{2,i}; Z_i | E = 0) - \sum_{i=1}^n I(W_{2,i}; U_i | E = 0) + n \cdot 2\varepsilon
\end{aligned} \tag{107}$$

$$\begin{aligned}
&= n \cdot \left(I(W_{1,T}; Y_T | T, E = 0) - I(W_{1,T}; S_T | T, E = 0) \right. \\
&\left. + I(W_{2,T}; Z_T | T, E = 0) - I(W_{2,T}; U_T | T, E = 0) + 2\varepsilon \right)
\end{aligned} \tag{108}$$

$$\begin{aligned}
&\leq n \cdot \left(I(W_{1,T}, T; Y_T | E = 0) - I(W_{1,T}, T; S_T | E = 0) \right. \\
&\left. + I(W_{2,T}, T; Z_T | E = 0) - I(W_{2,T}, T; U_T | E = 0) + 2\varepsilon \right)
\end{aligned} \tag{109}$$

$$\leq n \cdot \left(I(W_{1,T}, T; Y) - I(W_{1,T}, T; S) + I(W_{2,T}, T; Z) - I(W_{2,T}, T; U) + 3\varepsilon \right) \tag{110}$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y) + I(W_2; Z) - I(W_1; S) - I(W_2; U) + 3\varepsilon \right). \tag{111}$$

Equation (108) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (109) comes from the independence of T with S_T and with U_T as stated in Lemma 11 in the Appendix. This implies $I(T; S_T | E = 0) = I(T; U_T | E = 0) = 0$.

Equation (110) comes from replacing the mean random variables $(U_T, S_T, Z_T, X_T, Y_T, V_T)$ by the random variables (U, S, Z, X, Y, V) with probability distribution $\mathcal{Q}(u, z, v) \otimes \mathcal{Q}(s, x, y)$. Lemma 9 states that the probability distribution induced by the coding scheme $\mathcal{P}((S_T, U_T, Z_T, X_T, Y_T, V_T) = (s, u, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, z, v) \otimes \mathcal{Q}(s, x, y)$. It remains to apply the continuity of the entropy function stated pp. 33 in [10] to obtain this upper bound with the additional error term: $n \cdot \varepsilon$.

Equation (111) comes from identifying W_1 and W_2 with $(W_{1,T}, T)$ and $(W_{2,T}, T)$ and taking the maximum over the distributions $\mathcal{Q}(u, z, w_2, v) \otimes \mathcal{Q}(s, x, w_1, y) \in \mathbb{Q}_s$. Auxiliary random variables $W_1 = (W_{1,T}, T)$ and $W_2 = (W_{2,T}, T)$ satisfy the three Markov chains of the set \mathbb{Q}_s as stated in Lemma 7 in the Appendix:

$$Y \text{ --- } (X, S) \text{ --- } W_1, \tag{112}$$

$$Z \text{ --- } U \text{ --- } W_2, \tag{113}$$

$$V \text{ --- } (Z, W_2) \text{ --- } U. \tag{114}$$

Remark IV.3 By construction, W_1 and W_2 are correlated since $W_{1,i} = (U^n, Y^{i-1}, S_{i+1}^n)$ and $W_{2,i} = (Y^n, Z^{-i}, U^{i-1})$ have a common part, namely (U^{i-1}, Y^{i-1}) . However, the random variables of the channel (S, X, Y) are independent of ones of the source (U, Z, V) . Hence, the optimal distribution for maximizing the first term $I(W_1; Y) - I(W_1; S)$ of equation (111) does not depend on the distribution of random variables W_2 and (U, Z, V) . Similarly, the optimal distribution for maximizing the second term $I(W_2; Z) - I(W_2; U)$ of equation (111) does not depend on the distribution of random variables W_1 and (S, X, Y) .

$$\max_{\mathcal{Q}_{XY}} \left(\mathbb{E}[f(X)] + \mathbb{E}[g(Y)] \right) = \max_{\mathcal{Q}_X} \mathbb{E}[f(X)] + \max_{\mathcal{Q}_Y} \mathbb{E}[g(Y)] = \max_{\mathcal{Q}_X \otimes \mathcal{Q}_Y} \left(\mathbb{E}[f(X)] + \mathbb{E}[g(Y)] \right).$$

Hence, the optimal distribution $\mathcal{Q} \in \mathbb{Q}_s$ for the optimization problem of equation (111) is a product of independent probability distributions $\mathcal{Q}(u, z, w_2, v) \otimes \mathcal{Q}(s, x, w_1, y)$. The author would like to thank Pablo Piantanida, Matthieu Bloch and Claudio Weidmann for useful discussions about the independence of the auxiliary random variables (W_1, W_2) for this converse.

The probability distribution $\mathcal{Q} \in \mathbb{Q}_d$ decomposes as follows:

$$\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(w_1, w_2|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_2).$$

The supports of the auxiliary random variables (W_1, W_2) are bounded by: $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq (|\mathcal{B}| + 1) \cdot (|\mathcal{B}| + 2)$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark V.2 Note that the last Markov chain is different from the previous results for coordination since V is generated using $\mathcal{Q}(v|y, z, w_2)$ instead of $\mathcal{Q}(v|y, z, w_1, w_2)$.

Remark V.3 This result was already stated in [7] without considering state informations at the encoder S and at the decoder Z .

A. Achievability Proof

We consider a probability distribution $\mathcal{Q} \in \mathbb{Q}_d$ that satisfies equation (116). There exists a $\delta > 0$ and rates $R \geq 0, R_L \geq 0$ such that:

$$R \geq I(W_2; U, S) + \delta, \quad (118)$$

$$R_L \geq I(W_1; U, S, W_2) + \delta, \quad (119)$$

$$R + R_L \leq I(W_1; Y, Z, W_2) - \delta. \quad (120)$$

We consider a block-Markov random code $c \in \mathcal{C}(n)$ defined over $B \in \mathbb{N}$ blocs of length $n \in \mathbb{N}$. The total length of the code is denoted by $N = n \cdot B \in \mathbb{N}$.

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ sequences $W_2^n(m)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_{w_2}^{\otimes n}$ with index $m \in \mathcal{M}$. We generate $|\mathcal{M} \times \mathcal{M}_L| = 2^{n(R+R_L)}$ sequences $W_1^n(m, l)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_{w_1}^{\otimes n}$ with indexes $(m, l) \in \mathcal{M} \times \mathcal{M}_L$.
- *Encoding function.* At block $b \in \{2, \dots, B-1\}$, the encoder observes the sequence of symbols of source and state $(U_{b+1}^n, S_{b+1}^n) \in \mathcal{U}^n \times \mathcal{S}^n$ over the next block $b+1$. It finds an index $m \in \mathcal{M}$ such that the sequences $(U_{b+1}^n, S_{b+1}^n, W_{2,b+1}^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder observes the jointly typical sequences of symbols $(U_b^n, S_b^n, W_{2,b}^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{W}_2^n$ of the current block b and finds the index $l \in \mathcal{M}_L$ such that the sequences $(U_b^n, S_b^n, W_{2,b}^n, W_{1,b}^n(m, l)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence X_b^n drawn from the transition probability $\mathcal{Q}_{x|usw_1w_2}^{\otimes n}$ depending on the sequences $(U_b^n, S_b^n, W_{2,b}^n, W_{1,b}^n(m, l))$ of block $b \in \mathbb{N}$.
- *Decoding function.* At the end of block $b \in \{2, \dots, B-1\}$, the decoder observes the pair of sequences (Y_b^n, Z_b^n) and recalls the sequence $W_{2,b}^n$ of block $b \in \mathbb{N}$. It finds the indexes $(m, l) \in \mathcal{M} \times \mathcal{M}_L$ such that $(Y_b^n, Z_b^n, W_{2,b}^n, W_{1,b}^n(m, l)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. It deduces the sequence $W_{2,b+1}^n(m)$ corresponding to index $m \in \mathcal{M}$ over the block $b+1 \in \mathbb{N}$. In the next block $b+1 \in \mathbb{N}$, it returns the sequence V_{b+1}^n drawn from the transition probability $\mathcal{Q}_{v|yzw_2}^{\otimes n}$ depending on the sequences $(Y_{b+1}^n, Z_{b+1}^n, W_{2,b+1}^n(m))$.
- *First bloc.* Encoder finds the index $m \in \mathcal{M}$ such that the sequences $(U_{b_2}^n, S_{b_2}^n, W_{2,b_2}^n(m)) \in A_\varepsilon^{*n}$ are jointly typical in the second block b_2 . It sends index $m \in \mathcal{M}$ to the decoder using classical Gel'fand

Pinsker [3] coding scheme. Decoder returns an arbitrary sequence of symbols $V^n \in \mathcal{V}^n$.

Remark V.4 In the first bloc, Gel'fand Pinsker [3] coding scheme can be used to transmit index $m \in \mathcal{M}$. Consider (U, S) as state information for the encoder, Y as decoder output and Z as state information of the decoder. Equation (116) proves that there exists a distribution $\mathcal{Q}_{xw|us}$ such that $I(W; Y, Z) - I(W; U, S) > 0$.

$$0 < I(W_1; Y, Z, W_2) - I(W_1; U, S, W_2) - I(W_2; U, S) \quad (121)$$

$$= I(W_1; Y, Z|W_2) - I(W_1; U, S|W_2) - I(W_2; U, S) \quad (122)$$

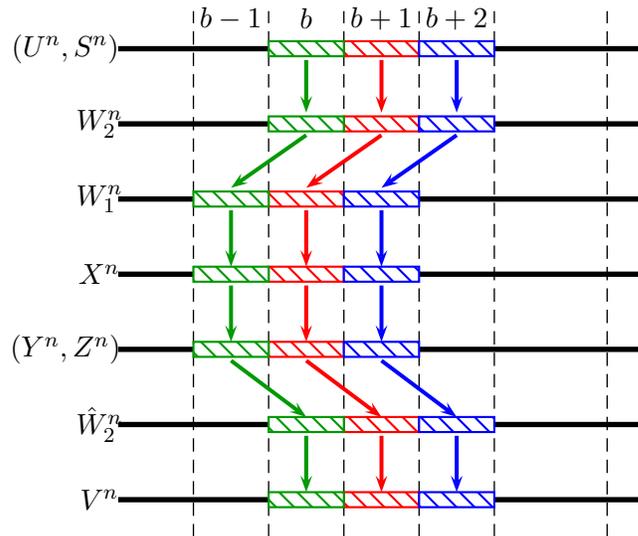
$$= I(W_1; Y, Z|W_2) - I(W_1, W_2; U, S) \quad (123)$$

$$= I(W_1, W_2; Y, Z) - I(W_1, W_2; U, S) - I(W_2; Y, Z) \quad (124)$$

$$\leq \max_{\mathcal{Q}_{xw|us}} \left(I(W; Y, Z) - I(W; U, S) \right) - I(W_2; Y, Z). \quad (125)$$

Hence Gel'fand Pinsker [3] coding scheme can be used at a rate strictly greater than $I(W_2; Y, Z)$. If it is necessary, the length $n \in \mathbb{N}$ of the first block can be adapted to transmit the index $m \in \mathcal{M}$ reliably.

- *Last bloc.* Encoder sends a sequence of symbols $X_B^n \in \mathcal{X}^n$ jointly typical with the sequences $(U_B^n, S_B^n, W_{2,B}^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{W}_2^n$. Decoder returns the sequence V_B^n drawn from the transition probability $\mathcal{Q}_{v|yzw_2}^{\otimes n}$ depending on the sequences $(Y_B^n, Z_B^n, W_{2,B}^n)$. The sequences $(U_B^n, S_B^n, Z_B^n, X_B^n, Y_B^n, V_B^n) \in \mathcal{A}_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical over the last block.



For each block $b \in \{2, \dots, B\}$, the properties of typical sequences, packing and covering lemmas stated in [9] pp. 27, 46 and 208, equations (118), (119), (120) imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{\star n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (126)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, \quad ((U_{b+1}^n, S_{b+1}^n, W_{2,b+1}^n(m)) \notin A_\varepsilon^{\star n}(\mathcal{Q})) \right) \right] \leq \varepsilon, \quad (127)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall l \in \mathcal{M}_L, \quad (U_b^n, S_b^n, W_{2,b}^n, W_{1,b}^n(m, l)) \notin A_\varepsilon^{\star n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (128)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists (m', l') \neq (m, l), \text{ s.t. } (Y_b^n, Z_b^n, W_{2,b}^n, W_{1,b}^n(m', l')) \in A_\varepsilon^{\star n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (129)$$

For each block $b \in \{2, \dots, B\}$, for all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U_b^n, S_b^n, Z_b^n, W_{1,b}^n, W_{2,b}^n, X_b^n, Y_b^n, V_b^n) \in A_\varepsilon^{\star n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{\text{usZ}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{Q}(w_1, w_2|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_2)$ with probability more than $1 - 4\varepsilon$.

We denote by $\tilde{Q}^N \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$, the empirical distribution of symbols over every blocs $b \in \{2, \dots, B\}$ removing the first bloc. We show \tilde{Q}^N is close to the empirical distribution Q^N over all the B blocks, for a number of block $B \in \mathbb{N}$ sufficiently large, *i.e.* for which $\frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \leq \varepsilon$. We denote by Q_1 , the empirical distribution of symbols over the first bloc.

$$\begin{aligned} & \left\| Q^N - \tilde{Q}^N \right\|_{\text{tv}} \\ &= \left\| \frac{1}{B} \cdot \left((B-1) \cdot \tilde{Q}^N + Q_1 \right) - \tilde{Q}^N \right\|_{\text{tv}} \\ &= \frac{1}{B} \cdot \left\| Q_1 - \tilde{Q}^N \right\|_{\text{tv}} \leq \frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \leq \varepsilon. \end{aligned}$$

Then, the expected error probability is bounded by ε .

$$\begin{aligned} & \mathbb{E}_c \left[\mathcal{P}_e(c) \right] \\ &= \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\ &= \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \tilde{Q}^N + \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\ &\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \tilde{Q}^N \right\| + \left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\ &\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon - \frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \right) \right] \\ &\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq \varepsilon \right) \right] \leq 1 - \left(1 - 4\varepsilon \right)^{B-1}. \end{aligned}$$

This implies the existence of a code $c^* \in \mathcal{C}(N)$ with an error probability below $(1 - 4\varepsilon)^{B-1}$ for all $N \geq B \cdot \bar{n}$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem V.1.

B. Converse Proof

We consider the joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ and we introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (130)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, v)$, *i.e.* for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ goes to zero. We have the following equations:

$$0 = \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}; U_i, S_i | U_{i+1}^n, S_{i+1}^n) \quad (131)$$

$$= \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}; U_{i+1}^n, S_{i+1}^n; U_i, S_i) \quad (132)$$

$$= \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | W_{2,i}) - \sum_{i=1}^n I(W_{1,i}, W_{2,i}; U_i, S_i). \quad (133)$$

Equation (131) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (132) comes from the i.i.d. property of the information source (U, S) that implies $I(U_i, S_i; U_{i+1}^n, S_{i+1}^n) = 0$ for all $i \in \{1, \dots, n\}$.

Equation (133) comes from the introduction of the auxiliary random variables $W_{1,i} = (U_{i+1}^n, S_{i+1}^n)$ and $W_{2,i} = (Y^{i-1}, Z^{i-1})$. The two random variables $(W_{1,i}, W_{2,i})$ satisfy the Markov Chains corresponding to the set of probability distributions \mathbb{Q}_d :

$$Z_i \circlearrowleft (U_i, S_i) \circlearrowleft (X_i, Y_i, W_{1,i}, W_{2,i}), \quad (134)$$

$$Y_i \circlearrowleft (X_i, S_i) \circlearrowleft (U_i, Z_i, W_{1,i}, W_{2,i}), \quad (135)$$

$$V_i \circlearrowleft (Y_i, Z_i, W_{2,i}) \circlearrowleft (U_i, S_i, X_i, W_{1,i}). \quad (136)$$

- The first Markov chain comes from i.i.d. property of the source and the fact that Z_i does not belong to $(W_{1,i}, W_{2,i})$.
- The second Markov chain comes from memoryless property of the channel and the fact that Y_i does not belong to $(W_{1,i}, W_{2,i})$.
- The third Markov chain comes from the causal decoding: V_i is a function of (Y^i, Z^i) that corresponds to $(Y_i, Z_i, Y^{i-1}, Z^{i-1}) = (Y_i, Z_i, W_{2,i})$.

$$0 = \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | W_{2,i}) - \sum_{i=1}^n I(W_{1,i}, W_{2,i}; U_i, S_i) \quad (137)$$

$$= n \cdot \left(I(W_{1,T}; Y_T, Z_T | W_{2,T}, T) - I(W_{1,T}, W_{2,T}; U_T, S_T | T) \right) \quad (138)$$

$$= n \cdot \left(I(W_{1,T}; Y_T, Z_T | W_{2,T}, T) - I(W_{1,T}, W_{2,T}, T; U_T, S_T) \right) \quad (139)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | W_2) - I(W_1, W_2; U_T, S_T) \right) \quad (140)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_1, W_2; U_T, S_T | E = 0) + \varepsilon \right) \quad (141)$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1; Y, Z | W_2) - I(W_1, W_2; U, S) + 2\varepsilon \right). \quad (142)$$

Equation (138) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (139) comes from the i.i.d. property of the information source that implies $I(T; U_T, S_T) = 0$.

Equation (140) comes from identifying W_1 and W_2 with $W_{1,T}$ and $(W_{2,T}, T)$ and taking the maximum over the probability distributions that belong to \mathbb{Q}_d . This is made possible since the random variables $W_{1,T}$ and $(W_{2,T}, T)$ satisfies the three Markov chains of the set of probability distributions \mathbb{Q}_d , as stated in Lemma 7 in the Appendix.

Equation (141) comes from the empirical coordination requirement as stated in Lemma 8 in the Appendix. Sequences are not jointly typical with small error probability $\mathcal{P}(E = 1)$.

Equation (142) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, Y_T, V_T) = (u, s, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes.

This concludes the converse proof of Theorem V.1.

VI. CAUSAL ENCODING

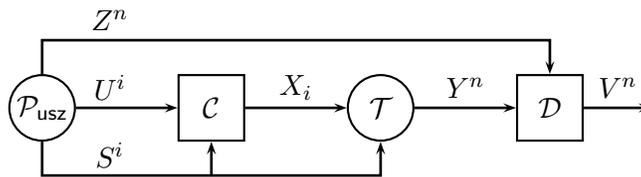


Fig. 6. Causal encoding function $f_i : \mathcal{U}^i \times \mathcal{S}^i \rightarrow \mathcal{X}$, for all $i \in \{1, \dots, n\}$ and non-causal decoding function $g : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \mathcal{V}^n$

Theorem VI.1 (Causal Encoding)

1) Joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ is achievable if and only if it decomposes as follows:

$$\left\{ \begin{array}{l} \mathcal{Q}(u, s, z) = \mathcal{P}_{\text{usz}}(u, s, z), \\ \mathcal{Q}(y|x, s) = \mathcal{T}(y|x, s), \\ Y \perp\!\!\!\perp (X, S) \perp\!\!\!\perp (U, Z), \\ Z \perp\!\!\!\perp (U, S) \perp\!\!\!\perp (X, Y). \end{array} \right. \quad (143)$$

and $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is achievable.

2) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_e} \left(I(W_1, W_2; Y, Z) - I(W_2; U, S|W_1) \right) > 0, \quad (144)$$

3) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathbb{Q}_e} \left(I(W_1, W_2; Y, Z) - I(W_2; U, S|W_1) \right) < 0, \quad (145)$$

where \mathbb{Q}_e is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ with

auxiliary random variables (W_1, W_2) that satisfies:

$$\left\{ \begin{array}{l} \sum_{(w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2} \mathcal{Q}(u, s, z, w_1, w_2, x, y, v) \\ = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y), \\ (U, S) \text{ independent of } W_1, \\ X \text{ --- } (U, S, W_1) \text{ --- } W_2, \\ Y \text{ --- } (X, S) \text{ --- } (U, Z, W_1, W_2), \\ Z \text{ --- } (U, S) \text{ --- } (X, Y, W_1, W_2), \\ V \text{ --- } (Y, Z, W_1, W_2) \text{ --- } (U, S, X). \end{array} \right.$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}_e$ decomposes as follows:

$$\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(w_1) \otimes \mathcal{Q}(w_2|u, s, w_1) \otimes \mathcal{Q}(x|u, s, w_1) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_1, w_2).$$

The supports of the auxiliary random variables (W_1, W_2) are bounded by $\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq (|\mathcal{B}| + 1) \cdot (|\mathcal{B}| + 2)$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark VI.2 This result was already stated in [1] without considering state informations at the encoder S and at the decoder Z .

A. Achievability Proof

We consider a probability distribution $\mathcal{Q} \in \mathbb{Q}_e$ that achieves the maximum in equation (144). There exists a $\delta > 0$ and a rate $\mathbf{R} > 0$ such that:

$$\mathbf{R} \geq I(W_2; U, S|W_1) + \delta, \quad (146)$$

$$\mathbf{R} \leq I(W_1; Y, Z) + I(W_2; Y, Z|W_1) - \delta = I(W_1, W_2; Y, Z) - \delta. \quad (147)$$

We consider a block-Markov random code $c \in \mathcal{C}(n)$ defined over $B \in \mathbb{N}$ blocs of length $n \in \mathbb{N}$. The total length of the code is denoted by $N = n \cdot B \in \mathbb{N}$ and \mathbf{R} denotes the rate of the code.

- *Random codebook.* We generate $|\mathcal{M}| = 2^{n\mathbf{R}}$ sequences $W_1^n(m)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_{w_1}^{\otimes n}$ with index $m \in \mathcal{M}$. For each index $m \in \mathcal{M}$, we generate the same number $|\mathcal{M}| = 2^{n\mathbf{R}}$ of sequences $W_2^n(m, \hat{m})$ with index $\hat{m} \in \mathcal{M}$, drawn from the i.i.d. conditional probability distribution $\mathcal{Q}_{w_2|w_1}^{\otimes n}$ depending on sequence $W_1^n(m)$.
- *Encoding function.* At the beginning of block $b \in \{2, \dots, B-1\}$, the encoder observes the sequences of source symbols $(U_{b-1}^n, S_{b-1}^n) \in \mathcal{U}^n \times \mathcal{S}^n$ of the previous block $b-1$. It also recalls the index $m_{b-1} \in \mathcal{M}$ of the sequence $W_1^n(m_{b-1}) \in \mathcal{W}_1^n$ over block $b-1$. It finds index $m_b \in \mathcal{M}$ such that the sequences $(U_{b-1}^n, S_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. It deduces the sequence $W_1^n(m_b)$ corresponding to the current block $b \in \{2, \dots, B-1\}$. Encoder sends the sequence X_b^n drawn from the conditional probability $\mathcal{Q}_{x|usw_1}^{\otimes n}$ depending on sequences $W_1^n(m_b)$ and (U_b^n, S_b^n) observed causally on the current block $b \in \{2, \dots, B-1\}$.

Remark VI.3 An alternative encoder can choose index $m_b \in \mathcal{M}$ such that the sequences $(U_{b-1}^n, S_{b-1}^n, X_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical, including with sequence X_{b-1}^n . This alternative encoder introduces more correlation between the random variables and the Markov chain $X \ominus (U, S, W_1) \ominus W_2$ is removed. However the corresponding rate constraint writes $I(W_2; U, S, X|W_1) = I(W_2; U, S|W_1) + I(W_2; X|W_1, U, S)$. Hence, the price of removing the Markov chain $X \ominus (U, S, W_1) \ominus W_2$ is equal to $I(W_2; X|W_1, U, S)$ and the converse proof concludes it is not optimal.

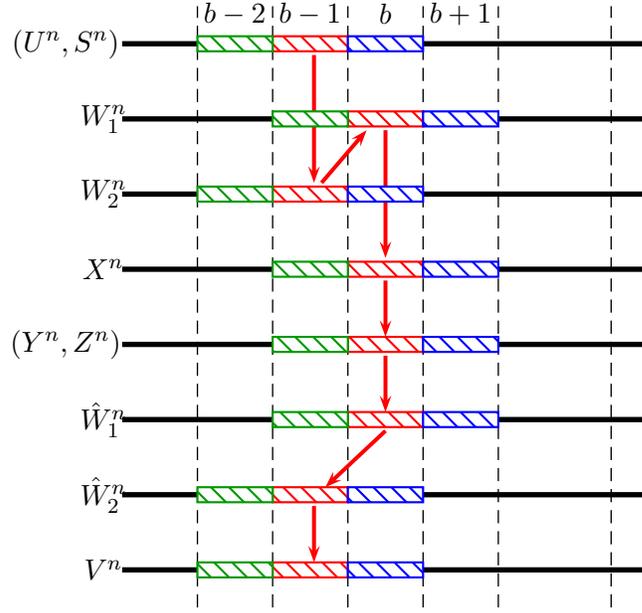
- *Decoding function.* At the end of block $b \in \{2, \dots, B-1\}$, the decoder recalls sequences (Y_{b-1}^n, Z_{b-1}^n) and the index $m_{b-1} \in \mathcal{M}$ corresponding to the sequence $W_1^n(m_{b-1})$. It observes

the sequences (Y_b^n, Z_b^n) and finds index $m_b \in \mathcal{M}$ such that $(Y_b^n, Z_b^n, W_1^n(m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ and $(Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Decoder returns sequence V_{b-1}^n drawn from the conditional probability distribution $\mathcal{Q}_{v|yzw_1w_2}^{\otimes n}$ depending on sequences $(Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m_b))$.

- *First bloc at the encoder.* An arbitrary index $m_1 \in \mathcal{M}$ of $W_1^n(m_1) \in \mathcal{W}_1^n$ is given to both encoder \mathcal{C} and decoder \mathcal{D} . Encoder sends the sequence $X_{b_1}^n$ drawn from the conditional probability distribution $\mathcal{Q}_{x|usw_1}^{\otimes n}$ depending on sequences $W_1^n(m_1)$ and $(U_{b_1}^n, S_{b_1}^n)$ observed causally on the current block b_1 . At the beginning of the second bloc b_2 , encoder recalls $(U_{b_1}^n, S_{b_1}^n, W_1^n(m_1))$ and finds index m_2 such that sequences $(U_{b_1}^n, S_{b_1}^n, W_1^n(m_1), W_2^n(m_1, m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence $X_{b_2}^n$ drawn from the conditional probability $\mathcal{Q}_{x|usw_1}^{\otimes n}$ depending on sequences $W_1^n(m_2)$ and $(U_{b_2}^n, S_{b_2}^n)$ observed causally on the second block b_2 .
- *First bloc at the decoder.* At the end of second block b_2 , the decoder finds the index m_2 such that $(Y_{b_2}^n, Z_{b_2}^n, W_1^n(m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ and $(Y_{b_1}^n, Z_{b_1}^n, W_1^n(m_1), W_2^n(m_1, m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Over the first bloc, decoder \mathcal{D} returns $V_{b_1}^n \in \mathcal{V}^n$ drawn from the conditional probability $\mathcal{Q}_{v|yzw_1w_2}^{\otimes n}$ depending on sequences $(Y_{b_1}^n, Z_{b_1}^n, W_1^n(m_1), W_2^n(m_1, m_2))$. Sequences $(U_{b_1}^n, S_{b_1}^n, Z_{b_1}^n, W_1^n(m_1), W_2^n(m_1, m_2), X_{b_1}^n, Y_{b_1}^n, V_{b_1}^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical over the first block b_1 .
- *Last bloc.* Encoder \mathcal{C} and decoder \mathcal{D} choose arbitrary sequences X_B^n and V_B^n . Sequences are not jointly typical on the last block.

Remark VI.4 The source and state (U, S) are jointly encoded using two streams of information represented by the auxiliary random variables W_1 and W_2 . Auxiliary random variable W_2 is used to quantify the past source and the past codeword (U, S, W_1) . The quantification index is transmitted in the next block using codeword W_1 . This process works in this way from one block to another.

For each block $b \in \{1, \dots, B-1\}$, the properties of typical sequences, packing and covering lemmas stated in [9] pp. 27, 46 and 208, equations (146), (147) imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected



probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (148)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, \quad (U_{b-1}^n, S_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (149)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } \left\{ (Y_b^n, Z_b^n, W_1^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \cap \left\{ (Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \right) \right] \leq \varepsilon. \quad (150)$$

Lemma 2 proves that equation (147) implies equation (150).

For each block $b \in \{1, \dots, B-1\}$, for all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U_b^n, S_b^n, Z_b^n, W_1^n(m_b), W_2^n(m_b, m_{b+1}), X_b^n, Y_b^n, V_b^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for the probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(w_1) \otimes \mathcal{Q}(w_2|u, s, w_1) \otimes \mathcal{Q}(x|u, s, w_1) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, w_1, w_2)$ with probability more than $1 - 3\varepsilon$.

We denote by $\tilde{Q}^N \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$, the empirical distribution of symbols over every blocs $b \in \{1, \dots, B-1\}$ removing the last bloc. We show \tilde{Q}^N is close to the empirical distribution Q^N over all the B blocks, for a number of blocks $B \in \mathbb{N}$ sufficiently large, *i.e.* for which $\frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \leq \varepsilon$.

We denote by Q_B , the empirical distribution of symbols over the last bloc.

$$\begin{aligned}
& \left\| Q^N - \tilde{Q}^N \right\|_{\text{tv}} \\
&= \left\| \frac{1}{B} \cdot \left((B-1) \cdot \tilde{Q}^N + Q_B \right) - \tilde{Q}^N \right\|_{\text{tv}} \\
&= \frac{1}{B} \cdot \left\| Q_B - \tilde{Q}^N \right\|_{\text{tv}} \leq \frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \leq \varepsilon.
\end{aligned}$$

Then, the expected error probability is bounded by ε .

$$\begin{aligned}
& \mathbb{E}_c \left[\mathcal{P}_e(c) \right] \\
&= \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\
&= \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \tilde{Q}^N + \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\
&\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| Q^N - \tilde{Q}^N \right\| + \left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon \right) \right] \\
&\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq 2\varepsilon - \frac{2}{B} \cdot |\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}| \right) \right] \\
&\leq \mathbb{E}_c \left[\mathcal{P} \left(\left\| \tilde{Q}^N - \mathcal{Q} \right\|_{\text{tv}} \geq \varepsilon \right) \right] \leq 1 - \left(1 - 3\varepsilon \right)^{B-1}.
\end{aligned}$$

This implies the existence of a code $c^* \in \mathcal{C}(N)$ with an error probability below $(1 - 3\varepsilon)^{B-1}$ for all $N \geq B \cdot \bar{n}$.

The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem VI.1.

Lemma 2 (Packing Lemma)

$$R \leq I(W_1; Y, Z) + I(W_2; Y, Z|W_1) - \delta = I(W_1, W_2; Y, Z) - \delta, \quad (151)$$

For each block $b \in \{1, \dots, B-1\}$ and for each previous message m_{b-1} , equation (151) implies that for all $\varepsilon > 0$, there exists a $\bar{n} \in \mathbb{N}$ such that for all $n \geq \bar{n}$, the expected error probability is upper bounded by $\varepsilon > 0$, as stated in equation (152).

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } \left\{ (Y_b^n, Z_b^n, W_1^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \cap \left\{ (Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \right) \right] \leq \varepsilon. \quad (152)$$

Proof VI.5 (Lemma 2) Consider $\varepsilon > 0$ that satisfies $\delta > 7\varepsilon$. Hence, we have:

$$R - I(Y, Z; W_1) - I(Y, Z; W_2|W_1) + 6\varepsilon \leq -\delta + 6\varepsilon < -\varepsilon. \quad (153)$$

We have the following equations:

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } \left\{ (Y_b^n, Z_b^n, W_1^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \cap \left\{ (Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \right) \right] \quad (154)$$

$$\leq \sum_{m' \neq m} \mathbb{E}_c \left[\mathcal{P} \left(\left\{ (Y_b^n, Z_b^n, W_1^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \cap \left\{ (Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \right) \right] \quad (155)$$

$$= \sum_{m' \neq m} \mathbb{E}_c \left[\mathcal{P} \left((Y_b^n, Z_b^n, W_1^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \times \mathbb{E}_c \left[\mathcal{P} \left((Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \quad (156)$$

$$\leq \sum_{m' \neq m} \sum_{(y^n, z^n, w_1^n) \in A_\varepsilon^{*n}(\mathcal{Q})} \mathbb{E}_c \left[\mathcal{P} \left((Y_b^n, Z_b^n, W_1^n(m')) = (y^n, z^n, w_1^n) \right) \right] \times \sum_{(y^n, z^n, w_1^n, w_2^n) \in A_\varepsilon^{*n}(\mathcal{Q})} \mathbb{E}_c \left[\mathcal{P} \left((Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) = (y^n, z^n, w_1^n, w_2^n) \right) \right]. \quad (157)$$

Equation (154) comes from the definition of the error event for block $b \in \{1, \dots, B-1\}$. Notations Y^n , Z^n stand for $Y^n(m)$, $Z^n(m)$ corresponding to the correct index $m \in \mathcal{M}$.

Equation (155) comes from Boole's inequality.

Equation (156) comes from the independence of the random variables $(Y_b^n, Z_b^n, W_1^n(m'))$ over block b with the random variables $(Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m'))$ over block $b-1$. This comes from the i.i.d. property of the source, the codebook with independent codewords and the block-Markov coding process.

Equation (157) comes from Boole's inequality.

$$\begin{aligned}
& \sum_{m' \neq m} \sum_{(y^n, z^n, w_1^n) \in A_\varepsilon^{*n}(\mathcal{Q})} \mathbb{E}_c \left[\mathcal{P} \left((Y_b^n, Z_b^n, W_1^n(m')) = (y^n, z^n, w_1^n) \right) \right] \\
& \times \sum_{(y^n, z^n, w_1^n, w_2^n) \in A_\varepsilon^{*n}(\mathcal{Q})} \mathbb{E}_c \left[\mathcal{P} \left((Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1}), W_2^n(m_{b-1}, m')) = (y^n, z^n, w_1^n, w_2^n) \right) \right] \\
& \leq \sum_{m' \neq m} \sum_{(y^n, z^n, w_1^n) \in A_\varepsilon^{*n}(\mathcal{Q})} \mathbb{E}_c \left[\mathcal{P} \left((Y_b^n, Z_b^n) = (y^n, z^n) \right) \right] \times \mathbb{E}_c \left[\mathcal{P} \left(W_1^n(m') = w_1^n \right) \right] \\
& \times \sum_{\substack{(y^n, z^n, w_1^n, w_2^n) \\ \in A_\varepsilon^{*n}(\mathcal{Q})}} \mathbb{E}_c \left[\mathcal{P} \left((Y_{b-1}^n, Z_{b-1}^n, W_1^n(m_{b-1})) = (y^n, z^n, w_1^n) \right) \right] \\
& \times \mathbb{E}_c \left[\mathcal{P} \left(W_2^n(m_{b-1}, m') = w_2^n \mid W_1^n(m_{b-1}) = w_1^n \right) \right] \tag{158}
\end{aligned}$$

$$\begin{aligned}
& \leq \sum_{m' \neq m} \sum_{(y^n, z^n, w_1^n) \in A_\varepsilon^{*n}(\mathcal{Q})} 2^{-n(H(Y,Z)-\varepsilon)} \times 2^{-n(H(W_1)-\varepsilon)} \\
& \times \sum_{\substack{(y^n, z^n, w_1^n, w_2^n) \\ \in A_\varepsilon^{*n}(\mathcal{Q})}} 2^{-n(H(Y,Z,W_1)-\varepsilon)} \times 2^{-n(H(W_2|W_1)-\varepsilon)} \tag{159}
\end{aligned}$$

$$\leq \sum_{m' \neq m} 2^{n(H(Y,Z,W_1)-H(Y,Z)-H(W_1)+3\varepsilon)} \times 2^{n(H(Y,Z,W_1,W_2)-H(Y,Z,W_1)-H(W_2|W_1)+3\varepsilon)} \tag{160}$$

$$\leq 2^{nR} \times 2^{n(-I(Y,Z;W_1)-I(Y,Z;W_2|W_1)+6\varepsilon)} \tag{161}$$

$$= 2^{n(R-I(Y,Z;W_1)-I(Y,Z;W_2|W_1)+6\varepsilon)} \tag{162}$$

$$\leq 2^{-n \cdot \varepsilon}. \tag{163}$$

Equation (158) comes from the independence of the sequences (Y_b^n, Z_b^n) with $W_1^n(m')$. The codebook with superposition induces that the sequence $W_2^n(m_{b-1}, m')$ depends on $W_1^n(m_{b-1})$.

Equations (159), (160), (161), (162) comes from the properties of the typical sequences and of the mutual information.

Equation (163) comes from the choice of $\varepsilon > 0$ that satisfies $\delta > 7\varepsilon$.

This proves that there exists a \bar{n} such that for all $n \geq \bar{n}$, equation (152) is satisfied. This concludes the proof of Lemma 2.

B. Converse Proof

We consider the joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ and we introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (164)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, v)$, i.e. for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ goes to zero. The converse is based on the following equations:

$$0 = \sum_{i=1}^n I(U^{i-1}, S^{i-1}; Y_i, Z_i | Y_{i+1}^n, Z_{i+1}^n) - \sum_{i=1}^n I(Y_{i+1}^n, Z_{i+1}^n; U_i, S_i | U^{i-1}, S^{i-1}) \quad (165)$$

$$\leq \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n; Y_i, Z_i) - \sum_{i=1}^n I(Y_{i+1}^n, Z_{i+1}^n; U_i, S_i | U^{i-1}, S^{i-1}) \quad (166)$$

$$= \sum_{i=1}^n I(W_{1,i}, W_{2,i}; Y_i, Z_i) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | W_{1,i}). \quad (167)$$

Equation (165) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (166) comes from the properties of the mutual information.

Equation (167) comes from the introduction of the auxiliary random variables $W_{1,i} = (U^{i-1}, S^{i-1})$ and $W_{2,i} = (Y_{i+1}^n, Z_{i+1}^n)$. For all $i \in \{1, \dots, n\}$, the auxiliary random variables $W_{1,i}$ and $W_{2,i}$ satisfy the properties corresponding to the set of probability distributions \mathbb{Q}_e :

$$(U_i, S_i) \text{ are independent of } W_{1,i}, \quad (168)$$

$$X_i \text{---} (U_i, S_i, W_{1,i}) \text{---} W_{2,i}, \quad (169)$$

$$Y_i \text{---} (X_i, S_i) \text{---} (U_i, Z_i, W_{1,i}, W_{2,i}), \quad (170)$$

$$Z_i \text{---} (U_i, S_i) \text{---} (X_i, Y_i, W_{1,i}, W_{2,i}), \quad (171)$$

$$V_i \text{---} (Y_i, Z_i, W_{1,i}, W_{2,i}) \text{---} (U_i, S_i, X_i). \quad (172)$$

- Equation (168) comes from the i.i.d. property of the source and states (U, S) that implies that $W_{1,i} = (U^{i-1}, S^{i-1})$ is independent of (U_i, S_i) .
- Equation (169) comes from the causal encoding function that implies that X_i is a deterministic function of $(U_i, S_i, U^{i-1}, S^{i-1})$ which is equal to $(U_i, S_i, W_{1,i})$.
- Equation (170) comes from the memoryless property of the channel and the fact that Y_i is not included in $(W_{1,i}, W_{2,i})$.

- Equation (171) comes from the i.i.d. property of the source and states (U_i, S_i, Z_i) and the fact that Z_i is not included in $(W_{1,i}, W_{2,i})$.
- Equation (172) comes from the causal encoding and the non-causal decoding as stated in Lemma 3.

Equation (167) gives:

$$\begin{aligned}
0 &\leq \sum_{i=1}^n I(W_{1,i}, W_{2,i}; Y_i, Z_i) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | W_{1,i}) \\
&= n \cdot \left(I(W_{1,T}, W_{2,T}; Y_T, Z_T | T) - I(W_{2,T}; U_T, S_T | W_{1,T}, T) \right) \tag{173}
\end{aligned}$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1, W_2; Y_T, Z_T) - I(W_2; U_T, S_T | W_1) \right) \tag{174}$$

$$\leq n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1, W_2; Y_T, Z_T | E = 0) - I(W_2; U_T, S_T | W_1, E = 0) + \varepsilon \right) \tag{175}$$

$$= n \cdot \max_{\mathcal{Q} \in \mathbb{Q}} \left(I(W_1, W_2; Y, Z) - I(W_2; U, S | W_1) + 2\varepsilon \right). \tag{176}$$

Equation (173) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (174) comes from identifying W_1 and W_2 with $(W_{1,T}, T)$ and $W_{2,T}$ and taking the maximum over the probability distributions that belong to the set \mathbb{Q}_ε . This is possible since the random variables $(W_{1,T}, T)$ and $W_{2,T}$ satisfy the properties of the set of probability distributions \mathbb{Q}_ε as stated in Lemma 7 in the Appendix.

Equation (175) comes from the empirical coordination requirement as stated in Lemma 8 in the Appendix. Sequences are not jointly typical with small error probability $\mathcal{P}(E = 1)$.

Equation (176) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, Y_T, V_T) = (u, s, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes.

This concludes the proof of Theorem VI.1.

Lemma 3 Markov chain $V_i \ominus (Y_i, Z_i, W_{1,i}, W_{2,i}) \ominus (U_i, S_i, X_i)$ is satisfied for all $i \in \{1, \dots, n\}$.

Proof VI.6 (Lemma 3) We evaluate the following probability:

$$\begin{aligned}
& \mathcal{P}(V_i | Y_i, Z_i, W_{1,i}, W_{2,i}, U_i, S_i, X_i) \\
&= \mathcal{P}(V_i | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(V_i, X^{i-1}, Y^{i-1}, Z^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i) \quad (177) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(Z^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i) \\
&\times \mathcal{P}(X^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}) \\
&\times \mathcal{P}(Y^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}, X^{i-1}) \\
&\times \mathcal{P}(V_i | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}, X^{i-1}, Y^{i-1}). \quad (178)
\end{aligned}$$

We can remove (U_i, S_i, X_i) , in the four conditional probability distributions:

$$\mathcal{P}(Z^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i) = \mathcal{P}(Z^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n), \quad (179)$$

$$\mathcal{P}(X^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}) = \mathcal{P}(X^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}), \quad (180)$$

$$\mathcal{P}(Y^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}, X^{i-1}) = \mathcal{P}(Y^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}, X^{i-1}), \quad (181)$$

$$\mathcal{P}(V_i | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, X_i, Z^{i-1}, X^{i-1}, Y^{i-1}) = \mathcal{P}(V_i | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}, X^{i-1}, Y^{i-1}). \quad (182)$$

Equation (179) comes from the i.i.d. property of the information source: Z^{i-1} only depends on (U^{i-1}, S^{i-1}) .

Equation (180) comes from the causal encoding: X^{i-1} is a deterministic function of (U^{i-1}, S^{i-1}) .

Equation (181) comes from the memoryless property of the channel: Y^{i-1} only depends on (X^{i-1}, S^{i-1}) .

Equation (182) comes from the non-causal decoding: V_i is a deterministic function of $(Y^n, Z^n) = (Y^{i-1}, Y_i, Y_{i+1}^n, Z^{i-1}, Z_i, Z_{i+1}^n)$.

Hence we have:

$$\begin{aligned}
& \mathcal{P}(V_i | Y_i, Z_i, W_{1,i}, W_{2,i}, U_i, S_i, X_i) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(V_i, X^{i-1}, Y^{i-1}, Z^{i-1} | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) \quad (183)
\end{aligned}$$

$$= \mathcal{P}(V_i | Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) \quad (184)$$

$$= \mathcal{P}(V_i | Y_i, Z_i, W_{1,i}, W_{2,i}). \quad (185)$$

The above equation corresponds to the Markov chain $V_i \ominus (Y_i, Z_i, W_{1,i}, W_{2,i}) \ominus (U_i, S_i, X_i)$ and it concludes the proof of Lemma 3.

VII. STRICTLY CAUSAL DECODING

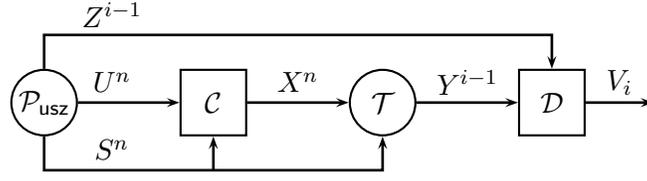


Fig. 7. Non-causal encoding function $f : \mathcal{U}^n \times \mathcal{S}^n \rightarrow \mathcal{X}^n$ and strictly causal decoding function $g_i : \mathcal{Y}^{i-1} \times \mathcal{Z}^{i-1} \rightarrow \mathcal{V}$ for all $i \in \{1, \dots, n\}$.

Theorem VII.1 (Strictly Causal Decoding)

1) Joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ is achievable if and only if it decomposes as follows:

$$\begin{cases} \mathcal{Q}(u, s, z) = \mathcal{P}_{\text{usz}}(u, s, z), \\ \mathcal{Q}(y|x, s) = \mathcal{T}(y|x, s), \\ Y \text{---} \text{---} (X, S) \text{---} \text{---} (U, Z, V), \\ Z \text{---} \text{---} (U, S) \text{---} \text{---} (X, Y, V). \end{cases} \quad (186)$$

and $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{T}(y|x, s)$ is achievable.

2) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{T}(y|x, s)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathcal{Q}_{\text{sd}}} \left(I(W_1; Y, Z|V) - I(W_1, V; U, S) \right) > 0, \quad (187)$$

3) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{T}(y|x, s)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathcal{Q}_{\text{sd}}} \left(I(W_1; Y, Z|V) - I(W_1, V; U, S) \right) < 0, \quad (188)$$

where \mathcal{Q}_{sd} is the set of distributions $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{W}_1 \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ with auxiliary random variable W_1 that satisfies:

$$\begin{cases} \sum_{w_1 \in \mathcal{W}_1} \mathcal{Q}(u, s, z, w_1, x, y, v) \\ = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{T}(y|x, s), \\ Y \text{---} \text{---} (X, S) \text{---} \text{---} (U, Z, W_1, V), \\ Z \text{---} \text{---} (U, S) \text{---} \text{---} (X, Y, W_1, V). \end{cases}$$

The probability distribution $\mathcal{Q} \in \mathcal{Q}_{\text{sd}}$ decomposes as follows:

$$\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{Q}(w_1|u, s, x, v) \otimes \mathcal{T}(y|x, s).$$

The supports of the auxiliary random variable W_1 is bounded by: $|\mathcal{W}_1| \leq |\mathcal{B}| + 1$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark VII.2 Strictly causal decoding requires the output of the decoder V is not directly correlated with (Y, Z) . Hence, the general probability distributions $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ reduces to $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{T}(y|x, s)$.

Remark VII.3 This result was already stated in [7] without considering state informations at the encoder S and at the decoder Z .

A. Achievability Proof

The achievability proof is very similar to the one of Theorem V.1 for causal decoding replacing W_2 by V . We consider a probability distribution $\mathcal{Q} \in \mathcal{Q}_{\text{sd}}$ that satisfies equation (187). There exists a $\delta > 0$ and rates $R \geq 0$, $R_L \geq 0$ such that:

$$R \geq I(V; U, S) + \delta, \quad (189)$$

$$R_L \geq I(W_1; U, S, V) + \delta, \quad (190)$$

$$R + R_L \leq I(W_1; Y, Z, V) - \delta. \quad (191)$$

We consider a block-Markov random code $c \in \mathcal{C}(n)$ defined over $B \in \mathbb{N}$ blocs of length $n \in \mathbb{N}$. The total length of the code is denoted by $N = n \cdot B \in \mathbb{N}$.

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ sequences $V^n(m)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_V^{\otimes n}$ with index $m \in \mathcal{M}$. We generate $|\mathcal{M} \times \mathcal{M}_L| = 2^{n(R+R_L)}$ sequences $W_1^n(m, l)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_{W_1}^{\otimes n}$ with indexes $(m, l) \in \mathcal{M} \times \mathcal{M}_L$.
- *Encoding function.* At block $b \in \{2, \dots, B-1\}$, the encoder observes the sequence of symbols of source and state $(U_{b+1}^n, S_{b+1}^n) \in \mathcal{U}^n \times \mathcal{S}^n$ of the next block $b+1$. It finds an index $m \in \mathcal{M}$ such that the sequences $(U_{b+1}^n, S_{b+1}^n, V_{b+1}^n(m)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder observes the jointly typical sequences of symbols $(U_b^n, S_b^n, V_b^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{V}^n$ of the current block b and finds the index $l \in \mathcal{M}_L$ such that the sequences $(U_b^n, S_b^n, V_b^n, W_{1,b}^n(m, l)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence X_b^n drawn from the transition probability $\mathcal{Q}_{X|USVW_1}^{\otimes n}$ depending on the sequences $(U_b^n, S_b^n, V_b^n, W_{1,b}^n(m, l))$ of block $b \in \mathbb{N}$.
- *Decoding function.* At the end of block $b \in \{2, \dots, B-1\}$, the decoder observes the pair of sequences (Y_b^n, Z_b^n) and recalls the sequence V_b^n of block $b \in \mathbb{N}$. It finds the indexes $(m, l) \in \mathcal{M} \times \mathcal{M}_L$ such that $(Y_b^n, Z_b^n, V_b^n, W_{1,b}^n(m, l)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. In the next block $b+1 \in \mathbb{N}$, it returns the sequence $V_{b+1}^n(m)$ corresponding to index $m \in \mathcal{M}$ over the block $b+1 \in \mathbb{N}$.
- *First bloc.* Encoder finds the index $m \in \mathcal{M}$ such that the sequences $(U_{b_2}^n, S_{b_2}^n, V_{b_2}^n(m)) \in A_\varepsilon^{*n}$ are jointly typical in the second block b_2 . It sends index $m \in \mathcal{M}$ to the decoder using classical Gel'fand Pinsker [3] coding scheme. Decoder returns an arbitrary sequence of symbols $V^n \in \mathcal{V}^n$.

Remark VII.4 In the first bloc, Gel'fand Pinsker [3] coding scheme can be used to transmit index $m \in \mathcal{M}$. Consider (U, S) as state information of the encoder, Y as decoder output and Z as state information of the decoder. Equation (187) proves that there exists a distribution $\mathcal{Q}_{xw|us}$ such that $I(W; Y, Z) - I(W; U, S) > 0$.

$$0 < I(W_1; Y, Z, V) - I(W_1; U, S, V) - I(V; U, S) \quad (192)$$

$$= I(W_1; Y, Z|V) - I(W_1; U, S|V) - I(V; U, S) \quad (193)$$

$$= I(W_1; Y, Z|V) - I(W_1, V; U, S) \quad (194)$$

$$= I(W_1, V; Y, Z) - I(W_1, V; U, S) - I(V; Y, Z) \quad (195)$$

$$\leq \max_{\mathcal{Q}_{xw|us}} \left(I(W; Y, Z) - I(W; U, S) \right) - I(V; Y, Z). \quad (196)$$

Hence Gel'fand Pinsker [3] coding scheme can be used at a rate strictly greater than $I(V; Y, Z)$. If it is necessary, the length $n \in \mathbb{N}$ of the first block can be adapted to transmit the index $m \in \mathcal{M}$ reliably.

- *Last bloc.* Encoder sends a sequence of symbols $X_B^n \in \mathcal{X}^n$ jointly typical with the sequences $(U_B^n, S_B^n, V_B^n) \in \mathcal{U}^n \times \mathcal{S}^n \times \mathcal{V}^n$. Decoder returns the sequence V_B^n . The sequences $(U_B^n, S_B^n, Z_B^n, X_B^n, Y_B^n, V_B^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical over the last block.

For each block $b \in \{2, \dots, B\}$, the properties of typical sequences, packing and covering lemmas stated in [9] pp. 27, 46 and 208, equations (189), (190), (191) imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (197)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, \quad ((U_{b+1}^n, S_{b+1}^n, V_{b+1}^n(m)) \notin A_\varepsilon^{*n}(\mathcal{Q})) \right) \right] \leq \varepsilon, \quad (198)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall l \in \mathcal{M}_L, \quad (U_b^n, S_b^n, V_b^n, W_{1,b}^n(m, l)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (199)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists (m', l') \neq (m, l), \text{ s.t. } (Y_b^n, Z_b^n, V_b^n, W_{1,b}^n(m', l')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon. \quad (200)$$

For each block $b \in \{2, \dots, B\}$, for all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U_b^n, S_b^n, Z_b^n, W_{1,b}^n, X_b^n, Y_b^n, V_b^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for distribution $\mathcal{P}_{usz}(u, s, z) \otimes \mathcal{Q}(x, v|u, s) \otimes \mathcal{Q}(w_1|u, s, x, v) \otimes \mathcal{T}(y|x, s)$ with probability more than $1 - 4\varepsilon$.

The remaining of the proof is very similar to the achievability part of Theorem V.1. The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem VII.1.

B. Converse Proof

We consider the joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ and we introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{\star n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{\star n}(\mathcal{Q}). \end{cases} \quad (201)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, v)$, i.e. for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ goes to zero. We have the following equations:

$$0 = \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}; U_i, S_i | U_{i+1}^n, S_{i+1}^n) \quad (202)$$

$$= \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n; U_i, S_i) \quad (203)$$

$$= \sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}, V_i) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n, V_i; U_i, S_i) \quad (204)$$

$$\leq \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | V_i) - \sum_{i=1}^n I(Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n, V_i; U_i, S_i) \quad (205)$$

$$= \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | V_i) - \sum_{i=1}^n I(W_{1,i}, V_i; U_i, S_i). \quad (206)$$

Equation (202) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (203) comes from the i.i.d. property of the information source (U, S) that implies $I(U_i, S_i; U_{i+1}^n, S_{i+1}^n) = 0$ for all $i \in \{1, \dots, n\}$.

Equation (204) comes from the strictly causal decoding that implies V_i is a deterministic function of (Y^{i-1}, Z^{i-1}) . Hence, for all $i \in \{1, \dots, n\}$ we have:

$$I(V_i; U_i, S_i | Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n) = 0, \quad (207)$$

$$I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) - I(U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}, V_i) \quad (208)$$

$$= I(V_i; U_{i+1}^n, S_{i+1}^n; Y_i, Z_i | Y^{i-1}, Z^{i-1}) = 0. \quad (209)$$

Equation (205) comes from the properties of the mutual information.

Equation (206) comes from the introduction of the auxiliary random variable $W_{1,i} = (Y^{i-1}, Z^{i-1}, U_{i+1}^n, S_{i+1}^n)$ that satisfies the Markov Chains corresponding to the set of probability distributions \mathbb{Q}_{sd} :

$$Z_i \text{---} (U_i, S_i) \text{---} (X_i, Y_i, W_{1,i}, V_i), \quad (210)$$

$$Y_i \text{---} (X_i, S_i) \text{---} (U_i, Z_i, W_{1,i}, V_i). \quad (211)$$

- The first Markov chain comes from i.i.d. property of the source and the fact that Z_i does not belong to $W_{1,i}$.
- The second Markov chain comes from memoryless property of the channel and the fact that Y_i does not belong to $W_{1,i}$.

$$0 = \sum_{i=1}^n I(W_{1,i}; Y_i, Z_i | V_i) - \sum_{i=1}^n I(W_{1,i}, V_i; U_i, S_i) \quad (212)$$

$$= n \cdot \left(I(W_{1,T}; Y_T, Z_T | V_T, T) - I(W_{1,T}, V_T; U_T, S_T | T) \right) \quad (213)$$

$$\leq n \cdot \left(I(W_{1,T}, T; Y_T, Z_T | V_T) - I(W_{1,T}, T, V_T; U_T, S_T) \right) \quad (214)$$

$$\leq n \cdot \max_{\mathbb{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | V) - I(W_1, V; U_T, S_T) \right) \quad (215)$$

$$\leq n \cdot \max_{\mathbb{Q} \in \mathbb{Q}} \left(I(W_1; Y_T, Z_T | V, E = 0) - I(W_1, V; U_T, S_T | E = 0) + \varepsilon \right) \quad (216)$$

$$\leq n \cdot \max_{\mathbb{Q} \in \mathbb{Q}} \left(I(W_1; Y, Z | V) - I(W_1, V; U, S) + 2\varepsilon \right). \quad (217)$$

Equation (213) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, X_T, Y_T, V_T$.

Equation (214) comes from the i.i.d. property of the information source that implies $I(T; U_T, S_T) = 0$.

Equation (215) comes from identifying W_1 with $(W_{1,T}, T)$ and taking the maximum over the probability distributions that belong to \mathbb{Q}_{sd} . This is made possible since the pair of random variables $(W_{1,T}, T)$ satisfies the three Markov chains of the set of probability distributions \mathbb{Q}_{sd} , as stated in Lemma 7 in the Appendix.

Equation (216) comes from the empirical coordination requirement as stated in Lemma 8 in the Appendix. Sequences are not jointly typical with small error probability $\mathcal{P}(E = 1)$.

Equation (217) comes from Lemma 9 that states the probability distribution induced by the coding scheme $\mathcal{P}\left((U_T, S_T, Z_T, X_T, Y_T, V_T) = (u, s, z, x, y, v) | E = 0\right)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes.

This concludes the converse proof of Theorem VII.1.

VIII. STRICTLY CAUSAL ENCODING

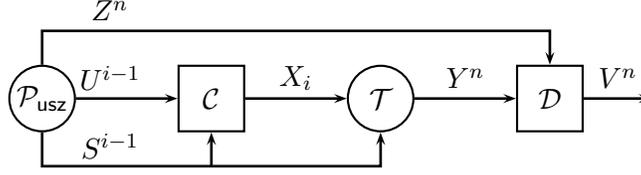


Fig. 8. Strictly causal encoding function $f_i : \mathcal{U}^{i-1} \times \mathcal{S}^{i-1} \rightarrow \mathcal{X}$, for all $i \in \{1, \dots, n\}$ and non-causal decoding function $g : \mathcal{Y}^n \times \mathcal{Z}^n \rightarrow \mathcal{V}^n$.

Theorem VIII.1 (Strictly Causal Encoding)

1) Joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ is achievable if and only if it decomposes as follows:

$$\left\{ \begin{array}{l} \mathcal{Q}(u, s, z) = \mathcal{P}_{\text{usz}}(u, s, z), \\ \mathcal{Q}(y|x, s) = \mathcal{T}(y|x, s), \\ (U, S) \text{ independent of } X, \\ Y \text{ --- } (X, S) \text{ --- } (U, Z), \\ Z \text{ --- } (U, S) \text{ --- } (X, Y). \end{array} \right. \quad (218)$$

and $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is achievable.

2) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is achievable if:

$$\max_{\mathcal{Q} \in \mathcal{Q}_{\text{se}}} \left(I(X, W_2; Y, Z) - I(W_2; U, S|X) \right) > 0, \quad (219)$$

3) Joint probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ is not achievable if:

$$\max_{\mathcal{Q} \in \mathcal{Q}_{\text{se}}} \left(I(X, W_2; Y, Z) - I(W_2; U, S|X) \right) < 0, \quad (220)$$

where \mathcal{Q}_{se} is the set of probability distributions $\mathcal{Q} \in \Delta(\mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{W}_2 \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V})$ with auxiliary random variable W_2 that satisfies:

$$\left\{ \begin{array}{l} \sum_{w_2 \in \mathcal{W}_2} \mathcal{Q}(u, s, z, w_2, x, y, v) \\ = \mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y), \\ (U, S) \text{ independent of } X, \\ Y \text{ --- } (X, S) \text{ --- } (U, Z, W_2), \\ Z \text{ --- } (U, S) \text{ --- } (X, Y, W_2), \\ V \text{ --- } (Y, Z, X, W_2) \text{ --- } (U, S). \end{array} \right.$$

The probability distribution $\mathcal{Q} \in \mathbb{Q}_{\text{se}}$ decomposes as follows:

$$\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{Q}(w_2|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, x, w_2).$$

The supports of the auxiliary random variable W_2 is bounded by $|\mathcal{W}_2| \leq |\mathcal{B}| + 1$ with $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$.

Remark VIII.2 Strictly causal encoding requires the channel input X to be independent of the source and state (U, S) . Hence, the general probability distributions $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x|u, s) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, z, x, y)$ reduces to $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|u, s, x, y, z)$.

Remark VIII.3 This result was already stated in [1] without considering state informations at the encoder S and at the decoder Z .

A. Achievability Proof

The achievability proof is very similar to the one of Theorem VI.1 for causal encoding replacing W_1 by X . We consider a probability distribution $\mathcal{Q} \in \mathcal{Q}_{\text{se}}$ that achieves the maximum in equation (219). There exists a $\delta > 0$ and a rate $R > 0$ such that:

$$R \geq I(W_2; U, S|X) + \delta, \quad (221)$$

$$R \leq I(X; Y, Z) + I(W_2; Y, Z|X) - \delta = I(X, W_2; Y, Z) - \delta. \quad (222)$$

We consider a block-Markov random code $c \in \mathcal{C}(n)$ defined over $B \in \mathbb{N}$ blocs of length $n \in \mathbb{N}$. The total length of the code is denoted by $N = n \cdot B \in \mathbb{N}$ and R denotes the rate of the code.

- *Random codebook.* We generate $|\mathcal{M}| = 2^{nR}$ sequences $X^n(m)$ drawn from the i.i.d. probability distribution $\mathcal{Q}_x^{\otimes n}$ with index $m \in \mathcal{M}$. For each index $m \in \mathcal{M}$, we generate the same number $|\mathcal{M}| = 2^{nR}$ of sequences $W_2^n(m, \hat{m})$ with index $\hat{m} \in \mathcal{M}$, drawn from the i.i.d. conditional probability distribution $\mathcal{Q}_{w_2|x}^{\otimes n}$ depending on sequence $X^n(m)$.
- *Encoding function.* At the beginning of block $b \in \{2, \dots, B-1\}$, the encoder observes the sequences of source symbols $(U_{b-1}^n, S_{b-1}^n) \in \mathcal{U}^n \times \mathcal{S}^n$ of the previous block $b-1$. It also recalls the index $m_{b-1} \in \mathcal{M}$ of the sequence $X^n(m_{b-1}) \in \mathcal{X}^n$ over block $b-1$. It finds index $m_b \in \mathcal{M}$ such that the sequences $(U_{b-1}^n, S_{b-1}^n, X^n(m_{b-1}), W_2^n(m_{b-1}, m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the sequence $X^n(m_b)$ corresponding to the current block $b \in \{2, \dots, B-1\}$.
- *Decoding function.* At the end of block $b \in \{2, \dots, B-1\}$, the decoder recalls sequences (Y_{b-1}^n, Z_{b-1}^n) and the index $m_{b-1} \in \mathcal{M}$ corresponding to the sequence $X^n(m_{b-1})$. It observes the sequences (Y_b^n, Z_b^n) and finds index $m_b \in \mathcal{M}$ such that $(Y_b^n, Z_b^n, X^n(m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ and $(Y_{b-1}^n, Z_{b-1}^n, X^n(m_{b-1}), W_2^n(m_{b-1}, m_b)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Decoder returns the sequence V_{b-1}^n drawn from the conditional probability distribution $\mathcal{Q}_{v|yzxw_2}^{\otimes n}$ depending on sequences $(Y_{b-1}^n, Z_{b-1}^n, X^n(m_{b-1}), W_2^n(m_{b-1}, m_b))$.
- *First bloc at the encoder.* An arbitrary index $m_1 \in \mathcal{M}$ is given to both encoder \mathcal{C} and decoder \mathcal{D} . Encoder sends the corresponding sequence $X^n(m_1) \in \mathcal{X}^n$. At the beginning of the second bloc b_2 , encoder recalls $(U_{b_1}^n, S_{b_1}^n, X^n(m_1))$ and finds index m_2 such that sequences $(U_{b_1}^n, S_{b_1}^n, X^n(m_1), W_2^n(m_1, m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Encoder sends the corresponding

sequence $X^n(m_2)$ on the second block b_2 .

- *First bloc at the decoder.* At the end of second block b_2 , the decoder finds the index m_2 such that $(Y_{b_2}^n, Z_{b_2}^n, X^n(m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ and $(Y_{b_1}^n, Z_{b_1}^n, X^n(m_1), W_2^n(m_1, m_2)) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical. Over the first bloc, decoder \mathcal{D} returns $V_{b_1}^n \in \mathcal{V}^n$ drawn from the conditional probability distribution $\mathcal{Q}_{v|yzxw_2}^{\otimes n}$ depending on sequences $(Y_{b_1}^n, Z_{b_1}^n, X^n(m_1), W_2^n(m_1, m_2))$. Sequences $(U_{b_1}^n, S_{b_1}^n, Z_{b_1}^n, X^n(m_1), W_2^n(m_1, m_2), Y_{b_1}^n, V_{b_1}^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical over the first block b_1 .
- *Last bloc.* Encoder \mathcal{C} and decoder \mathcal{D} choose arbitrary sequences X_B^n and V_B^n . Sequences are not jointly typical on the last block.

For each block $b \in \{1, \dots, B-1\}$, the properties of typical sequences, packing and covering lemmas stated in [9] pp. 27, 46 and 208, equations (221), (221) imply there exists a $\bar{n} \in \mathbb{N}$ such that the expected probability of error events are bounded by ε for all $n \geq \bar{n}$:

$$\mathbb{E}_c \left[\mathcal{P} \left((U^n, S^n) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (223)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\forall m \in \mathcal{M}, (U_{b-1}^n, S_{b-1}^n, X^n(m_{b-1}), W_2^n(m_{b-1}, m)) \notin A_\varepsilon^{*n}(\mathcal{Q}) \right) \right] \leq \varepsilon, \quad (224)$$

$$\mathbb{E}_c \left[\mathcal{P} \left(\exists m' \neq m, \text{ s.t. } \left\{ (Y_b^n, Z_b^n, X^n(m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \cap \left\{ (Y_{b-1}^n, Z_{b-1}^n, X^n(m_{b-1}), W_2^n(m_{b-1}, m')) \in A_\varepsilon^{*n}(\mathcal{Q}) \right\} \right) \right] \leq \varepsilon. \quad (225)$$

For each block $b \in \{1, \dots, B-1\}$, for all $n \geq \bar{n}$, there exists a code $c^* \in \mathcal{C}(n)$ such that sequences $(U_b^n, S_b^n, Z_b^n, X^n(m_b), W_2^n(m_b, m_{b+1}), Y_b^n, V_b^n) \in A_\varepsilon^{*n}(\mathcal{Q})$ are jointly typical for the probability distribution $\mathcal{P}_{\text{usz}}(u, s, z) \otimes \mathcal{Q}(x) \otimes \mathcal{Q}(w_2|u, s, x) \otimes \mathcal{T}(y|x, s) \otimes \mathcal{Q}(v|y, z, x, w_2)$ with probability more than $1 - 3\varepsilon$.

The remaining of the proof is very similar to the achievability part of Theorem VI.1. Lemma 2 also proves that equation (222) implies equation (225). The cardinality bound is stated in Lemma 6 in the Appendix. This concludes the achievability proof of Theorem VIII.1.

B. Converse Proof

We consider the joint probability distribution $\mathcal{Q}(u, s, z, x, y, v)$ and we introduce the random event of error $E \in \{0, 1\}$ defined as follows:

$$E = \begin{cases} 0 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} \leq \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \in A_\varepsilon^{*n}(\mathcal{Q}), \\ 1 & \text{if } \|Q^n - \mathcal{Q}\|_{\text{tv}} > \varepsilon \iff (U^n, S^n, Z^n, X^n, Y^n, V^n) \notin A_\varepsilon^{*n}(\mathcal{Q}). \end{cases} \quad (226)$$

Consider a sequence of code $c(n) \in \mathcal{C}$ that achieves the probability distribution $\mathcal{Q}(u, s, z, x, y, v)$, i.e. for which the probability of error $\mathcal{P}_e(c) = \mathcal{P}(E = 1)$ goes to zero. The converse is based on the following equations:

$$0 = \sum_{i=1}^n I(U^{i-1}, S^{i-1}; Y_i, Z_i | Y_{i+1}^n, Z_{i+1}^n) - \sum_{i=1}^n I(Y_{i+1}^n, Z_{i+1}^n; U_i, S_i | U^{i-1}, S^{i-1}) \quad (227)$$

$$\leq \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n; Y_i, Z_i) - \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n; U_i, S_i) \quad (228)$$

$$= \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, X_i; Y_i, Z_i) - \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, X_i; U_i, S_i) \quad (229)$$

$$\leq \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, X_i; Y_i, Z_i) - \sum_{i=1}^n I(U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n; U_i, S_i | X_i) \quad (230)$$

$$\leq \sum_{i=1}^n I(X_i, W_{2,i}; Y_i, Z_i) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | X_i). \quad (231)$$

Equation (227) comes from Csiszár Sum Identity stated pp. 25 in [9].

Equation (228) comes from the i.i.d. property of the information source (U, S) that implies $I(U^{i-1}, S^{i-1}; U_i, S_i) = 0$ for all $i \in \{1, \dots, n\}$.

Equation (229) comes from the strictly causal encoding function $X_i = f_i(U^{i-1}, S^{i-1})$ that implies $I(X_i; Y_i, Z_i | U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) = I(X_i; U_i, S_i | U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) = 0$ for all $i \in \{1, \dots, n\}$.

Equation (230) comes from the properties of the mutual information.

Equation (236) comes from the introduction of the auxiliary random variable $W_{2,i} = (U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n)$. For all $i \in \{1, \dots, n\}$, auxiliary random variable $W_{2,i}$ satisfies the properties corresponding to the set of probability distributions \mathbb{Q}_{se} :

$$(U_i, S_i) \text{ are independent of } X_i, \quad (232)$$

$$Y_i \text{ --- } (X_i, S_i) \text{ --- } (U_i, Z_i, W_{2,i}), \quad (233)$$

$$Z_i \text{ --- } (U_i, S_i) \text{ --- } (X_i, Y_i, W_{2,i}), \quad (234)$$

$$V_i \text{ --- } (Y_i, Z_i, X_i, W_{2,i}) \text{ --- } (U_i, S_i). \quad (235)$$

- Equation (232) comes from the strictly causal encoding property that implies X_i is independent of (U_i, S_i) .
- Equation (233) comes from the memoryless property of the channel and the fact that Y_i is not included in $W_{2,i}$.
- Equation (234) comes from the i.i.d. property of the source and states (U_i, S_i, Z_i) and the fact that Z_i is not included in $W_{2,i}$.
- Equation (235) comes from the strictly causal encoding and the non-causal decoding as stated in Lemma 4.

Equation (230) gives:

$$0 \leq \sum_{i=1}^n I(X_i, W_{2,i}; Y_i, Z_i) - \sum_{i=1}^n I(W_{2,i}; U_i, S_i | X_i)$$

$$= n \cdot \left(I(X_T, W_{2,T}; Y_T, Z_T | T) - I(W_{2,T}; U_T, S_T | X_T, T) \right) \quad (236)$$

$$\leq n \cdot \left(I(X_T, W_{2,T}, T; Y_T, Z_T) - I(W_{2,T}; U_T, S_T | X_T, T) \right) \quad (237)$$

$$= n \cdot \left(I(X_T, W_{2,T}, T; Y_T, Z_T) - I(W_{2,T}, X_T, T; U_T, S_T) \right) \quad (238)$$

$$\leq n \cdot \left(I(X_T, W_{2,T}, T; Y_T, Z_T) - I(W_{2,T}, T; U_T, S_T | X_T) \right) \quad (239)$$

$$\leq n \cdot \max_{Q \in \mathbb{Q}} \left(I(X_T, W_2; Y_T, Z_T) - I(W_2; U_T, S_T | X_T) \right) \quad (240)$$

$$\leq n \cdot \max_{Q \in \mathbb{Q}} \left(I(X_T, W_2; Y_T, Z_T | E = 0) - I(W_2; U_T, S_T | X_T, E = 0) + \varepsilon \right) \quad (241)$$

$$= n \cdot \max_{Q \in \mathbb{Q}} \left(I(X, W_2; Y, Z) - I(W_2; U, S | X) + 2\varepsilon \right). \quad (242)$$

Equation (236) comes from the introduction of the uniform random variable T over $\{1, \dots, n\}$ and the introduction of the corresponding mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (237) comes from the properties of the mutual information.

Equation (238) comes from the independence between (U_T, S_T) and (X_T, T) that implies $I(X_T, T; U_T, S_T) = 0$ as stated in Lemma 5.

Equation (239) comes from the properties of the mutual information.

Equation (240) comes from identifying W_2 with $(W_{2,T}, T)$ and taking the maximum over the probability distributions that belong to the set \mathbb{Q}_{se} . This is possible since the pair of random variables $(W_{2,T}, T)$ satisfies the properties of the set of probability distributions \mathbb{Q}_{se} as stated in Lemma 7 in the Appendix.

Equation (241) comes from the empirical coordination requirement as stated in Lemma 8 in the Appendix. Sequences are not jointly typical with small error probability $\mathcal{P}(E = 1)$.

Equation (242) comes from Lemma 9 that states that the probability distribution induced by the coding scheme $\mathcal{P}((U_T, S_T, Z_T, X_T, Y_T, V_T) = (u, s, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(u, s, z, x, y, v)$. The continuity of the entropy function stated pp. 33 in [10] concludes.

This concludes the proof of Theorem VIII.1.

Lemma 4 Markov chain $V_i \text{---} (Y_i, Z_i, X_i, W_{2,i}) \text{---} (U_i, S_i)$ is satisfied for all $i \in \{1, \dots, n\}$.

Proof VIII.4 (Lemma 4) We evaluate the following probability:

$$\begin{aligned}
& \mathcal{P}(V_i | X_i, Y_i, Z_i, W_{2,i}, U_i, S_i) \\
&= \mathcal{P}(V_i | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(V_i, X^{i-1}, Y^{i-1}, Z^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i) \quad (243) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(Z^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i) \\
&\times \mathcal{P}(X^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}) \\
&\times \mathcal{P}(Y^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}, X^{i-1}) \\
&\times \mathcal{P}(V_i | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}, X^{i-1}, Y^{i-1}). \quad (244)
\end{aligned}$$

We can remove (U_i, S_i) in the four probability distributions:

$$\mathcal{P}(Z^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i) = \mathcal{P}(Z^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n), \quad (245)$$

$$\mathcal{P}(X^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}) = \mathcal{P}(X^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}), \quad (246)$$

$$\mathcal{P}(Y^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}, X^{i-1}) = \mathcal{P}(Y^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}, X^{i-1}), \quad (247)$$

$$\mathcal{P}(V_i | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, U_i, S_i, Z^{i-1}, X^{i-1}, Y^{i-1}) = \mathcal{P}(V_i | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n, Z^{i-1}, X^{i-1}, Y^{i-1}) \quad (248)$$

Equation (245) comes from the i.i.d. property of the information source: Z^{i-1} only depends on (U^{i-1}, S^{i-1}) .

Equation (246) comes from the strictly causal encoding: X^{i-1} is a deterministic function of (U^{i-2}, S^{i-2}) included in (U^{i-1}, S^{i-1}) .

Equation (247) comes from the memoryless property of the channel: Y^{i-1} only depends only on (X^{i-1}, S^{i-1}) .

Equation (248) comes from the non-causal decoding: V_i is a deterministic function of $(Y^n, Z^n) = (Y^{i-1}, Y_i, Y_{i+1}^n, Z^{i-1}, Z_i, Z_{i+1}^n)$.

Hence we have:

$$\begin{aligned}
& \mathcal{P}(V_i | X_i, Y_i, Z_i, W_{2,i}, U_i, S_i) \\
&= \sum_{X^{i-1}, Y^{i-1}, Z^{i-1}} \mathcal{P}(V_i, X^{i-1}, Y^{i-1}, Z^{i-1} | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) \quad (249)
\end{aligned}$$

$$= \mathcal{P}(V_i | X_i, Y_i, Z_i, U^{i-1}, S^{i-1}, Y_{i+1}^n, Z_{i+1}^n) \quad (250)$$

$$= \mathcal{P}(V_i | X_i, Y_i, Z_i, W_{2,i}). \quad (251)$$

The above equation corresponds to the Markov chain $V_i \ominus (Y_i, Z_i, X_i, W_{2,i}) \ominus (U_i, S_i)$ and it concludes the proof of Lemma 4.

Lemma 5 *Independence of random variable (X_T, T) with (U_T, S_T) induces the following equation:*

$$I(X_T, T; U_T, S_T) = 0. \quad (252)$$

Proof VIII.5 (Lemma 5) The i.i.d. property of the source (U, S) implies that for all $i \in \{1, \dots, n\}$, we have:

$$\mathcal{P}\left((U_T, S_T) = (u, s) | T = i, X_T = x\right) = \mathcal{P}\left((U_i, S_i) = (u, s) | T = i, X_i = x\right) \quad (253)$$

$$= \mathcal{P}\left((U_i, S_i) = (u, s)\right) \quad (254)$$

$$= \mathcal{P}\left((U_T, S_T) = (u, s)\right). \quad (255)$$

Equation (253) comes from the definition of the mean random variables (U_T, S_T, X_T) and of T .

Equation (254) comes from the i.i.d. property of the information source and the strictly causal encoding that induces the independence between (U_i, S_i) and (T, X_i) .

Equation (255) comes from the definition of the mean random variables (U_T, S_T) .

This implies directly: $I(X_T, T; U_T, S_T) = 0$.

APPENDIX

Lemma 6 (Cardinality Bound) We denote by \mathcal{B} the discrete set $\mathcal{B} = \mathcal{U} \times \mathcal{S} \times \mathcal{Z} \times \mathcal{X} \times \mathcal{Y} \times \mathcal{V}$. We consider the following information constraint with one auxiliary random variable $W = (W_1, W_2)$:

$$\max_{Q \in \mathbb{Q}} \left(I(W; Y, Z) - I(W; U, S) \right) > 0.$$

The cardinality of the support $|\mathcal{W}|$ of the auxiliary random variable W is bounded by

$$|\mathcal{W}| \leq |\mathcal{B}| + 1.$$

This result is based on the Lemma of Fenchel-Eggleston-Carathéodory. More details are provided pp. 631 in [9]. Considering the case of two auxiliary random variables W_1 and W_2 , the cardinality of the supports $|\mathcal{W}_1|$ and $|\mathcal{W}_2|$ can be bounded by:

$$\max(|\mathcal{W}_1|, |\mathcal{W}_2|) \leq (|\mathcal{B}| + 1) \cdot (|\mathcal{B}| + 2).$$

Proof A.1 Lemma 6

We denote by $d = |\mathcal{B}| + 1$, the cardinality of the family $\{h_i\}_{i=1}^d$ of continuous functions defined from $\Delta(\mathcal{B})$ into \mathbb{R} as follows:

$$h_i \left(\mathcal{P}_{\text{uszyv}|w} \right) = \begin{cases} \mathcal{P}_{\text{uszyv}|w}(i), & \text{for } i \in \{1, \dots, |\mathcal{B}| - 1\}, \\ H(Y, Z|W = w), & \text{for } i = |\mathcal{B}|, \\ H(U, S|W = w), & \text{for } i = |\mathcal{B}| + 1. \end{cases}$$

Support Lemma stated pp. 631 in [9], implies that there exists an auxiliary random variable $W' \sim \mathcal{P}_{w'}$ defined on a set \mathcal{W}' with finite cardinality $|\mathcal{W}'| \leq d + 1$ such that for all $i \in \{1, \dots, d\}$ we have:

$$\int_{\mathcal{W}} h_i \left(\mathcal{P}_{\text{uszyv}|w} \right) dF(w) = \sum_{w' \in \mathcal{W}'} h_i \left(\mathcal{P}_{\text{uszyv}|w'} \right) \mathcal{P}(w').$$

This implies that the probability $\mathcal{P}_{\text{uszyv}}$ is preserved and $H(Y, Z|W)$ and $H(U, S|W)$ are equal to $H(Y, Z|W')$ and $H(U, S|W')$:

$$\begin{aligned} \mathcal{P}_{\text{uszyv}}(i) &= \int_{\mathcal{W}} \mathcal{P}_{\text{uszyv}|w}(i) dF(w) = \sum_{w' \in \mathcal{W}'} \mathcal{P}_{\text{uszyv}|w'}(i) \cdot \mathcal{P}(w'), \quad \text{for } i \in \{1, \dots, |\mathcal{B}| - 1\} \\ H(Y, Z|W) &= \int_{\mathcal{W}} H(Y, Z|W = w) dF(w) = \sum_{w' \in \mathcal{W}'} H(Y, Z|W' = w') \cdot \mathcal{P}(w') = H(Y, Z|W'), \\ H(U, S|W) &= \int_{\mathcal{W}} H(U, S|W = w) dF(w) = \sum_{w' \in \mathcal{W}'} H(U, S|W' = w') \cdot \mathcal{P}(w') = H(U, S|W'). \end{aligned}$$

Hence the information constraint remains equal with $|\mathcal{W}'| \leq d = |\mathcal{B}| + 1$.

$$\begin{aligned}
 & I(W; Y, Z) - I(W; U, S) \\
 = & H(Y, Z) - H(Y, Z|W) - H(U, S) + H(U, S|W) \\
 = & H(Y, Z) - H(Y, Z|W') - H(U, S) + H(U, S|W') \\
 = & I(W'; Y, Z) - I(W'; U, S),
 \end{aligned}$$

This concludes the proof of the bound on the cardinality of the support of the auxiliary random variable.

Lemma 7 *The random variables $W_1 = (W_{1,T}, T)$ and $W_2 = W_{2,T}$ satisfy the properties of the set of probability distributions \mathbb{Q} corresponding to Theorem I.1.*

This result extends to Theorem II.1, III.1, IV.1, V.1 and VI.1.

Proof A.2 (Lemma 7) The Markov chains stated in equations (13) - (15) are valid for all $i \in \{1, \dots, n\}$. The definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$, the i.i.d. property of the source and the memoryless property of the channel implies directly the result of Lemma 7. The details are provided below.

$$Y_T \text{---} (X_T, S_T) \text{---} (U_T, Z_T, W_{1,T}, T, W_{2,T}), \quad (256)$$

$$Z_T \text{---} (U_T, S_T) \text{---} (X_T, Y_T, W_{1,T}, T, W_{2,T}), \quad (257)$$

$$V_T \text{---} (Y_T, Z_T, W_{1,T}, T, W_{2,T}) \text{---} (U_T, S_T, X_T). \quad (258)$$

- Equation (256) comes from the memoryless property of the channel and the fact that Y_T is not included in $(W_{1,T}, T, W_{2,T})$ for all realization $T = i \in \{1, \dots, n\}$:

$$\begin{aligned} & \mathcal{P}(Y_T = y | X_T = x, S_T = s, U_T = u, Z_T = z, W_{1,T} = w_1, T = i, W_{2,T} = w_2) \\ &= \mathcal{P}(Y_T = y | X_T = x, S_T = s). \end{aligned} \quad (259)$$

- Equation (257) comes from the i.i.d. property of the source and the fact that Z_T is not included in $(W_{1,T}, W_{2,T})$ for all realization $T = i \in \{1, \dots, n\}$:

$$\begin{aligned} & \mathcal{P}(Z_T = z | U_T = u, S_T = s, X_T = x, Y_T = y, W_{1,T} = w_1, T = i, W_{2,T} = w_2) \\ &= \mathcal{P}(Z_T = z | U_T = u, S_T = s). \end{aligned} \quad (260)$$

- Equation (258) comes from the following equations:

$$\begin{aligned} & \mathcal{P}(V_T = v | Y_T = y, Z_T = z, W_{1,T} = w_1, T = i, W_{2,T} = w_2, U_T = u, S_T = s, X_T = x) \\ &= \mathcal{P}(V_i = v | Y_i = y, Z_i = z, W_{1,i} = w_1, T = i, W_{2,i} = w_2, U_i = u, S_i = s, X_i = x) \end{aligned} \quad (261)$$

$$= \mathcal{P}(V_i = v | Y_i = y, Z_i = z, W_{1,i} = w_1, T = i, W_{2,i} = w_2) \quad (262)$$

$$= \mathcal{P}(V_T = v | Y_T = y, Z_T = z, W_{1,T} = w_1, T = i, W_{2,T} = w_2). \quad (263)$$

Equation (261) comes from the definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (262) comes from the Markov chain property $V_i \ominus (Y_i, Z_i, W_{1,i}, W_{2,i}) \ominus (U_i, S_i, X_i)$ that is valid for all $i \in \{1, \dots, n\}$.

Equation (263) comes from the definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

This concludes the proof of Lemma 7.

Regarding the extension to Theorem VI.1, the Markov chains stated in equations (168) - (169) are valid for all $i \in \{1, \dots, n\}$.

$$(U_T, S_T) \text{ are independent of } (W_{1,T}, T), \quad (264)$$

$$X_T \ominus (U_T, S_T, W_{1,T}) \ominus W_{2,T}. \quad (265)$$

- Equation (264) comes from the following equations:

$$\mathcal{P}(U_T = u, S_T = s | W_{1,T} = w_1, T = i) = \mathcal{P}(U_i = u, S_i = s | W_{1,i} = w_1, T = i) \quad (266)$$

$$= \mathcal{P}(U_i = u, S_i = s) \quad (267)$$

$$= \mathcal{P}(U_T = u, S_T = s). \quad (268)$$

Equation (266) comes from the definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (267) comes from the independence of the source (U_i, S_i) with respect to T and $W_{1,i}$.

Equation (268) comes from the i.i.d. property of the source.

- Equation (265) comes from the following equations:

$$\begin{aligned} & \mathcal{P}(X_T = x | U_T = u, S_T = s, W_{1,T} = w_1, T = i, W_{2,T} = w_2) \\ &= \mathcal{P}(X_i = x | U_i = u, S_i = s, W_{1,i} = w_1, T = i, W_{2,i} = w_2) \end{aligned} \quad (269)$$

$$= \mathcal{P}(X_i = x | U_i = u, S_i = s, W_{1,i} = w_1, T = i) \quad (270)$$

$$= \mathcal{P}(X_T = x | U_T = u, S_T = s, W_{1,T} = w_1, T = i). \quad (271)$$

Equation (269) comes from the definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Equation (270) comes from the Markov chain property $X_i \ominus (U_i, S_i, W_{1,i}) \ominus W_{2,i}$ that is valid for all $i \in \{1, \dots, n\}$.

Equation (271) comes from the definition of the mean random variables $U_T, S_T, Z_T, W_{1,T}, W_{2,T}, X_T, Y_T, V_T$.

Lemma 8 Fix a probability distribution $\mathcal{Q} \in \mathbb{Q}$ and suppose that the error probability $\mathcal{P}(E = 1)$ is small enough such that $\mathcal{P}(E = 1) \cdot \log_2 |\mathcal{Y} \times \mathcal{Z}| + 2 \cdot h_b(\mathcal{P}(E = 1)) \leq \varepsilon$. Then we have:

$$I(W_1; Y_T, Z_T | W_2) - I(W_2; U_T, S_T | W_1) \quad (272)$$

$$\leq I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_2; U_T, S_T | W_1, E = 0) + \varepsilon. \quad (273)$$

The proof of Lemma 8 extends to the following equations that intervene in the proof of Theorems V.1, VI.1, VII.1 and VIII.1.

Proof A.3 Lemma 8 comes from the properties of the mutual information.

$$I(W_1; Y_T, Z_T | W_2) - I(W_2; U_T, S_T | W_1) \quad (274)$$

$$= I(W_1; Y_T, Z_T | W_2, E) - I(W_2; U_T, S_T | W_1, E) \quad (275)$$

$$+ I(E; Y_T, Z_T | W_2) - I(E; Y_T, Z_T | W_1, W_2) - I(E; U_T, S_T | W_1) + I(E; U_T, S_T | W_1, W_2) \quad (276)$$

$$\leq I(W_1; Y_T, Z_T | W_2, E) - I(W_2; U_T, S_T | W_1, E) + 2H(E) \quad (277)$$

$$= \mathcal{P}(E = 0) \cdot \left(I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_2; U_T, S_T | W_1, E = 0) \right) \\ + \mathcal{P}(E = 1) \cdot \left(I(W_1; Y_T, Z_T | W_2, E = 1) - I(W_2; U_T, S_T | W_1, E = 1) \right) + 2H(E) \quad (278)$$

$$\leq I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_2; U_T, S_T | W_1, E = 0) \\ + \mathcal{P}(E = 1) \cdot \log_2 |\mathcal{Y} \times \mathcal{Z}| + 2 \cdot h_b(\mathcal{P}(E = 1)) \quad (279)$$

$$\leq I(W_1; Y_T, Z_T | W_2, E = 0) - I(W_2; U_T, S_T | W_1, E = 0) + \varepsilon. \quad (280)$$

This concludes the proof of Lemma 8.

Lemma 9 Probability distribution $\mathcal{P}((S_T, U_T, Z_T, X_T, Y_T, V_T) = (s, u, z, x, y, v) | E = 0)$ is closed to the target probability distribution $\mathcal{Q}(s, u, z, x, y, v)$:

$$\left| \mathcal{P}((S_T, U_T, Z_T, X_T, Y_T, V_T) = (s, u, z, x, y, v) | E = 0) - \mathcal{Q}(s, u, z, x, y, v) \right| \leq \varepsilon. \quad (281)$$

Proof A.4 (Proof of Lemma 9) We evaluate the probability $\mathcal{P}(S_T = s | E = 0)$ and we show it is closed to the desired probability $\mathcal{P}_s(s)$:

$$\mathcal{P}(S_T = s | E = 0) = \sum_{s^n \in A_\varepsilon^{*n}} \sum_{i=1}^n \mathcal{P}(S^n = s^n, T = i, S_T = s | E = 0) \quad (282)$$

$$\begin{aligned} &= \sum_{s^n \in A_\varepsilon^{*n}} \sum_{i=1}^n \mathcal{P}(S^n = s^n | E = 0) \\ &\times \mathcal{P}(T = i | S^n = s^n, E = 0) \cdot \mathcal{P}(S_T = s | S^n = s^n, T = i, E = 0) \end{aligned} \quad (283)$$

$$\begin{aligned} &= \sum_{s^n \in A_\varepsilon^{*n}} \sum_{i=1}^n \mathcal{P}(S^n = s^n | E = 0) \\ &\times \mathcal{P}(T = i) \cdot \mathcal{P}(S_T = s | S^n = s^n, T = i, E = 0) \end{aligned} \quad (284)$$

$$= \sum_{s^n \in A_\varepsilon^{*n}} \sum_{i=1}^n \mathcal{P}(S^n = s^n | E = 0) \cdot \mathcal{P}(T = i) \cdot \mathbb{1}_{\{s_T = s\}} \quad (285)$$

$$= \sum_{s^n \in A_\varepsilon^{*n}} \mathcal{P}(S^n = s^n | E = 0) \cdot \sum_{i=1}^n \frac{1}{n} \cdot \mathbb{1}_{\{s_T = s\}} \quad (286)$$

$$= \sum_{s^n \in A_\varepsilon^{*n}} \mathcal{P}(S^n = s^n | E = 0) \cdot \frac{N(s | s^n)}{n}. \quad (287)$$

Equation (284) comes from the independence of event $\{T = i\}$ with events $\{S^n = s^n\}$ and $\{E = 0\}$.

Equation (287) comes from the definition of the number of occurrence $N(s | s^n) = \sum_{i=1}^n \mathbb{1}_{\{s_T = s\}}$.

Since the sequences $s^n \in A_\varepsilon^{*n}$ are typical, we have the following equation:

$$\mathcal{P}_s(s) - \varepsilon \leq \frac{N(s | s^n)}{n} \leq \mathcal{P}_s(s) + \varepsilon. \quad (288)$$

This provides an upper bound and a lower bound on $\mathcal{P}(S_T = s|E = 0)$:

$$\mathcal{P}_s(s) - \varepsilon = \sum_{s^n \in A_\varepsilon^{*n}} \mathcal{P}(S^n = s^n|E = 0) \cdot (\mathcal{P}_s(s) - \varepsilon) \quad (289)$$

$$\leq \mathcal{P}(S_T = s|E = 0) \quad (290)$$

$$\leq \sum_{s^n \in A_\varepsilon^{*n}} \mathcal{P}(S^n = s^n|E = 0) \cdot (\mathcal{P}_s(s) + \varepsilon) \quad (291)$$

$$= \mathcal{P}_s(s) + \varepsilon, \quad (292)$$

$$\Leftrightarrow \left| \mathcal{P}(S_T = s|E = 0) - \mathcal{P}_s(s) \right| \leq \varepsilon. \quad (293)$$

Using the same arguments, we prove that $\mathcal{P}\left((S_T, U_T, Z_T, X_T, Y_T, V_T) = (s, u, z, x, y, v) \mid E = 0\right)$ is closed to the target probability distribution $\mathcal{Q}(s, u, z, x, y, v)$:

$$\left| \mathcal{P}\left((S_T, U_T, Z_T, X_T, Y_T, V_T) = (s, u, z, x, y, v) \mid E = 0\right) - \mathcal{Q}(s, u, z, x, y, v) \right| \leq \varepsilon. \quad (294)$$

This concludes the proof of Lemma 9.

Lemma 10 *The i.i.d. property of the information source and state informations (U, S, Z) induces the following equation:*

$$\sum_{i=1}^n I(S_i; U^{-i}, S_{i+1}^n | U_i, E = 0) \leq n \cdot \varepsilon. \quad (295)$$

Remark A.5 The proof of Lemma 10 extends to the following equations that intervene in the proofs of Theorems II.1, III.1 and IV.1:

$$\sum_{i=1}^n I(U_{i+1}^n, S_{i+1}^n; U_i, S_i | E = 0) \leq n \cdot \varepsilon, \quad (296)$$

$$\sum_{i=1}^n H(Z_i | X^n, Z^{-i}, U_{i+1}^n, S_{i+1}^n, U_i, S_i, E = 0) - \sum_{i=1}^n H(Z_i | U_i, S_i, E = 0) \leq n \cdot \varepsilon, \quad (297)$$

$$\sum_{i=1}^n H(Z_i | U_i, S_i, E = 0) - H(Z^n | U^n, S^n, E = 0) \leq n \cdot \varepsilon, \quad (298)$$

$$\sum_{i=1}^n I(U_i; Z^{-i}, U^{i-1} | Z_i, E = 0) \leq n \cdot \varepsilon. \quad (299)$$

In the proof of Theorem IV.1, the random variables U and S are independent. Hence this gives the following equation:

$$\sum_{i=1}^n I(S_i; U^n, S_{i+1}^n | E = 0) \leq n \cdot \varepsilon. \quad (300)$$

Proof A.6 (Proof of Lemma 10)

$$\sum_{i=1}^n I(S_i; U^{-i}, S_{i+1}^n | U_i, E = 0) = \sum_{i=1}^n H(S_i | U_i, E = 0) - \sum_{i=1}^n H(S_i | U^n, S_{i+1}^n, E = 0) \quad (301)$$

$$= \sum_{i=1}^n H(S_i | U_i, E = 0) - H(S^n | U^n, E = 0). \quad (302)$$

Equations (301) and (302) come from the properties of the mutual information.

$$H(S^n | U^n, E = 0) = \frac{1}{\mathcal{P}(E = 0)} \cdot \left[H(S^n | U^n, E) - \mathcal{P}(E = 1) \cdot H(S^n | U^n, E = 1) \right] \quad (303)$$

$$\geq H(S^n | U^n, E) - \mathcal{P}(E = 1) \cdot H(S^n | U^n, E = 1) \quad (304)$$

$$\geq H(S^n | U^n) - I(E; S^n | U^n) - \mathcal{P}(E = 1) \cdot n \cdot \log_2 |\mathcal{S}| \quad (305)$$

$$\geq H(S^n | U^n) - H(E) - \mathcal{P}(E = 1) \cdot n \cdot \log_2 |\mathcal{S}| \quad (306)$$

$$\geq H(S^n | U^n) - 1 - \mathcal{P}(E = 1) \cdot n \cdot \log_2 |\mathcal{S}| \quad (307)$$

$$= n \cdot \left(H(S | U) - \frac{1}{n} - \mathcal{P}(E = 1) \cdot \log_2 |\mathcal{S}| \right). \quad (308)$$

Equation (303) comes from the definition of the entropy. Note that $\mathcal{P}(E = 0) \neq 0$, since error probability $\mathcal{P}(E = 1) < 1$ is low.

Equation (304) comes from $\mathcal{P}(E = 0) \leq 1$ that implies $\frac{1}{\mathcal{P}(E=0)} \geq 1$.

Equations (305), (306), (307) come from the properties of the mutual information.

Equation (308) comes from the i.i.d. property of the sequences of source U^n and states S^n .

$$H(S_i|U_i, E = 0) = \frac{1}{\mathcal{P}(E = 0)} \cdot \left[H(S_i|U_i, E) - \mathcal{P}(E = 1) \cdot H(S_i|U_i, E = 1) \right] \quad (309)$$

$$\leq \frac{1}{\mathcal{P}(E = 0)} \cdot H(S_i|U_i, E) \quad (310)$$

$$= H(S_i|U_i, E) + \frac{1 - \mathcal{P}(E = 0)}{\mathcal{P}(E = 0)} \cdot H(S_i|U_i, E) \quad (311)$$

$$\leq H(S_i|U_i,) + \frac{1 - \mathcal{P}(E = 0)}{\mathcal{P}(E = 0)} \cdot \log_2 |\mathcal{S}| \quad (312)$$

$$= H(S|U) + \frac{\mathcal{P}(E = 1)}{1 - \mathcal{P}(E = 1)} \cdot \log_2 |\mathcal{S}|. \quad (313)$$

Equations (309), (310), (311), (312) come from the properties of the entropy.

Equation (313) comes from the i.i.d. property of the source U_i and states S_i , for all $i \in \{1, \dots, n\}$.

$$\sum_{i=1}^n I(S_i; U^{-i}, S_{i+1}^n | U_i, E = 0) \quad (314)$$

$$= \sum_{i=1}^n H(S_i|U_i, E = 0) - H(S^n|U^n, E = 0) \quad (315)$$

$$\leq n \cdot \left(H(S|U) + \frac{\mathcal{P}(E = 1)}{1 - \mathcal{P}(E = 1)} \cdot \log_2 |\mathcal{S}| - H(S|U) + \frac{1}{n} + \mathcal{P}(E = 1) \cdot \log_2 |\mathcal{S}| \right) \quad (316)$$

$$= n \cdot \left(\frac{\mathcal{P}(E = 1)^2}{1 - \mathcal{P}(E = 1)} \cdot \log_2 |\mathcal{S}| + \frac{1}{n} \right) \quad (317)$$

$$\leq n \cdot \varepsilon. \quad (318)$$

Equation (315) comes from equations (308) and (313).

Equation (318) comes from the hypothesis of small error probability $\mathcal{P}(E = 1)$ and large length of codewords $n \in \mathbb{N}$, hence $\varepsilon \geq \frac{\mathcal{P}(E=1)^2}{1-\mathcal{P}(E=1)} \cdot \log_2 |\mathcal{S}| + \frac{1}{n}$.

This concludes the proof of Lemma 10.

Lemma 11 Independence of random variable T with (U_T, S_T) induces the following equation:

$$I(T; S_T, U_T | E = 0) = 0. \quad (319)$$

This implies: $I(T; S_T | U_T, E = 0) = I(T; S_T | E = 0) = I(T; U_T | E = 0) = 0$.

Proof A.7 (Proof of Lemma 11) The i.i.d. property of the source (U, S) implies that for all $i, j \in \{1, \dots, n\}$, we have:

$$\mathcal{P}\left((U_T, S_T) = (u, s) | T = i, E = 0\right) = \mathcal{P}\left((U_T, S_T) = (u, s) | T = j, E = 0\right) \quad (320)$$

$$\implies \mathcal{P}\left((U_T, S_T) = (u, s) | T = i, E = 0\right) = \mathcal{P}\left((U_T, S_T) = (u, s) | E = 0\right), \forall i \in \{1, \dots, n\} \quad (321)$$

$$\implies I(T; S_T, U_T | E = 0) = 0. \quad (322)$$

REFERENCES

- [1] P. Cuff and C. Schieler, “Hybrid codes needed for coordination over the point-to-point channel,” in *49th Annual Allerton Conference on Communication, Control, and Computing*, pp. 235–239, Sept 2011.
- [2] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with state information at the decoder,” *IEEE Transactions on Information Theory*, vol. 22, pp. 1–11, Jan. 1976.
- [3] S. I. Gel’fand and M. S. Pinsker, “Coding for channel with random parameters,” *Problems of Control and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [4] M. Le Treust, “Correlation between channel state and information source with empirical coordination constraint,” *IEEE Proc. of the Information Theory Workshop (ITW), Hobart, Australia*, pp. 272–276, Nov. 2014.
- [5] M. Le Treust, “Correlation between channel state and information source with empirical coordination constraint – technical report,” tech. rep., <https://sites.google.com/site/maellestreust/B.pdf?attredirects=0&d=1>, 2014.
- [6] N. Merhav and S. Shamai, “On joint source-channel coding for the wyner-ziv source and the gel’fand-pinsker channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 11, pp. 2844–2855, 2003.
- [7] M. Le Treust, “Empirical coordination for the joint source-channel coding problem,” *submitted to IEEE Trans. on Information Theory*, 2014.
- [8] P. Cuff, “Distributed channel synthesis,” *IEEE Transactions on Information Theory*, vol. 59, pp. 7071–7096, Nov 2013.
- [9] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, Dec. 2011.
- [10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. 1981.